A MODIFIED GENERALIZED ESTIMATOR FOR EFFICIENT POPULATION PARAMETER ESTIMATION IN TWO-PHASE STRATIFIED RANDOM SAMPLING

[1]Adenekan, Q. A., [1]Akintunde. A. A., [1]Olayiwola, O. A., [2]Osinuga, I. A., [1]Wale-Orojo, O. A., [1]Ajayi, A. O.

1. Department of Statistics, Federal University of Agriculture, Abeokuta, Nigeria.

2. Department of Mathematics, Federal University of Agriculture, Abeokuta, Nigeria.

Abstract

Sampling is often conducted in multiple phases to optimize resources, reduce costs, and improve accuracy. Two-phase sampling leverages on auxiliary information to enhance efficiency. However, most existing estimators primarily focused on homogeneous populations. Hence, leaving a gap in addressing heterogeneous populations. This study introduced a modified generalized estimator within a stratified two-phase sampling framework, incorporated auxiliary information at both sampling and estimation stages. The motivation behind this study was to develop relatively more efficient estimators with reduced mean square error (MSE) for heterogeneous populations. The objective was to derive and evaluate the performance of the proposed estimators using real-life datasets and Monte Carlo simulations. Two real-life datasets were utilized: the first dataset from the Joint Admissions and Matriculation Board (JAMB) records on applicants by gender for 2017 and 2018, with 2018 serving as the variable of interest and 2017 as the auxiliary variable. The second dataset consists of the enrolments of public primary school pupils and number of teachers by gender and local government for the 2018/2019 academic session. The Taylor series expansion up to the second-degree approximation was applied in deriving the MSE of the proposed estimators. The efficiency of the estimators was assessed using real-life datasets and a Markov Chain Monte Carlo (MCMC) simulation across varying sample sizes (5, 10, and 50). The result demonstrated decreasing MSE values with increasing sample sizes and established that the proposed modified generalized estimators effectively minimized MSE and maximized efficiency in two-phase stratified random sampling as compared to the existing work proposed by Ashish *et al* (2023) and the classical mean estimator under two-phase stratified sampling.

Keywords: Two-phase sampling, Stratified sampling, Auxiliary Variable, Mean square error, Percentage relative efficiency.

1.0 Introduction

The goal of conducting sample survey is to estimate population quantities more efficiently, faster and at reduced cost. The use of auxiliary information from a variable correlated to the variable of interest together with the study variable helps increase precision, consistency and performance of the Estimators. The process of using information gathered from a subset of the population to make inferences, decisions, or conclusions about the population parameter is known as estimation. The basic procedures for estimating the population parameters by using auxiliary information are the ratio, product and regression methods of estimations. In the listed methods, the prior information on the auxiliary variable(s) is utilized at the estimation stage but in the absence of such prior information. Neyman (1934) originally presented the potent approach of two-phase sampling for the purpose of stratification, thereby establishing the use of auxiliary information. Regression and ratio estimation methods are employed in two-phase sampling to determine the finite population mean. Cochran (1940) emphasized how important it has become to increase the precision of the estimators of population parameters like the mean and variance of a variable under study with the use of auxiliary variable. A lot of studies have been carried out in order to increase the precision of estimators in which Sarndal et al. (1992) created a framework for estimating population parameters in two-phase sampling with the use of calibration technique which modifies the sample weight. Also, Okafor (2002) refined the use of auxiliary variable in increasing precision of the estimator of a population parameter. Singh and Ruiz (2007) suggested double sampling ratio-product estimator of a finite population mean in sample surveys. Singh and Vishwakarma (2008) proposed a family of estimators of population mean using auxiliary information in stratified sampling, Koyuncu and Kadilar (2009) proposed a family of estimators of population mean using two auxiliary variables in stratified random sampling. Choudhury and Singh (2012) worked on a class of chain ratio-product type estimators with auxiliary variables under double sampling scheme while Hamad *et al* (2013) proposed a regression type estimator with two auxiliary variables for two-phase sampling. Sanuullah *et al* (2014) proposed a generalized exponential chain ratio estimator under stratified two-phase random sampling. Vishwakarma and Kumar (2015) suggested an efficient class of estimators for the mean of a finite population in two-phase sampling using multi-auxiliary variates. Kumar and Vishwakarma (2017) proposed estimation of mean in double sampling using exponential technique on multi-auxiliary variates. Shabbir and Gupta (2017) looked into generalised exponential chain ratio estimators under stratified two-phase random sampling.

Recently, Olayiwola et al (2020) and many others used the auxiliary information to increase the level of precision of estimates. In 2016, Olayiwola *et al* proposed a modified regression estimator in double sampling. Kumar and Vishwakarma (2020) proposed a generalized classes of regression-cum-ratio estimators of population mean in stratified random sampling, Zeeshan *et al* (2021) developed an efficient variant of dual to product and ratio estimators in sample surveys. Zaman and Kadilar (2021) proposed exponential ratio and product type estimators of the mean in stratified two phase sampling. Also, Kumar and Tiwari (2022) suggested a composite class of ratio estimators for a finite population mean in two-phase sampling. Olayiwola *et al* (2021) established a modified exponential-type estimator for population mean with two auxiliary variables in two phase sampling,

While various survey sampling techniques are currently in use, gaps remain in their ability to provide precise and accurate estimates for specific demographic factors (population parameters). The two-phase sampling design was utilized for the estimation of concerned population parameter (*Ashish et al*, 2023). Although much research has been conducted on two-phase sampling and stratified sampling, relatively few studies have explored two-phase stratified sampling using ratio-type estimators.

This work extended the work of Ashish *et al* (2023) and proposed modified generalized estimators for estimating population parameters in stratified two-phase sampling, with the goal of achieving more precise, accurate, and reliable estimates for certain population parameters.

ISSN NUMBER: 1116-249X

The objective is to derive and evaluate the performance of the proposed estimators using real datasets and Monte Carlo simulations.

2.0 Methodology

In this study, two-phase stratified random sampling was adopted in such a way that the first phase sample was taken for the estimation of the auxiliary characteristics of the population in the various strata. In the second phase, sub-samples are taken from the first phase sample by selecting the observations of the study variable and the auxiliary variable from each stratum.

2.1 Sampling Design

To estimate the mean $\bar{Y}$ of a variable of interest with set of values $Y_1,\ Y_2, \dots ,\ Y_N$ and the mean $\bar{X}$ of an auxiliary variable with the set of values $X_1,\ X_2, \dots ,\ X_N$ correlated to the study variable Y of a finite population, it is assumed that the population is stratified, which implies that it has been divided into k non-overlapping strata or groups of sizes $N_1,\ N_2, \dots ,\ N_h$ such that $\sum_{i=1}^{k} N_h$. Let $\bar{y}_{st} = \sum_{h=1}^{k} W_h \bar{y}_h$ and $\bar{x}_{st} = \sum_{h=1}^{k} W_h \bar{x}_h$ be the stratified means of study variable and auxiliary variable respectively and let $W_h = \frac{N_h}{N}$ be the stratum weight. Let the i^{th} sample unit of the study variable and auxiliary variable in the h^{th} stratum be denoted by y_{hi} and x_{hi} respectively. A first-phase stratified sample of size n_h' was selected from h^{th} stratum to estimate $\bar{x}_h$ and another stratified sample of size n_h was selected from the first-phase sample to obtain information on the study variable as well as the auxiliary variable for the second-phase. Let $\bar{y}_h = n_h^{-1} y_{hi}$and $\bar{x}_h = n_h^{-1} x_{hi}$ denote the h^{th} stratum sample means for the study and auxiliary variable. Then for the h^{th} stratum, let us define

$$\varepsilon_0' = \frac{\bar{y}_{st} - \bar{Y}}{\bar{Y}} \tag{1}$$

$$\varepsilon_1 = \frac{\bar{x}_{st} - \bar{X}}{\bar{X}} \tag{2}$$

$$\varepsilon_1' = \frac{\bar{x}_{st}' - \bar{X}}{\bar{X}} \tag{3}$$

The sample variance of the auxiliary variable is

$$S_{x_h}^2 = \frac{1}{n-1} \sum_{h=1}^{n} \left(x_h - \bar{X}_h \right)^2 \tag{4}$$

The sample variance of the study variable is

$$S_{y_h}^2 = \frac{1}{n-1} \sum_{h=1}^{n} \left(y_h - \bar{Y}_h \right)^2 \tag{5}$$

ISSN NUMBER: 1116-249X

The covariance between y_h and x_h is

$$S_{y_h x_h} = \frac{1}{n-1}\sum_{h=1}^{n}\left(x_h - \overline{X}_h\right)\left(y_h - \overline{Y}_h\right) \tag{6}$$

The correlation between y_h and x_h is

$$\rho_{y_h x_h} = \frac{S_{yx}}{S_y S_x} \tag{7}$$

The coefficient of variation of y_h is

$$C_{y_h} = \frac{S_{y_h}}{\overline{Y}_h} \tag{8}$$

The coefficient of variation of x_h is

$$C_{x_h} = \frac{S_{x_h}}{\overline{X}_h} \tag{9}$$

Such that

$$E(\in_0) = E\left(\in_1\right) = E(\in_2) = 0 \tag{10}$$

$$E(\varepsilon_1 \varepsilon_1') = \sum W_h^2 \theta_h' C_{x_h}^2 \tag{11}$$

$$\mathrm{V}(\overline{y}_{st}) = \sum \mathrm{W}_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) s_y^2 = \sum \mathrm{W}_h^2 \theta_h s_y^2 \tag{12}$$

$$E(\varepsilon_0^2) = \sum W_h^2 \theta_h C_{y_h}^2 , \tag{13}$$

$$E(\varepsilon_1^2) = \sum W_h^2 \theta_h C_{x_h}^2 , \tag{14}$$

$$E(\varepsilon_1'^2) = \sum W_h^2 \theta_h' C_{x_h}^2 , \tag{15}$$

ISSN NUMBER: 1116-249X

$$E(\varepsilon_0\varepsilon_1) = \sum W_h^2 \theta_h \rho_{y_h x_h} C_{x_h} C_{y_h} \tag{16}$$

$$E(\varepsilon_0\varepsilon_1') = \sum W_h^2 \theta_h' \rho_{y_h x_h} C_{x_h} C_{y_h} \tag{17}$$

Where

$$\theta_h = \frac{1}{n_h} - \frac{1}{N_h} \tag{18}$$

$$\theta_h' = \frac{1}{n_h'} - \frac{1}{N_h} \tag{19}$$

$$\theta_d = \theta_h - \theta_h' = \frac{1}{n_h} - \frac{1}{n_h'} \tag{20}$$

The classical ratio and product estimators in stratified two-phase sampling are given by

$$\bar{y}_{st}^{R} = \bar{y}_{st} \frac{\bar{x}_{st}'}{\bar{x}_{st}} = \sum W_h \bar{y}_h \left(\frac{\sum W_h x_h'}{\sum W_h x_h} \right) \tag{21}$$

and

$$\bar{y}_{st}^{P} = \bar{y}_{st} \frac{\bar{x}_{st}}{\bar{x}_{st}'} = \sum W_h \bar{y}_h \left(\frac{\sum W_h x_h}{\sum W_h x_h'} \right) \tag{22}$$

the MSE of $\bar{y}_{st}^{R}$ and $\bar{y}_{st}^{P}$ are given by

$$MSE(\bar{y}_{st}^{R}) = \bar{Y}^2 \left[(\sum W_h^2 \theta_h C_{y_h}^2 - 2(\sum W_h^2 \theta_d \rho_{y_h x_h} c_{x_h} c_{y_h}) + (\sum W_h^2 \theta_d c_{x_h}^2) \right] \tag{23}$$

$$MSE(\bar{y}_{st}^{P}) = \bar{Y}^2 \left[(\sum W_h^2 \theta_h C_{y_h}^2 + 2(\sum W_h^2 \theta_d \rho_{y_h x_h} c_{x_h} c_{y_h}) + (\sum W_h^2 \theta_d c_{x_h}^2) \right] \tag{24}$$

Ashish et al (2023) formulated a generalized class of ratio-cum-product estimator for the estimation of a finite population mean in simple random two-phase sampling. Their proposed estimator follows:

$$T = \bar{y} \left[\left\{ \frac{\alpha \bar{x}' + \gamma}{\alpha \bar{x} + \gamma} \right\} + (1-k) \left\{ \frac{\alpha \bar{x} + \gamma}{\alpha \bar{x}' + \gamma} \right\} \right] \tag{25}$$

ISSN NUMBER: 1116-249X

The approximate MSE of *T* was given by

$$MSE(T) = \overline{Y}^2 \left[f_1 C_Y^2 + (2k-1) f_3 \psi \left\{ (2k-1)\psi C_X^2 - 2\rho_{YX} C_Y C_X \right\} \right] \tag{26}$$

Where $\psi = \dfrac{\alpha \bar{x}}{\alpha \bar{x} + y}$

And optimum value which minimizes the MSE of T, was obtained as:

$$\psi_{opt} = \frac{1}{(2k-1)} \frac{\rho_{YX} c_Y}{c_X} \tag{27}$$

the minimum attainable MSE of T was given as:

$$\text{MSE(T)}_{\min} = \overline{Y}^2 C_Y^2 (f_1 - f_3 \rho_{YX}^2) \tag{28}$$

Such that the notations used were defined as follows:

$$f_1 = \left(\frac{1}{n} - \frac{1}{N}\right),\ f_2 = \left(\frac{1}{n'} - \frac{1}{N}\right),\ f_3 = f_1 - f_2 = \left(\frac{1}{n} - \frac{1}{n'}\right).$$

$$C_Y = \frac{S_y}{\overline{Y}},\ C_X = \frac{S_x}{\overline{X}},\ \rho_{YX} = \frac{S_{yx}}{S_y S_x},$$

$$S_Y^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (y_i - \overline{Y})^2,\ S_x^2 = \frac{1}{(N-1)} \sum_{i=1}^{N} (x_i - \overline{X})^2,\ \text{and } S_{yx} = \frac{1}{(N-1)} \sum_{i=1}^{N} (y_i - \overline{Y})(x_i - \overline{X}).$$

2.2 Proposed Generalised Estimator and its properties

Motivated by Ashish *et al* (2023), a generalised ratio-cum-product mean estimator was formulated under stratified two-phase sampling to be used to estimate the mean of variable of interest using auxiliary variable given by

$$\text{T}_{st_2} = \bar{y}_{st} \left[c \left\{ \frac{\alpha \bar{x}'_{st} + b}{\alpha \bar{x}_{st} + b} \right\} + (1-c) \left\{ \frac{\alpha \bar{x}_{st} + b}{\alpha \bar{x}'_{st} + b} \right\} \right]^g \tag{29}$$

Where c, x, b are scalar quantities. The optimum values of these scalars can be obtained by minimizing the MSE of the proposed estimator T_{st}. In addition, g is a suitable constant, in which when g=0, we have the classical statified ratio estimator also g and c may take different values to get classical ratio or product type estimator respectively, which can be: g=1, c=1, α=1, b=0

$$\text{T}_{st_2} = \bar{y}_{st} \left[\frac{\bar{x}'_{st}}{\bar{x}_{st}} \right]$$

g=1, c=0, α=1, b=0

ISSN NUMBER: 1116-249X

$$\mathrm{T}_{st_2} = \bar{y}_{st}\left[\frac{\bar{x}_{st}}{\bar{x}'_{st}}\right]$$

To obtain the MSE of the proposed estimator, we may represent the proposed generalized estimator given in Eq. (9) in term of ε as follows

$$\mathrm{T}_{st_2} = \bar{Y}(1+\varepsilon_0)\left[c\left\{\frac{\alpha\bar{X}(1+\varepsilon_1^{'})+b}{\alpha\bar{X}(1+\varepsilon_1)+b}\right\}+(1-c)\left\{\frac{\alpha\bar{X}(1+\varepsilon_1)+b}{\alpha\bar{X}(1+\varepsilon_1^{'})+b}\right\}\right]^g \tag{30}$$

$$(T_{st_2}) = \bar{Y}(1+\varepsilon_0)\left[c\left\{\frac{\alpha\bar{X}+\alpha\bar{X}\varepsilon_1^{'}+b}{\alpha\bar{X}+\alpha\bar{X}\varepsilon_1)+b}\right\}+(1-c)\left\{\frac{\alpha\bar{X}+\alpha\bar{X}\varepsilon_1+b}{\alpha\bar{X}+\alpha\bar{X}\varepsilon_1^{'})+b}\right\}\right]^g$$

$$(T_{st_2}) = \bar{Y}(1+\varepsilon_0)\left[c\left\{\frac{(\alpha\bar{X}+b)+\alpha\bar{X}\varepsilon_1^{'}}{(\alpha\bar{X}+b)+\alpha\bar{X}\varepsilon_1}\right\}+(1-c)\left\{\frac{(\alpha\bar{X}+b)+\alpha\bar{X}\varepsilon_1}{(\alpha\bar{X}+b)+\alpha\bar{X}\varepsilon_1^{'}}\right\}\right]^g$$

$$(\mathrm{T}_{st_2}) = \bar{Y}(1+\varepsilon_0)\left[c\left(\frac{\frac{\alpha\bar{X}+b}{\alpha\bar{X}+b}+\frac{\alpha\bar{X}\varepsilon_1^{'}}{\alpha\bar{X}+b}}{\frac{\alpha\bar{X}+b}{\alpha\bar{X}+b}+\frac{\alpha\bar{X}\varepsilon_1}{\alpha\bar{X}+b}}\right)+(1-c)\left(\frac{\frac{\alpha\bar{X}+b}{\alpha\bar{X}+b}+\frac{\alpha\bar{X}\varepsilon_1}{\alpha\bar{X}+b}}{\frac{\alpha\bar{X}+b}{\alpha\bar{X}+b}+\frac{\alpha\bar{X}\varepsilon_1^{'}}{\alpha\bar{X}+b}}\right)\right]^g$$

Put $\frac{\alpha\bar{X}}{\alpha\bar{X}+b} = \varphi$

Using Taylor series to expand

where, $T_{st2} = \bar{y}_{st}\, f(\varphi)$; $\bar{y}_{st} = (1+\varepsilon_0)\bar{Y}$ *so,* $T_{st2} = \bar{Y}(1+\varepsilon_0)\, f(\varphi)$

and $\mathrm{f}(\varphi) = \mathrm{f(a)} + \frac{(\varphi-a)}{1!} f'(a) + \frac{(\varphi-a)^2}{2!} f''(a)$

$$f(\varphi) = [c(1+\varphi\varepsilon_1^{'})(1+\varphi\varepsilon_1)^{-1} + (1-c)(1+\varphi\varepsilon_1)(1+\varphi\varepsilon_1^{'})^{-1}]^g \tag{32}$$

ISSN NUMBER: 1116-249X

$$f'(\varphi) = g\begin{bmatrix}[c(1+\varphi\varepsilon_1^{'})(1+\varphi\varepsilon_1)^{-1}+(1-c)(1+\varphi\varepsilon_1)(1+\varphi\varepsilon_1^{'})^{-1}]^{g-1}\times \\ [c(\dfrac{(1+\varphi\varepsilon_1)\varepsilon_1^{'}-(1+\varphi\varepsilon_1^{'})\varepsilon_1}{(1+\varphi\varepsilon_1)^2}+(1-c)\dfrac{(1+\varphi\varepsilon_1^{'})\varepsilon_1-(1+\varphi\varepsilon_1)\varepsilon_1^{'}}{(1+\varphi\varepsilon_1^{'})^2}\end{bmatrix}$$

$$= g\begin{bmatrix}[c(1+\varphi\varepsilon_1^{'})(1+\varphi\varepsilon_1)^{-1}+(1-c)(1+\varphi\varepsilon_1)(1+\varphi\varepsilon_1^{'})^{-1}]^{g-1}\times \\ \left[c\left(\dfrac{\varepsilon_1^{'}-\varepsilon_1}{(1+\varphi\varepsilon_1)^2}\right)+(1-c)\left(\dfrac{\varepsilon_1-\varepsilon_1^{'}}{(1+\varphi\varepsilon_1^{'})^2}\right)\right]\end{bmatrix} \tag{33}$$

$$f''(\varphi) = V\frac{\partial u}{\partial\varphi}+U\frac{\partial v}{\partial\varphi}$$

$$U = g\left[c(1+\varphi\varepsilon_1^{'})(1+\varphi\varepsilon_1)^{-1}+(1-c)(1+\varphi\varepsilon_1)(1+\varphi\varepsilon_1^{'})^{-1}\right]^{g-1}$$

$$\frac{\partial u}{\partial\varphi} = g(g-1)\left[c(1+\varphi\varepsilon_1^{'})(1+\varphi\varepsilon_1)^{-1}+(1-c)(1+\varphi\varepsilon_1)(1+\varphi\varepsilon_1^{'})^{-1}\right]^{g-2}\times$$

$$\left[c\left(\frac{(1+\varphi\varepsilon_1)\varepsilon_1^{'}-(1+\varphi\varepsilon_1^{'})\varepsilon_1}{(1+\varphi\varepsilon_1)^2}\right)+(1-c)\left(\frac{(1+\varphi\varepsilon_1^{'})\varepsilon_1-(1+\varphi\varepsilon_1)\varepsilon_1^{'}}{(1+\varphi\varepsilon_1^{'})^2}\right)\right]$$

$$V = c\left[\left(\frac{\varepsilon_1^{'}-\varepsilon_1}{(1+\varphi\varepsilon_1)^2}\right)+(1-c)\left(\frac{\varepsilon_1-\varepsilon_1^{'}}{(1+\varphi\varepsilon_1^{'})^2}\right)\right]$$

$$\frac{\partial v}{\partial\varphi} = c(\varepsilon_1^{'}-\varepsilon_1)-2\in_1(1+\varphi\varepsilon_1)^{-3}+(1-c)(\varepsilon_1-\varepsilon_1^{'})-2\varepsilon_1^{'}(1+\varphi\varepsilon_1^{'})^{-3}$$

$$=\left[\frac{-2c\in_1(\varepsilon_1^{'}-\varepsilon_1)}{(1+\varphi\varepsilon_1)^3}+\frac{-2(1-c)\varepsilon_1^{'}(\varepsilon_1-\varepsilon_1^{'})}{(1+\varphi\varepsilon_1^{'})^3}\right]$$

ISSN NUMBER: 1116-249X

$$f''(\varphi)=\left[\begin{array}{l}\left(c\left(\dfrac{\varepsilon_1'-\varepsilon_1}{(1+\varphi\varepsilon_1)^2}\right)+(1-c)\left(\dfrac{\varepsilon_1'-\varepsilon_1}{(1+\varphi\varepsilon_1)^2}\right)\right)\times g(g-1)[c(1+\varphi\varepsilon_1')(1+\varphi\varepsilon_1)^{-1}+(1-c)(1+\varphi\varepsilon_1)\\ (1+\varphi\varepsilon_1')^{-1}]^{g-2}\times c\left(\dfrac{(1+\varphi\varepsilon_1)\varepsilon_1'-(1+\varphi\varepsilon_1')\varepsilon_1}{(1+\varphi\varepsilon_1)^2}\right)+(1-c)\dfrac{(1+\varphi\varepsilon_1')\varepsilon_1-(1+\varphi\varepsilon_1)\varepsilon_1'}{(1+\varphi\varepsilon_1')^2}\end{array}\right]+$$

$$\left[g\left[c\left(1+\varphi\varepsilon_1'\right)(1+\varphi\varepsilon_1)^{-1}+(1-c)(1+\varphi\varepsilon_1)(1+\varphi\varepsilon_1')^{g-1}\right]\times\frac{-2c\varepsilon_1(\varepsilon_1'-\varepsilon_1)}{(1+\varphi\varepsilon_1)^3}-\frac{-2(1-c)\varepsilon_1'(\varepsilon_1-\varepsilon_1')}{(1+\varphi\varepsilon_1')^3}\right] \quad (34)$$

$$f(0)=[c(1+0)(1+0)^{-1}+(1-c)(1+0)(1+0)^{-1}]^g=[c+1-c]^g=1^g=1$$

$$f'(0)=[c(1+0)(1+0)^{-1}+(1-c)(1+0)(1+0)^{-1}]^g\times\left[c\left(\frac{\varepsilon_1'-\varepsilon_1}{(1+0)^2}\right)+(1-c)\left(\frac{\varepsilon_1-\varepsilon_1'}{(1+0)^2}\right)\right]$$

$$=g[c+1-c]^{g-1}\times[c(\varepsilon_1'-\varepsilon_1)+(1-c)(\varepsilon_1-\varepsilon_1')]$$

$$=g[c\varepsilon_1'-c\varepsilon_1+\varepsilon_1-\varepsilon_1'-c\varepsilon_1+c\varepsilon_1']$$

$$=g[2c\varepsilon_1'-2k\varepsilon_1+\varepsilon_1-\varepsilon_1']=g[2c(\varepsilon_1'-\varepsilon_1)-1(\varepsilon_1'-\varepsilon_1)]$$

$$\boldsymbol{f'(0)=g[(2c-1)(\varepsilon_1'-\varepsilon_1)]} \quad \mathbf{(35)}$$

$$f''(0)=\left\{c(\varepsilon_1'-\varepsilon_1)+(1-c)(\varepsilon_1-\varepsilon_1')\right\}\times g(g-1)[c+1(1-c)]^{g-2}\times\left\{c(\varepsilon_1'-\in_1)+(1-c)(\varepsilon_1-\varepsilon_1')\right\}$$
$$+g[c(1-c)]^{g-1}\times\left\{-2c\varepsilon_1(\varepsilon_1'-\in_1)-2(1-c)\varepsilon_1'(\varepsilon_1-\varepsilon_1')\right\}$$

$$=[c\varepsilon_1'-c\varepsilon_1+\in_1-\varepsilon_1'-c\varepsilon_1+c\varepsilon_1']\times g(g-1)\times[c\varepsilon_1'-c\varepsilon_1+\varepsilon_1-\varepsilon_1'-c\varepsilon_1+c\varepsilon_1']+$$
$$g[c(1-c)]^{g-1}\times(-2c\varepsilon_1(\varepsilon_1'-\varepsilon_1)-2(1-c)\varepsilon_1'(\varepsilon_1-\varepsilon_1'))$$

$$=[2c\varepsilon_1'-\varepsilon_1)-1(\varepsilon_1'-\varepsilon_1)]\times g(g-1)\times[2c(\varepsilon_1'-\varepsilon_1)-1(\varepsilon_1'-\varepsilon_1)]+g[(2c\varepsilon_1'^2-2\varepsilon_1'\varepsilon_1+\varepsilon_1'^2-2c\varepsilon_1'^2)]$$

$$=\left[2c\varepsilon_1'-2c\varepsilon_1+\varepsilon_1-\varepsilon_1'\right]\times g(g-1)\left[2c\varepsilon_1'-2c\varepsilon_1+\varepsilon_1-\varepsilon_1'\right]+\left[g\times(2c\varepsilon_1'^2-2\varepsilon_1'\varepsilon_1+\varepsilon_1'^2-2c\varepsilon_1'^2)\right]$$

$$=g(g-1)\left[(2c-1)(\varepsilon_1'-\varepsilon_1)\times(2c-1)(\varepsilon_1'-\varepsilon_1)\right]+g\left[2c\varepsilon_1^2-2\varepsilon_1'\varepsilon_1+\varepsilon_1'^2-2c\varepsilon_1'^2\right]$$

$$f''(0)=g(g-1)[(2c-1)^2(\varepsilon_1'-\varepsilon_1)^2]+g[2c\varepsilon_1^2-2\varepsilon_1'\varepsilon_1+\varepsilon_1'^2-2c\varepsilon_1'^2] \quad (36)$$

ISSN NUMBER: 1116-249X

$$So, T_{st} = \overline{y}_{st}\,(f(\varphi))$$

$$where\ \overline{y}_{st} = (1+\varepsilon_0)\overline{Y} \quad \Rightarrow \quad T_{st} = \overline{Y}(1+\varepsilon_0)f(\varphi)$$

$$and\ \ f(\varphi) = f(a) + \frac{(\varphi-a)}{1!}f'(a) + \frac{(\varphi-a)^2}{2!}f''(a)$$

$$= f(0) + \frac{(\varphi-0)}{1!}f'(0) + \frac{(\varphi-0)^2}{2!}f''(0)$$

Therefore,

$$=1+(\varphi-0)g(2c-1)(\in_1'-\in_1)+(\varphi-0)^2\left[g(g-1)(2c-1)^2(\in_1'-\in_1)^2+g\left\{2c\in_1^2-2\in_1'\in_1+2\in_1'^2-2c\in_1'^2\right\}\right]$$

$$T_{st2} = \overline{Y}(1+\varepsilon_0)\left[1+\varphi g(2c-1)(\varepsilon_1^{'}-\varepsilon_1)+\varphi^2\left\{\begin{matrix} g(g-1)(2c-1)^2(\varepsilon_1^{'}-\varepsilon_1)^2+g(2c\varepsilon_1^{'2}-2\varepsilon_1^{'}\varepsilon_1 \\ +2\varepsilon_1^{'2}-2c\varepsilon_1^{'2}) \end{matrix}\right\}\right] \quad (37)$$

$$T_{st2} = \overline{Y}\left[(1+\varepsilon_0)+\varphi g(2c-1)(1+\varepsilon_0)(\varepsilon_1^{'}-\varepsilon_1)+\varphi^2(1+\varepsilon_0)\left\{\begin{matrix} g(g-1)(2c-1)^2(\varepsilon_1^{'}-\varepsilon_1)^2 \\ +2g(c\varepsilon_1^2-\varepsilon_1^{'}\varepsilon_1+\varepsilon_1^{'2}-c\varepsilon_1^{'2}) \end{matrix}\right\}\right] \quad (38)$$

$$T_{st2} = \overline{Y}+\overline{Y}\left[\varepsilon_0+\varphi g(2c-1)(1+\varepsilon_0)(\varepsilon_1^{'}-\varepsilon_1)+\varphi^2(1+\varepsilon_0)\left\{\begin{matrix} g(g-1)(2c-1)^2(\varepsilon_1^{'}-\varepsilon_1)^2 \\ +2g(c\varepsilon_1^2-\varepsilon_1^{'}\varepsilon_1+\varepsilon_1^{'2}-c\varepsilon_1^{'2}) \end{matrix}\right\}\right] \quad (39)$$

$$T_{st2}-\overline{Y} = \overline{Y}\left[\varepsilon_0+\varphi g(2c-1)(1+\varepsilon_0)(\varepsilon_1^{'}-\varepsilon_1)+\varphi^2(1+\varepsilon_0)\left\{\begin{matrix} g(g-1)(2c-1)^2(\varepsilon_1^{'}-\varepsilon_1)^2 \\ +2g(c\varepsilon_1^2-\varepsilon_1^{'}\varepsilon_1+\varepsilon_1^{'2}-c\varepsilon_1^{'2}) \end{matrix}\right\}\right]$$

$$(T_{st2}-\overline{Y})^2 = \overline{Y}^2\left[\varepsilon_0+\varphi g(2c-1)(1+\varepsilon_0)(\varepsilon_1^{'}-\varepsilon_1)+\varphi^2(1+\varepsilon_0)\left\{\begin{matrix} g(g-1)(2c-1)^2(\varepsilon_1^{'}-\varepsilon_1)^2 \\ +2g(c\varepsilon_1^2-\varepsilon_1^{'}\varepsilon_1+\varepsilon_1^{'2}-c\varepsilon_1^{'2}) \end{matrix}\right\}\right]^2 \quad (40)$$

ISSN NUMBER: 1116-249X

$$(T_{st}-\bar{Y})^2 = \bar{Y}^2\begin{bmatrix} \varepsilon_0^2+\varepsilon_0(1+\varepsilon_0)\varphi g(2c-1)(\varepsilon_1^{'}-\varepsilon_1)+\varepsilon_0\{(1+\varepsilon_0)\varphi^2(g(g-1)(2c-1)^2(\varepsilon_1^{'}-\varepsilon_1)^2 \\ +g(2c\varepsilon_1^2-2\varepsilon_1^{'}\varepsilon_1+2\varepsilon_1^{'2}-2c\varepsilon_1^{'2})\}+\varepsilon_0(1+\varepsilon_0)\varphi g(2c-1)(\varepsilon_1^{'}-\varepsilon_1) \\ +(1+\varepsilon_0)^2\left\{\varphi g(2c-1)(\varepsilon_1^{'}-\varepsilon_1)\right\}^2 \\ +\varphi g(2c-1)(1+\varepsilon_0)(\varepsilon_1^{'}-\varepsilon_1)*\varphi^2(1+\varepsilon_0)\begin{Bmatrix} g(g-1)(2c-1)^2(\varepsilon_1^{'}-\varepsilon_1)^2+ \\ g(2c\varepsilon_1^2-2\varepsilon_1^{'}\varepsilon_1+2\varepsilon_1^{'2}-2c\varepsilon_1^{'2})\end{Bmatrix} \\ +\varepsilon_0*\varphi^2(1+\varepsilon_0)\left\{g(g-1)(2c-1)^2(\varepsilon_1^{'}-\varepsilon_1)^2+g(2c\varepsilon_1^2-2\varepsilon_1^{'}\varepsilon_1+2\varepsilon_1^{'2}-2c\varepsilon_1^{'2})\right\} \\ +\varphi g(2c-1)(1+\varepsilon_0)(\varepsilon_1^{'}-\varepsilon_1)* \\ \varphi^2(1+\varepsilon_0)\left\{g(g-1)(2c-1)^2(\varepsilon_1^{'}-\varepsilon_1)^2+g(2c\varepsilon_1^2-2\varepsilon_1^{'}\varepsilon_1+2\varepsilon_1^{'2}-2c\varepsilon_1^{'2})\right\}+ \\ \varphi^4(1+\varepsilon_0)^2*\left\{g(g-1)(2c-1)^2(\varepsilon_1^{'}-\varepsilon_1)^2+g(2c\varepsilon_1^2-2\varepsilon_1^{'}\varepsilon_1+2\varepsilon_1^{'2}-2c\varepsilon_1^{'2})\right\}^2 \end{bmatrix} \quad (41)$$

$$= \bar{Y}^2\begin{bmatrix} \varepsilon_0^2+\left\{\varphi g(2c-1)(\varepsilon_1^{'}-\varepsilon_1)\right\}^2+\varphi^4(1+\varepsilon_0)^2*\begin{Bmatrix} g(g-1)(2c-1)^2(\varepsilon_1^{'}-\varepsilon_1)^2 \\ +g(2c\varepsilon_1^2-2\varepsilon_1^{'}\varepsilon_1+2\varepsilon_1^{'2}-2c\varepsilon_1^{'2})\end{Bmatrix}^2 \\ +2\varepsilon_0(1+\varepsilon_0)\varphi g(2c-1)(\varepsilon_1^{'}-\varepsilon_1)+2\varepsilon_0*(1+\varepsilon_0)\varphi^2(1+\varepsilon_0)\begin{Bmatrix} g(g-1)(2c-1)^2(\varepsilon_1^{'}-\varepsilon_1)^2+ \\ g(2c\varepsilon_1^2-2\varepsilon_1^{'}\varepsilon_1+2\varepsilon_1^{'2}-2c\varepsilon_1^{'2})\end{Bmatrix} \\ +\varphi g(2c-1)(1+\varepsilon_0)(\varepsilon_1^{'}-\varepsilon_1)*\varphi^2(1+\varepsilon_0)\begin{Bmatrix} g(g-1)(2c-1)^2(\varepsilon_1^{'}-\varepsilon_1)^2+ \\ g(2c\varepsilon_1^2-2\varepsilon_1^{'}\varepsilon_1+2\varepsilon_1^{'2}-2c\varepsilon_1^{'2})\end{Bmatrix} \end{bmatrix} \quad (42)$$

Recall that we neglect the power greater than 2

Therefore; we have

$$(T_{st2}-\bar{Y})^2 = \bar{Y}^2\left[\varepsilon_0^2+\left\{\varphi g(2c-1)(\varepsilon_1^{'}-\varepsilon_1)\right\}^2+0+2\varepsilon_0(1+\varepsilon_0)\varphi g(2c-1)(\varepsilon_1^{'}-\varepsilon_1)+0+0\right] \quad (43)$$

$$= \bar{Y}^2\left[\varepsilon_0^2+\begin{Bmatrix} \varphi g(2c-1)(\varepsilon_1^{'2}-2\varepsilon_1^{'}\varepsilon_1+\varepsilon_1^2+2\varepsilon_0\varepsilon_1^{'2} \\ -4\varepsilon_0\varepsilon_1^{'}\varepsilon_1+2\varepsilon_0\varepsilon_1^2+\varepsilon_0^2\varepsilon_1^{'2}-2\varepsilon_0\varepsilon_1^{'}\varepsilon_1+\varepsilon_0\varepsilon_1^{'2})\end{Bmatrix}+\varphi g(2c-1)(\varepsilon_0\varepsilon_1^{'}-\varepsilon_0\varepsilon_1+\varepsilon_0^2\varepsilon_1^{'}-\varepsilon_0^2\varepsilon_1)\right] \quad (44)$$

We neglect order higher than 2

$$(T_{st}-\bar{Y})^2 = \bar{Y}^2\left[\varepsilon_0^2+\left\{\varphi^2 g^2(2c-1)^2(\varepsilon_1^{'2}-2\varepsilon_1^{'}\varepsilon_1+\varepsilon_1^2)\right\}+2\varphi g(2c-1)(\varepsilon_0\varepsilon_1^{'}-\varepsilon_0\varepsilon_1)\right] \quad (45)$$

Taking the expectation of Eq. (45), such that

$$E(T_{st2}-\bar{Y})^2 = MSE(T_{st2})$$

ISSN NUMBER: 1116-249X

$$E(T_{st}-\bar{Y})^2 = E\left(\bar{Y}^2\left[\varepsilon_0^2+\left\{\varphi^2 g^2(2c-1)^2(\varepsilon_1^{'2}-2\varepsilon_1^{'}\varepsilon_1+\varepsilon_1^2)\right\}+2\varphi g(2c-1)(\varepsilon_0\varepsilon_1^{'}-\varepsilon_0\varepsilon_1)\right]\right)$$

$$MSE(T_{st2})=\bar{Y}^2\left(E\left[\varepsilon_0^2+\left\{\varphi^2 g^2(2c-1)^2(\varepsilon_1^{'2}-2\varepsilon_1^{'}\varepsilon_1+\varepsilon_1^2)\right\}+2\varphi g(2c-1)(\varepsilon_0\varepsilon_1^{'}-\varepsilon_0\varepsilon_1)\right]\right) \quad (46)$$

Taking the expectations and expanding equation (46) produces the following:

$$MSE(T_{st2})=\bar{Y}^2\left[E(\varepsilon_0^2)+\left\{\varphi^2 g^2(2c-1)^2 E(\varepsilon_1^{'2}-2\varepsilon_1^{'}\varepsilon_1+\varepsilon_1^2)\right\}+2\varphi g(2c-1)E(\varepsilon_0\varepsilon_1^{'}-\varepsilon_0\varepsilon_1)\right] \quad (47)$$

$$MSE(T_{st2})=\bar{Y}^2\begin{bmatrix}\sum W_h^2\theta_h C_{y_h}^2+\varphi^2 g^2(2c-1)^2\left\{\sum W_h^2\theta_h^{'}C_{x_h}^2-2\sum W_h^2\theta_h^{'}C_{x_h}^2+\sum W_h^2\theta_h C_{x_h}^2\right\}\\ +2\varphi g(2c-1)\left\{\sum W_h^2\theta_h^{'}\rho_{y_hx_h}C_{y_h}C_{x_h}-\sum W_h^2\theta_h\rho_{y_hx_h}C_{y_h}C_{x_h}\right\}\end{bmatrix}$$

$$=\bar{Y}^2\begin{bmatrix}\sum W_h^2\theta_h C_{y_h}^2+\varphi^2 g^2(2c-1)^2\left\{\sum W_h^2\theta_h C_{x_h}^2-\sum W_h^2\theta_h^{'}C_{x_h}^2\right\}+\\ 2\varphi g(2c-1)\left\{\sum W_h^2\theta_h^{'}\rho_{y_hx_h}C_{y_h}C_{x_h}-\sum W_h^2\theta_h\rho_{y_hx_h}C_{y_h}C_{x_h}\right\}\end{bmatrix}$$

$$=\bar{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2+\varphi^2 g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2(\theta_h-\theta_h^{'})\right\}+2\varphi g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}(\theta_h^{'}-\theta_h)\right\}\right]$$

$$=\bar{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2+\varphi^2 g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2(\theta_h-\theta_h^{'})\right\}-2\varphi g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}(\theta_h-\theta_h^{'})\right\}\right]$$

$$=\bar{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2+\varphi^2 g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2\theta_d\right\}-2\varphi g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}\right] \quad (48)$$

$$MSE(T_{st2})=\bar{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2+\left\{\varphi g(2c-1)\sum W_h^2 C_{x_h}^2-2\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\right\}\varphi g(2c-1)\theta_d\right]$$

$$MSE(T_{st2})=\bar{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2+\varphi g(2c-1)\theta_d\left\{\varphi g(2c-1)\sum W_h^2 C_{x_h}^2-2\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\right\}\right] \quad (49)$$

The optimum value of φ is as follows:

$$\frac{\partial MSE(T_{st})}{\partial\varphi}=0$$

From (28), such that

$$=\bar{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2+\varphi^2 g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2\theta_d\right\}-2\varphi g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}\right]$$

$$\frac{\partial MSE(T_{st2})}{\partial\varphi}=\bar{Y}^2\left[0+2\varphi g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2\theta_d\right\}-2g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}\right]=0$$

$$\Rightarrow\bar{Y}^2\left[2\varphi g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2\theta_d\right\}-2g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}\right]=0 \quad (50)$$

divide through with $\bar{Y}^2$ to have the following:

ISSN NUMBER: 1116-249X

$$2\varphi g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2\theta_d\right\}-2g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}=0$$

$$2\varphi g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2\theta_d\right\}=2g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}$$

$$\varphi=\frac{2g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}}{2g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2\theta_d\right\}} \tag{51}$$

$$\varphi_{opt}=\frac{\rho_{y_hx_h}C_{y_h}}{g(2c-1)C_{x_h}}$$

$$\varphi_{opt}=\frac{1}{g(2c-1)}.\frac{\rho_{y_hx_h}C_{y_h}}{C_{x_h}} \tag{52}$$

substituting the φ_{opt} in the above MSE, we have

$$MSE(T_{st2})_{\min}=\bar{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2+\rho_{y_hx_h}^2C_{y_h}^2\left\{\sum W_h^2\theta_d\right\}-2\rho_{y_hx_h}C_{y_h}\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}\theta_d\right\}\right] \tag{53}$$

$$MSE(T_{st2})_{\min}=\bar{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2+\sum W_h^2\rho_{y_hx_h}^2C_{y_h}^2\theta_d-2\sum W_h^2\rho_{y_hx_h}^2C_{y_h}^2\theta_d\right]$$

$$MSE(T_{st2})_{\min}=\bar{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2-\sum W_h^2\rho_{y_hx_h}^2C_{y_h}^2\theta_d\right]$$

$$MSE(T_{st2})_{\min}=\bar{Y}^2\sum W_h^2C_{yh}^2[\theta_h-\rho_{y_hx_h}^2\theta_d]$$

3.0 Comparing the Relative Efficiency

Here, comparison between the proposed generalized estimator with some existing estimators in two-phase stratified sampling to show:

i. Proposed estimator will give better performance than the combined mean estimator under two-phase stratified sampling if

$$MSE(T_{st2})<V(\bar{y}_{st})$$

Which implies

$$\bar{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2+\varphi^2 g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2\theta_d\right\}-2\varphi g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}\right]<\bar{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2\right]$$

$$\left[\sum W_h^2\theta_h C_{y_h}^2+\varphi^2 g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2\theta_d\right\}-2\varphi g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}\right]<\left[\sum W_h^2\theta_h C_{y_h}^2\right]$$

$$\varphi^2 g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2\theta_d\right\}-2\varphi g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}>0$$

ii. Proposed estimator is more efficient and will give better performance than the combined ratio estimator under two-phase stratified sampling if

$$MSE\left(T_{st2}\right)<MSE\left(\bar{y}_{st}^R\right)$$

Which implies

$$\bar{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2+\varphi^2 g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2\theta_d\right\}-2\varphi g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}\right]$$
$$<\bar{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2-2\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}+\sum W_h^2 C_{x_h}^2\theta_d\right]$$

$$\left[\sum W_h^2\theta_h C_{y_h}^2+\varphi^2 g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2\theta_d\right\}-2\varphi g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}\right]$$
$$<\left[\sum W_h^2\theta_h C_{y_h}^2-2\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}+\sum W_h^2 C_{x_h}^2\theta_d\right]$$

$$\varphi^2 g^2(2c-1)^2\left\{\sum W_h^2 C_{x_h}^2\theta_d\right\}-2\varphi g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}$$
$$<\sum W_h^2 C_{x_h}^2\theta_d-2\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}$$

$$\varphi^2 g^2(2c-1)^2-2\varphi g(2c-1)\left\{\frac{\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}<1-\left\{\frac{2\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}$$

$$\varphi^2 g^2(2c-1)^2-1<2\varphi g(2c-1)\left\{\frac{\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}-\left\{\frac{2\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}$$

$$\varphi^2 g^2(2c-1)^2-1<\left\{\frac{2\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}\left\{\varphi g(2c-1)-1\right\}$$

$$\left[\varphi g(2c-1)-1\right]\left[\varphi g(2c-1)+1\right]<\left\{\frac{2\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}\left\{\varphi g(2c-1)-1\right\}$$

ISSN NUMBER: 1116-249X

$$\left[\varphi g(2c-1)+1\right]<\left\{\frac{2\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}$$

$$\varphi g(2c-1)<\left\{\frac{2\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}-1$$

iii. Proposed estimator is more efficient and will give better performance than the combined product estimator under two-phase stratified sampling if

$$MSE(T_{st2})<MSE\left(\overline{y}_{st}^{P}\right)$$

Which implies

$$\overline{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2+\varphi^2g^2(2c-1)^2\left\{\sum W_h^2C_{x_h}^2\theta_d\right\}-2\varphi g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}\right]$$
$$<\overline{Y}^2\left[\sum W_h^2\theta_h C_{y_h}^2+2\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}+\sum W_h^2C_{x_h}^2\theta_d\right]$$

$$\left[\sum W_h^2\theta_h C_{y_h}^2+\varphi^2g^2(2c-1)^2\left\{\sum W_h^2C_{x_h}^2\theta_d\right\}-2\varphi g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}\right]$$
$$<\left[\sum W_h^2\theta_h C_{y_h}^2+2\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}+\sum W_h^2C_{x_h}^2\theta_d\right]$$

$$\varphi^2g^2(2c-1)^2\left\{\sum W_h^2C_{x_h}^2\theta_d\right\}-2\varphi g(2c-1)\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}$$
$$<\sum W_h^2C_{x_h}^2\theta_d+2\left\{\sum W_h^2\rho_{y_hx_h}C_{y_h}C_{x_h}\theta_d\right\}$$

$$\varphi^2g^2(2c-1)^2-2\varphi g(2c-1)\left\{\frac{\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}<1+\left\{\frac{2\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}$$

$$\varphi^2g^2(2c-1)^2-1<2\varphi g(2c-1)\left\{\frac{\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}+\left\{\frac{2\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}$$

$$\varphi^2g^2(2c-1)^2-1<\left\{\frac{2\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}\left\{\varphi g(2c-1)+1\right\}$$

$$\left[\varphi g(2c-1)-1\right]\left[\varphi g(2c-1)+1\right]<\left\{\frac{2\rho_{y_hx_h}C_{y_h}}{C_{x_h}}\right\}\left\{\varphi g(2c-1)+1\right\}$$

ISSN NUMBER: 1116-249X

$$\left[\varphi g(2c-1)-1\right] < \left\{\frac{2\rho_{y_h x_h} C_{y_h}}{C_{x_h}}\right\}$$

$$\varphi g(2c-1) < \left\{\frac{2\rho_{y_h x_h} C_{y_h}}{C_{x_h}}\right\} + 1$$

4.0 Numerical comparison

4.1 Real life Data

In this study, two real and a simulated dataset were considered, the first dataset is the number of pupils (variable of interest) and number of teachers (auxiliary variable) sourced from the Lagos State Bureau of Statistics (2020 Statistical Digest). The second dataset is the enrolment of Unified Tertiary Matriculation Examination for the year 2018 (variable of interest) and 2017 (auxiliary variable). Gender was used as stratifying variable for both datasets.

4.2 Simulated Data

For simulated dataset, data were simulated using Markov Chain Monte Carlo (MCMC) for sample size of 5, 10 and 50 to ascertain the consistency of the proposed estimator. The data were generated as follows;

$$X \sim SkN(150, 5, 6)$$

$$k \sim N(100, 5)$$

$$\text{w=N(120,10)}$$

$$\text{e=N(0,1)}$$

$$\text{y=b}_0 + \text{b}_1 * \text{X+b}_2 * \text{k+b}_3 * \text{w+e}$$

$$b_0 = 5,\ b_1 = -2, b_2 = 1,\ b_3 = 3.5$$

The proposed estimator was computed for each sample size, 10,000 iterations were use to obtain an accurate estimate of the population mean. The estimated mean square error of the mean estimate is

$$\hat{R}MSE(T_{st}) = \frac{1}{10000}\sum_{1}^{10000}(T_{st} - \bar{Y})^2$$

$$\hat{R}Bias(T_{st}) = \frac{1}{10000}\sum_{1}^{10000}(T_{st} - \bar{Y})$$

ISSN NUMBER: 1116-249X

4.3 Discussion of Result

The descriptive statistics of pupils and teachers in Lagos state for the year 2019 is as seen in Table 1. Female teachers greatly outnumber male teachers, this highlights the dominance of women in the teaching profession at the primary school level in Lagos State in the year 2019. Also, it was discovered in Table 2 that the number of Female UTME Applicants is higher than that of Male Applicant for the year 2017 and 2018 respectively.

The MSE's of the family of the Proposed Estimator are smaller than that of the classical mean (Stratified) and other existing estimators as shown in Table 3 and Table 5. Also, the values of the PRE's of the Proposed Estimators are higher than that of the Classical Mean (Stratified) and other existing estimators, hence, the Proposed Estimators are more efficient.

Also, the MSE's of the family of the Proposed Estimator are smaller than that of the other existing estimators as shown in Table 4 and Table 6, the Percentage Relative Efficiency's (PRE) also proved that the Proposed Estimator are more efficient.

The MCMC simulation shows that the estimator was consistent, as the sample size increases the MSE became smaller (Table 7)

Table 1: No of Pupils and Teachers according to Gender.

	Pupils (Y)		*Teachers (X)*	
	Male	*Female*	*Male*	*Female*
Mean	4005.052632	4161.508772	43.94736842	145.4912281
Standard Error	263.2231672	266.8284524	4.988798295	10.47380438
Median	3662	4064	32	132
Standard Deviation	1987.291332	2014.510638	37.66460116	79.07548898
Sample Variance	3949326.836	4058253.112	1418.62218	6252.932957
Kurtosis	-0.663923934	-0.594371894	8.808132377	1.40520747
Skewness	0.483032583	0.459090187	2.526950525	0.991053659
Number of local council development area	57	57	57	57

ISSN NUMBER: 1116-249X

Table 2: Unified Tertiary Matriculation Examination Registration for 2017 and 2018

	UTME REGISTRATION FOR 2018 (Y)		UTME REGISTRATION FOR 2017 (X)	
	MALE	FEMALE	MALE	FEMALE
Mean	6515.26	7952.18	6606.34	8307.32
Median	5973.50	7882.00	6141.50	7512.50
Variance	19230196.686	13605189.938	19178680.610	16319751.195
Std. Deviation	4385.225	3688.521	4379.347	4039.771
Interquartile Range	6869	5383	6459	5661
Skewness	.465	-.034	.433	.111
Kurtosis	-.683	-.485	-.696	-.447

Table 3: shows the MSE and PRE of the Proposed Estimator using data I

	ESTIMATOR	MSE	PRE
Classical Mean (Stratified sampling)	$\bar{y}_{st}$	361107	100
Family of Proposed Estimator	$T_{st2} = \bar{y}_{st}\left(\frac{\bar{x}_{st}}{\bar{x}'_{st}}\right)$	67552.35	534.5587
	$T_{st2} = \bar{y}_{st}\left(\frac{\bar{x}'_{st}}{\bar{\bar{x}}_{st}}\right)$	17805.201	2028.0982
	$T_{st2} = \bar{y}_{st}\left(\frac{\bar{x}_{st} + c_x}{\bar{x}'_{st} + c_x}\right)$	67542.285	534.6383
	$T_{st2} = \bar{y}_{st}\left(\frac{\bar{x}'_{st} + c_x}{\bar{\bar{x}}_{st} + c_x}\right)$	17798.034	2028.9148
	$T_{st2} = \bar{y}_{st}\left(\frac{c_x\bar{x}_{st} + 1}{c_x\bar{x}'_{st} + 1}\right)$	34936.115	1033.6208
	$T_{st2} = \bar{y}_{st}\left(\frac{c_x\bar{x}'_{st} + 1}{c_x\bar{\bar{x}}_{st} + 1}\right)$	8330.216	4334.9052
	$T_{st2} = \bar{y}_{st}\left(\frac{B_1\bar{x}_{st} + B_1}{B_1\bar{x}'_{st} + B_2}\right)$	65409.105	552.0744
	$T_{st2} = \bar{y}_{st}\left(\frac{B_1\bar{x}'_{st} + B_1}{B_1\bar{\bar{x}}_{st} + B_2}\right)$	16267.683	2219.7811

ISSN NUMBER: 1116-249X

Table 4: shows the performance of the proposed Estimators in comparison with the existing Estimator using Data I

	ESTIMATOR	MSE	PRE
Classical Mean Estimator	$\bar{y}$	3974532	100
Sukhatme 1962	$\bar{y}_{Rd} = \bar{y}\left(\frac{\bar{x}}{\bar{x}'}\right)$	288295.48	1378.631
	$\bar{y}_{Rd} = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)$	279417.52	1422.435
Kwathekar and Ajagonkar 1984	$\bar{y}_{Pd} = \bar{y}\left(\frac{\bar{x}+C_x}{\bar{x}'+C_x}\right)$	288201.83	1379.079
	$\bar{y}_{Rd} = \bar{y}\left(\frac{\bar{x}'+C_x}{\bar{x}+C_x}\right)$	279325.66	1422.903
Family of Existing Estimator (Ashish *et al*, 2023)	$T = \bar{y}\left(\frac{C_x\bar{x}+1}{C_x\bar{x}'+1}\right)$	77798.1	5108.778
	$T = \bar{y}\left(\frac{C_x\bar{x}'+1}{C_x\bar{x}+1}\right)$	75030.84	5297.198
	$T = \bar{y}\left(\frac{B_1\bar{x}+B_2}{B_1\bar{x}'+B_2}\right)$	268715.58	1479.085
	$T = \bar{y}\left(\frac{B_1\bar{x}'+B_2}{B_1\bar{x}+B_2}\right)$	260220.45	1527.371
	Estimators with g=0.25		
Family of proposed Estimator	$T_{st2} = \bar{y}_{st}\left(\frac{\bar{x}_{st}}{\bar{x}'_{st}}\right)$	67552.35	534.5587
	$T_{st2} = \bar{y}_{st}\left(\frac{\bar{x}'_{st}}{\bar{x}_{st}}\right)$	17805.201	2028.0982
	$T_{st2} = \bar{y}_{st}\left(\frac{\bar{x}_{st}+c_x}{\bar{x}'_{st}+c_x}\right)$	67542.285	534.6383

ISSN NUMBER: 1116-249X

	$T_{st2}=\bar{y}_{st}\left(\frac{\bar{x}'_{st}+c_x}{\bar{x}_{st}+c_x}\right)$	17798.034	2028.9148
	$T_{st2}=\bar{y}_{st}\left(\frac{c_x\bar{x}_{st}+1}{c_x\bar{x}'_{st}+1}\right)$	34936.115	1033.6208
	$T_{st2}=\bar{y}_{st}\left(\frac{c_x\bar{x}'_{st}+1}{c_x\bar{x}_{st}+1}\right)$	8330.216	4334.9052
	$T_{st2}=\bar{y}_{st}\left(\frac{B_1\bar{x}_{st}+B_1}{B_1\bar{x}'_{st}+B_2}\right)$	65409.105	552.0744
	$T_{st2}=\bar{y}_{st}\left(\frac{B_1\bar{x}'_{st}+B_1}{B_1\bar{x}_{st}+B_2}\right)$	16267.683	2219.7811

Table 5: shows the MSE and PRE of the family of the Proposed Estimator using data II

	ESTIMATOR	MSE	PRE
Classical Mean (Stratified sampling)	$\bar{y}_{st}$	483769.2	100
Family of Proposed Estimator	$T_{st2}=\bar{y}_{st}\left(\frac{\bar{x}_{st}}{\bar{x}'_{st}}\right)$	254570.84	190.0332
	$T_{st2}=\bar{y}_{st}\left(\frac{\bar{x}'_{st}}{\bar{x}_{st}}\right)$	35788.27	1351.7535
	$T_{st2}=\bar{y}_{st}\left(\frac{\bar{x}_{st}+c_x}{\bar{x}'_{st}+c_x}\right)$	254570.43	190.0335
	$T_{st2}=\bar{y}_{st}\left(\frac{\bar{x}'_{st}+c_x}{\bar{x}_{st}+c_x}\right)$	35787.93	1351.7662
	$T_{st2}=\bar{y}_{st}\left(\frac{c_x\bar{x}_{st}+1}{c_x\bar{x}'_{st}+1}\right)$	250233.08	193.3274
	$T_{st2}=\bar{y}_{st}\left(\frac{c_x\bar{x}'_{st}+1}{c_x\bar{x}_{st}+1}\right)$	32203.79	1502.2119
	$T_{st2}=\bar{y}_{st}\left(\frac{B_1\bar{x}_{st}+B_1}{B_1\bar{x}'_{st}+B_2}\right)$	254479.78	190.1012

ISSN NUMBER: 1116-249X

	$T_{st2} = \bar{y}_{st}\left(\frac{B_1\bar{x}'_{st} + B_1}{B_1\bar{x}_{st} + B_2}\right)$	35713.21	1354.5942

Table 6: shows the performance of the proposed Estimators in comparison with the existing Estimators using Data II

	ESTIMATOR	MSE	PRE
Classical Mean Estimator	$\bar{y}$	16721859	100
Sukhatme 1962	$\bar{y}_{Rd} = \bar{y}\left(\frac{\bar{x}}{\bar{x}'}\right)$	1765002.5	947.4128
	$\bar{y}_{Rd} = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)$	334336.9	5001.4997
Kwathekar and Ajagonkar 1984	$\bar{y}_{Pd} = \bar{y}\left(\frac{\bar{x}+C_x}{\bar{x}'+C_x}\right)$	1764997.2	947.4156
	$\bar{y}_{Rd} = \bar{y}\left(\frac{\bar{x}'+C_x}{\bar{x}+C_x}\right)$	334336.8	5001.501
Family o Existing Estimator (Ashish *et al*, 2023)	$T = \bar{y}\left(\frac{C_x\bar{x}+1}{C_x\bar{x}'+1}\right)$	1708980.4	978.4699
	$T = \bar{y}\left(\frac{C_x\bar{x}'+1}{C_x\bar{x}+1}\right)$	333080	5020.3724
	$T = \bar{y}\left(\frac{B_1\bar{x}+B_2}{B_1\bar{x}'+B_2}\right)$	1763820	948.0479
	$T = \bar{y}\left(\frac{B_1\bar{x}'+B_2}{B_1\bar{x}+B_2}\right)$	334318.3	5001.7777
	Estimators with g=0.25		
Family of proposed Estimator	$T_{st2} = \bar{y}_{st}\left(\frac{\bar{x}_{st}}{\bar{x}'_{st}}\right)$	254570.84	6568.647

	$T_{st2} = \bar{y}_{st}\left(\frac{\bar{x}'_{st}}{\bar{\bar{x}}_{st}}\right)$	35788.27	46724.415
	$T_{st2} = \bar{y}_{st}\left(\frac{\bar{x}_{st} + c_x}{\bar{x}'_{st} + c_x}\right)$	254570.43	6568.657
	$T_{st2} = \bar{y}_{st}\left(\frac{\bar{x}'_{st} + c_x}{\bar{\bar{x}}_{st} + c_x}\right)$	35787.93	46724.854
	$T_{st2} = \bar{y}_{st}\left(\frac{c_x\bar{x}_{st} + 1}{c_x\bar{x}'_{st} + 1}\right)$	250233.08	6682.513
	$T_{st2} = \bar{y}_{st}\left(\frac{c_x\bar{x}'_{st} + 1}{c_x\bar{\bar{x}}_{st} + 1}\right)$	32203.79	51925.128
	$T_{st2} = \bar{y}_{st}\left(\frac{B_1\bar{x}_{st} + B_1}{B_1\bar{x}'_{st} + B_2}\right)$	254479.78	6570.997
	$T_{st2} = \bar{y}_{st}\left(\frac{B_1\bar{x}'_{st} + B_1}{B_1\bar{\bar{x}}_{st} + B_2}\right)$	35713.21	46822.609

Table 7: Markov Chain Monte Carlo Simulation

Sample Size				5		10		50	
SN	α	ϒ	c	REELATIVE BIAS	MSE	REELATIV E BIAS	MSE	REELATIVE BIAS	MSE
1	1	0	0	0.0656	0.1467	-0.0835	0.1364	0.0013	0.0100
2	1	0	1	0.0805	0.1698	-0.0725	0.1386	0.0207	0.0012
3	1	C_x	0	0.0656	0.1467	-0.0835	0.1364	0.0127	0.0100
4	1	C_x	1	0.8049	0.1698	-0.0725	0.1386	0.0207	0.0117
5	Cx	B2x	0	0.0699	0.1513	-0.0806	0.1353	0.0149	0.0103
6	Cx	B2x	1	0.0761	0.1612	-0.0755	0.1363	0.0185	0.0111
7	B1x	B2x	0	0.0658	0.1469	-0.0834	0.1363	0.0128	0.0100
8	B1x	B2x	1	0.0803	0.1694	-0.0727	0.1385	0.0206	0.0117

6.0 Conclusion

In conclusion, this research has successfully devised an innovative, modified generalised research ratio-cum-product estimator tailored for finite population mean estimation in two-phase stratified

ISSN NUMBER: 1116-249X

sampling frameworks. Both theoretical and empirical evidence unequivocally demonstrate the proposed estimator's superiority over extant methodologies, thereby marking a significant contribution to the field of survey sampling.

The practical application of this estimator holds considerable promise for enhancing the precision of population mean estimates, particularly within the context of two-phase stratified sampling designs.

References

Ashish, T., Manish, K. and Sarvash, K.D. 2023. A Generalized Approach for Estimation of a Finite Population mean in Two-Phase Sampling. *Asian Journal of Probability Statistics* 21(3):45-58.

Cochran, W.G. 1940. The Estimation of the Yields of the Cereal Experiments by Sampling for the Ratio of Grain to Total Produce. *Journal of Agricultural Science* 30: 262-275.

http://dx.doi.org/10.1017/S0021859600048012

Choudhury, S., Singh, B.K. 2012: A class of chain ratio-product type estimators with two auxiliary variables under double sampling scheme. *Journal of Korean Statistical Society* 41:247–256

Hamad, N., Hanif, M., and Haider, H. 2013. A regression type estimator with two auxiliary variables for two-phase sampling. *Open Journal of Statistics* 3: 74–78.

Kumar, M. and Vishwakarma, G.K. 2017. Estimation of mean in double sampling using exponential technique on multi-auxiliary variates. *Communications in Mathematics and Statistics* 5(4): 429–445.

Kumar, M. and Tiwari, A. 2021. A generalized class of estimators of population mean in two-phase sampling using two auxiliary variables. *International Journal of Mathematics and Statistics* 22(3): 18-27.

Kumar, M. and Vishwakarma, G.K. 2020. Generalized classes of regression-cum-ratio estimators of population mean in stratified random sampling. Proceedings of the National Academy of Sciences, India Section A: Physical Sciences.90(5): 933-939.

Koyuncu, N. and Kadilar, C. 2009. Family of estimators of population mean using two auxiliary variables in stratified random sampling. *Communication in Statistics-Theory and Methods* 38(14): 2398–2417.

Neyman, Jerzy. 1934. On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society* 97: 558–625.

Okafor, F.C. 2002. Sample Survey Theory and Application. Afro-Orbis Publications Ltd., Nsukka, Nigeria.

Olayiwola, O.M., Popoola, R.A. and Bisira, H.O. 2016. A modified regression estimator for double sampling. *Annals Computer Science Series* 14(1); 44-49.

ISSN NUMBER: 1116-249X

https://scholar.google.com/citations?view_op=view_citation&hl=en&user=UCiRkS0AAAAJ&citation_for_view=UCiRkS0AAAAJ:kNdYIx-mwKoC

Olayiwola, O.M., Audu, A., Ishaq, O.O., Olawoore, S.A. and Ibrahim, A. 2020. A class of ratio estimators of a finite population mean using two auxillary variables under two-phase sample scheme. Royal Statistical Society Nigeria Local Group Annual Conference Proceedings pp80-95. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=UCiRkS0AAAAJ&citation_for_view=UCiRkS0AAAAJ:kNdYIx-mwKoC

Olayiwola, O.M., Akintunde, A.A., Olawoore, S.A., Ishaq, O.O., Ajayi, A.O. and Ojenike, O.T. 2021. A modified exponential-type estimator for population mean with two auxiliary variables in two phase sampling. Royal Statistical Society Nigeria Local Group Annual Conference Proceedings pp121-132. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=UCiRkS0AAAAJ&cstart=20&pagesize=80&citation_for_view=UCiRkS0AAAAJ:rO6llkc54NcC

Sarndal, C. E., Swensson, B. and Wretman, J. 1992. Model assisted survey sampling. Springer Series in Statistics, Springer-Verlag New York.

Singh, H.P. and Ruiz, E. M. 2007. Double sampling ratio-product estimator of a finite population mean in sample surveys. *Journal of Applied Statistics* 34(1): 71-85.

Singh, H.P. and Vishwakarma, G.K. 2008. A family of estimators of population mean using auxiliary information in stratified sampling. *Communication in Statistics- Theory and Methods* 37: 1038–1050.

Vishwakarma, G.K. and Kumar, M. 2015. An efficient class of estimators for the mean of a finite population in two phase sampling using multi-auxiliary variates. *Communications in Mathematics and Statistics* 3(4):477–489.

Zaman, T., Kadilar, C. 2021. Exponential ratio and product type estimators of the mean in stratified two-phase sampling. African Institute of Mathematics Sciences 6(5): 4265–4279

ISSN NUMBER: 1116-249X