NSUK Journal of Science & Technology (NJST), Vol. 7: No. 2, 2020. pp. 101-106. Print: ISSN: 2756- 696X
On-line: ISSN2756-6951
Website: www.nsukjst.org.ng

A MODIFIED GENERALIZED CHAIN RATIO IN REGRESSION ESTIMATOR

***Apantaku, F. S., Olayiwola, O. A., Ajayi, A. O. and Jaiyeola, O. S.**

Department of Statistics, College of Physical Sciences,
Federal University of Agriculture, PMB 2240, Abeokuta, Ogun State, Nigeria.

Corresponding author's email: fsapantaku@yahoo.com

ABSTRACT
Generalized Chain ratio in regression type estimator is efficient for estimating the population mean. Many authors have derived a Generalized Chain ratio in regression type estimator. However, the computation of its Mean Square Error (MSE) is cumbersome based on the fact that several iterations have to be done, hence the need for a modified generalized chain ratio in regression estimator with lower MSE. This study proposed a modified generalized chain ratio in regression estimator which is less cumbersome in its computation. Two data sets were used in this study. The first data were on tobacco production by tobacco producing countries with yield of tobacco (variable of interest), area of land and production in metric tons as the auxiliary variables. The second data were the number of graduating pupils (variable of interest) in Ado-Odo/Ota local government, Ogun state with the number of enrolled pupils in primaries one and five as the auxiliary variables. The mean square errors in the existing and proposed estimators were derived and relative efficiency was determined. The MSE for the existing estimator of tobacco production gave six values; 0.00803, 0.00794, 0.00802, 0.00828, 0.00872 and 0.00933 with 0.00794 as the minimum while the proposed estimator gave 0.00540. The MSEs for the existing estimator for the graduating pupils were 20.732, 11.077, 7.487, 9.961, 18.499 and 33.103 with 6.525 as the minimum while the proposed was 6.523. The results of this study showed that the proposed estimator gave lower MSE for the two data sets, hence it is more efficient.
Keyword: *chain ratio regression, regression estimator, relative efficiency, auxiliary variables*

INTRODUCTION
Regression estimation is a type of model assisted survey estimation approach. Model-assisted estimation (Sârndal, *et al*., 1992) provides inferences and the asymptotic framework which are design based, with the working model only used to improve efficiency. Thus, the regression estimators are model assisted and design based, but not model dependent. Typically, the linear models are used as a working model in regression estimation. Generalized regression (GREG) estimators including ratio and linear regression estimators (Cochran, 1977), best linear unbiased estimators) and post-stratification estimators (Holt and Smith, 1979), are all based on assumed linear models. Generalized Chain ratio in regression type estimator is efficient for estimating the population mean. This estimator uses two auxiliary variables. Several authors have worked on related topic. Khare et al. (2013) proposed a generalized chain ratio in regression estimator for population mean, they got their estimator using an iterative procedure and continued with the process until it converges. The MSE's were obtained. Conditions were also given where the estimators will be more efficient. This continuous iteration will take a lot of time, hence the need to develop another estimator which will not require an iterative procedure and will satisfy all conditions regardless of the population.

LITERATURE REVIEW
The regression type estimators of the population mean or total of y assume advance knowledge of either population mean $\bar{X}$ or total X of the auxiliary variable x. In the absence of such information a large one of size n' is selected to observe x and thereby to estimate X while a subsample of size n is drawn to measure y. Thus the two-phase regression type estimator of population mean Y is

$$\bar{y}_{lr} = \bar{y} + \hat{B}(\bar{x}' - \bar{x})$$

Now suppose that information on yet another auxiliary variable z is available on all units of the population, with population mean $\bar{Z}_N$. Mohanty (1967) suggested the following regression in ratio estimator assuming that the population mean of the second auxiliary variable (z) is known; x being the first auxiliary variable.

$$T_M = \left[\bar{y} + b_{yx}(\bar{x}' - \bar{x})\right]\frac{\bar{Z}}{\bar{z}} \quad (1)$$

Chand (1975) proposed the chain ratio type estimator with similar conditions.

$$T_C = \bar{y}\left(\frac{\bar{X}'}{\bar{x}}\right)\left(\frac{\bar{Z}}{\bar{z}'}\right) \quad (2)$$

Singh *et al*. (2007) suggested a transformed chain ratio type estimator for the population mean $\bar{Y}$ utilizing the known correlation coefficient (ρ_{xz}) of the second auxiliary character through a simple transformation for estimating the population mean of auxiliary character more precisely in the first phase (preliminary) sample as

$$T_S = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)\left(\frac{\bar{Z} + \rho_{xz}}{\bar{z}' + \rho_{xz}}\right) \quad (3)$$

Taking $V_i = z_i + \rho_{xz} (i = 1, 2, \dots, N)$ so that $\bar{V}' + \bar{Z}' + \rho_{xz}$ is the sample mean of the transformed variable v in the first-phase sample and $\bar{V} + \bar{Z} + \rho_{xz}$ is the corresponding population mean.

Srivasta et al. (1990) proposed a generalized two phase sampling estimator for estimating population mean $\bar{Y}$ using information on auxiliary character X, which is given as follows:

$$T_2 = \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right)^{\alpha_1} \quad (4)$$

where α_1 is unknown constant.

The mean square errors of the estimators T_1, T_2 and T_{l1} are given as follows:

$$MSE(T_1) = V(\bar{y}) + \bar{Y}^2\left(\frac{1}{n} - \frac{1}{n'}\right)\left(C_x^2 - 2\rho_{yx}C_yC_x\right) \quad (5)$$

$$MSE(T_2)_{\min} = V(\bar{y}) + \bar{Y}^2\left(\frac{1}{n} - \frac{1}{n'}\right)\rho_{yx}^2 C_y^2 \quad (6)$$

and

$$MSE(T_{l1}) = V(\bar{y}) - \bar{Y}^2\left(\frac{1}{n} - \frac{1}{n'}\right)\rho_{yx}^2 C_y^2 \quad (7)$$

where $V(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2$ (8)

In the case when the population mean $(\bar{X})$ of an auxiliary character x is not known but the population mean $(\bar{Z})$ of an additional auxiliary character z is known. It is suggested to take a large preliminary sample of size $n'(< N)$ from population of size N by using simple random sampling without replacement scheme of sampling and estimate the population mean $\bar{X}$ by $\hat{\bar{X}} = \bar{x}'\frac{\bar{z}}{\bar{z}'}$ which is more efficient in comparison to $\bar{x}'$ if $\rho_{xz} > \frac{1}{2}\frac{C_z}{C_x}$, where $\bar{x}'$ and $\bar{z}'$ are the preliminary sample means based on n' units. Further, a sub-sample of size $n(< n')$ is selected from large preliminary sample of size n' and compute $\bar{y}$ and $\bar{x}$ based on n units. In this case, Chand (1975) and Kiregyera (1984) proposed chain ratio type and ratio in regression estimators, which are given as follows:

$$T_3 = \bar{y}\frac{\bar{x}'}{\bar{x}}\frac{\bar{Z}}{\bar{z}'} \quad \text{and} \quad T_{l2} = \bar{y} + b_{yx}\left[\bar{x}'\frac{\bar{Z}}{\bar{z}'} - \bar{x}\right]$$

METHODOLOGY

Double Sampling Procedure

Double sampling is a sampling method which makes use of auxiliary data where the auxiliary information is obtained through sampling. More precisely, we first take a sample of units strictly to obtain auxiliary information, and then take a second sample where the variable(s) of interest are observed. It will often be the case that this second sample is a subsample of the preliminary sample used to acquire auxiliary information.

Notations used are:

N – Population Size
m – first phase sample size
n – second phase sample size
Y – variable of interest
X, Z – Auxiliary variables
$\bar{X}, \bar{Y}, \bar{Z}$ - population means
$\bar{x}, \bar{y}, \bar{z}$ - sample means
C_x, C_y, C_z – coefficient of variations
β - regression coefficient
ρ - correlation coefficient

Existing Estimator by Khare et al (2013)

Let $\bar{Y}$, $\bar{X}$ and $\bar{Z}$ denote the population mean of study character *y*, and auxiliary character *x* and additional auxiliary character *z* having *jth* values Y_j, X_j and Z_j : $j=1,2,3,\ldots,N$

In the case when population mean of auxiliary character is not known, we draw a large preliminary sample of size n' ($<N$) from the population of size N by using SRSWOR scheme of sampling and estimate the population mean $\bar{X}$ by first phase sample mean $\bar{x}'$ based on n' units. Further, we draw a sub-sample of size n($<n'$) from large preliminary sample size n' and compute $\bar{y}$ and $\bar{x}$ which are the sub-sample based on *n* units. In such case, the double sample ratio and regression are defined by

$$T_1 = \bar{y}\frac{\bar{x}'}{\bar{x}} \quad (9)$$

$$T_{l1} = \bar{y} + b_{yx}(\bar{x}' - \bar{x}) \quad (10)$$

where $b_{yx} = \frac{\hat{S}_{yx}}{\hat{S}_x^2}$, $\bar{x}' = \frac{1}{n'}\sum_{j=1}^{n'} x_j$, $\hat{S}_{yx}$ and $\hat{S}_x^2$ denote the estimates of S_{yx} and S_x^2 based on n units.

The mean square errors of the estimators T_1 and T_{l1} are given as follows:

$$MSE(T_1) = V(\bar{y}) + \bar{Y}^2\left(\frac{1}{n} - \frac{1}{n'}\right)\left(C_x^2 - 2\rho_{yx}C_yC_x\right) \quad (11)$$

$$MSE(T_{l1}) = V(\bar{y}) - \bar{Y}^2\left(\frac{1}{n} - \frac{1}{n'}\right)\rho_{yx}^2 C_y^2 \quad (12)$$

where $V(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2$

In the case when the population mean $(\bar{X})$ of an auxiliary character x is not known but the population mean $(\bar{Z})$ of additional auxiliary character z is known. It is suggested to take a large preliminary sample of size $n'(< N)$ from population of size N by using simple random sampling without replacement scheme of sampling and estimate the population mean X by $\hat{\bar{X}} = \bar{x}'\frac{\bar{z}}{\bar{z}'}$ which is more efficient in comparison to $\bar{x}'$ if $\rho_{xz} > \frac{1}{2}\frac{C_z}{C_x}$ where $\bar{x}'$ and $\bar{z}'$ are the preliminary sample means based on n' units. Further, a sub-sample of size $n(< n')$ is selected from large preliminary sample of size n' and compute $\bar{y}$ and $\bar{x}$ based on n units.

Motivated by Chand (1975) and Kiregyra (1984), Khare et al (2013) now proposed a generalized chain ratio in regression estimator for population mean by using auxiliary characters which is given as follows:

$$T_{l3} = \bar{y} + b_{yx}\left[\bar{x}'\left(\frac{\bar{Z}}{\bar{z}'}\right)^{\alpha} - \bar{x}\right] \quad (13)$$

Using the large sample approximation, the expression for bias of the estimator T_{l3} was given as follows:

$$Bias(T_{l3}) = \theta\left[-\mu_{14} + \mu_{15} + \mu_{24} - \mu_{25} - \alpha\mu_{34} + \alpha\mu_{35} + \frac{f'}{n'}\left\{\frac{\alpha(\alpha+1)}{2}C_z^2 - \alpha\rho_{xz}C_xC_z\right\}\right] \quad (14)$$

where

$$\theta = \bar{X}\frac{S_{yx}}{S_x^2},\ \mu_{14} = Cov(\bar{x}, \hat{S}_{yx}), \mu_{15} = Cov(\bar{x}, \hat{S}_x^2),\ \mu_{24} = Cov(\bar{x}', \hat{S}_{yx}),$$

$$\mu_{25} = Cov(\bar{x}', \hat{S}_x^2),\ \mu_{34} = Cov(\bar{z}', \hat{S}_{yx}),\ \mu_{35} = Cov(\bar{z}', \hat{S}_x^2),$$

$$f' = 1 - \frac{n'}{N},\ C_x = \frac{S_x}{\bar{X}},\ C_z = \frac{S_z}{\bar{Z}};$$

$$S_y^2 = \frac{1}{N-1}\sum_{j=1}^{N}(Y_j - \bar{Y})^2$$

$$S_z^2 = \frac{1}{N-1}\sum_{j=1}^{N}(Z_j - \bar{Z})^2$$

$$S_x^2 = \frac{1}{N-1}\sum_{j=1}^{N}(X_j - \bar{X})^2$$

According to Singh (2003)

Let us define

$$\in_0 = \frac{\bar{y}}{\bar{Y}} - 1,\ \in_1 = \frac{\bar{x}}{\bar{X}} - 1,\ \in_2 = \frac{\bar{x}'}{\bar{X}} - 1,\ \in_3 = \frac{\bar{z}}{\bar{Z}} - 1 \text{ and } \in_4 = \frac{\bar{z}'}{\bar{Z}} - 1$$

such that

$$E(\in_j) = 0,\ j = 0,1,2,3,4.$$

$$E(\in_0^2) = \left(\frac{1}{n} - \frac{1}{N}\right)C_y^2,\ E(\in_1^2) = \left(\frac{1}{n} - \frac{1}{N}\right)C_x^2,\ E(\in_2^2) = \left(\frac{1}{m} - \frac{1}{N}\right)C_x^2,\ E(\in_3^2) = \left(\frac{1}{n} - \frac{1}{N}\right)C_z^2,$$

$$E(\in_4^2) = \left(\frac{1}{m} - \frac{1}{N}\right)C_z^2,\ E(\in_0\in_1) = \left(\frac{1}{n} - \frac{1}{N}\right)\rho_{xy}C_xC_y,\ E(\in_0\in_2) = \left(\frac{1}{m} - \frac{1}{N}\right)\rho_{xy}C_xC_y,$$

$$E(\in_0\in_3) = \left(\frac{1}{n} - \frac{1}{N}\right)\rho_{yz}C_yC_z,\ E(\in_0\in_4) = \left(\frac{1}{m} - \frac{1}{N}\right)\rho_{yz}C_yC_z,\ E(\in_1\in_2) = \left(\frac{1}{m} - \frac{1}{N}\right)C_x^2$$

$$E(\in_1\in_3) = \left(\frac{1}{n} - \frac{1}{N}\right)\rho_{xz}C_xC_z,\ E(\in_1\in_4) = \left(\frac{1}{m} - \frac{1}{N}\right)\rho_{xz}C_xC_z,\ E(\in_2\in_3) = \left(\frac{1}{m} - \frac{1}{N}\right)\rho_{xz}C_xC_z$$

$$E(\in_2\in_4) = \left(\frac{1}{m} - \frac{1}{N}\right)\rho_{xz}C_xC_z, \text{ and } E(\in_3\in_4) = \left(\frac{1}{m} - \frac{1}{N}\right)C_z^2$$

Using the large sample approximation, the mean square error MSE and Bias were derived as follows

$$T = \bar{Y}(1+\in_0) + b_{yx}\left[(1+\in_2)(1+\in_4)^{-\alpha} - (1+\in_1)\right]$$

$$\approx \bar{Y}(1+\in_0) + \beta\bar{X}\left[(1+\in_2)\left(1 - \alpha\in_4 + \frac{\alpha(\alpha+1)}{2}\in_4^2\right) - (1+\in_1)\right]$$

$$\approx \bar{Y}(1+\in_0) + \beta\bar{X}\left[1 - \alpha\in_4 + \frac{\alpha(\alpha+1)}{2}\in_4^2 + \in_2 - \alpha\in_2\in_4 - 1 - \in_1\right] \quad (15)$$

To obtain the biasness, subtract $\bar{Y}$ from both sides of equation 15 and take expectation

$$(T - \bar{Y}) = \bar{Y}\in_0 + \beta\bar{X}\left[1 - \alpha\in_4 + \frac{\alpha(\alpha+1)}{2}\in_4^2 + \in_2 - \alpha\in_2\in_4 - 1 - \in_1\right]$$

$$= \bar{Y}\in_0 + \beta\bar{X}\left[\in_2 - \alpha\in_4 - \in_1 - \alpha\in_2\in_4 + \frac{\alpha(\alpha+1)}{2}\in_4^2\right] \quad (16)$$

$$Bias(T) = E(T - \bar{Y}) = \beta\bar{X}\left[-\alpha\left(\frac{1}{m} - \frac{1}{N}\right)\rho_{xz}C_xC_z + \frac{\alpha(\alpha+1)}{2}\left(\frac{1}{m} - \frac{1}{N}\right)C_z^2\right] \quad (17)$$

Obtaining the MSE, square both side of equation 16 and retaining terms up to the second powers of $\in$'s

$$(T-\bar{Y})^2 = \bar{Y}^2 \in_0^2 + 2\beta\bar{X}\bar{Y}\left[\in_0\in_2 - \alpha \in_0\in_4 - \in_0\in_1\right] +$$
$$\bar{X}^2\beta^2\left[\in_2^2 - 2\alpha \in_2\in_4 - 2\in_1\in_2 + \alpha^2 \in_4^2 + 2\alpha \in_1\in_4 + \in_1^2\right]$$

$$MSE(T) = E(T-\bar{Y})^2 = \bar{Y}^2\left(\frac{1}{n}-\frac{1}{N}\right)C_y^2 + 2\beta\bar{X}\bar{Y}\left[\begin{array}{l}\left(\frac{1}{m}-\frac{1}{N}\right)\rho_{xy}C_xC_y - \alpha\left(\frac{1}{m}-\frac{1}{N}\right)\rho_{yz}C_yC_z \\ -\left(\frac{1}{n}-\frac{1}{N}\right)\rho_{xy}C_xC_y\end{array}\right] +$$

$$\bar{X}^2\beta^2\left[\begin{array}{l}\left(\frac{1}{m}-\frac{1}{N}\right)C_x^2 - 2\alpha\left(\frac{1}{m}-\frac{1}{N}\right)\rho_{xz}C_xC_z - 2\left(\frac{1}{m}-\frac{1}{N}\right)C_x^2 + \alpha^2\left(\frac{1}{m}-\frac{1}{N}\right)C_z^2 + \\ 2\alpha\left(\frac{1}{m}-\frac{1}{N}\right)\rho_{xz}C_xC_z + \left(\frac{1}{n}-\frac{1}{N}\right)C_x^2\end{array}\right]$$

$$MSE(T) = E(T-\bar{Y})^2 = \bar{Y}^2\left(\frac{1}{n}-\frac{1}{N}\right)C_y^2 + 2\beta\bar{X}\bar{Y}\left[\left(\frac{1}{m}-\frac{1}{n}\right)\rho_{xy}C_xC_y - \alpha\left(\frac{1}{m}-\frac{1}{N}\right)\rho_{yz}C_yC_z\right] +$$
$$\beta^2\bar{X}^2\left[\left(\frac{1}{n}-\frac{1}{m}\right)C_x^2 + \alpha^2\left(\frac{1}{m}-\frac{1}{N}\right)C_z^2\right]$$

(18)

But $\beta = \dfrac{\rho_{yx}S_x}{S_y}$

Then equation 18 can be re-written as

$$E(T-\bar{Y})^2 = \bar{Y}^2\left(\frac{1}{n}-\frac{1}{N}\right)C_y^2 - \bar{Y}^2\left(\frac{1}{n}-\frac{1}{m}\right)\rho_{xy}^2C_y^2 + \left(\frac{1}{m}-\frac{1}{n}\right)$$
$$+\bar{Y}^2\left[\frac{\alpha^2\rho_{xy}^2C_y^2C_z^2}{C_x^2} - \frac{2\alpha\rho_{xy}\rho_{yz}C_y^2C_z}{C_x}\right]$$

(19)

At various transformation of α say

$\alpha = 1$, the MSE becomes a chain ratio in regression estimator

$\alpha = -1$, the MSE becomes a product in regression estimator

$\alpha = 0$, the MSE becomes a regression estimator

Proposed Estimator

Based on the notations and derivation of Khare et al,

Proposing that if $\alpha = \frac{1}{2}$, we have a ratio type square root transformation

If $\alpha = 2$, we have a ratio type square transformation

At the various transformation of α, certain conditions about the population must be met.

For example if $\alpha = 1$, the relationship between y and z must be highly positively correlated

if $\alpha = -1$, the relationship between y and z must be linear and negatively correlated

if $\alpha = \frac{1}{2}$ or 2, the population must be skewed.

Testing for various value of α in the population, it will take time and these condition about the population may be hypothetical

Prasad (1989) proposed a ratio in simple random sampling

$$\bar{y}_p = k\bar{y}_r = k\frac{\bar{y}}{\bar{x}}\bar{X} \qquad (20)$$

Also Kadilar and Cingi (2005) in stratified random sampling using the same constant, suggested $\bar{y}_{stp} = K^*\bar{y}_{RC}$ (21)

The value of constant K will not vary

Motivated by these, this work introduced the constant to Khare et al (2013) estimator. Now the proposed estimator at $\alpha = 1$ is $T_P = K\left[\bar{y} + b_{yx}\left[\bar{x}'\left(\frac{\bar{Z}}{\bar{z}'}\right)^{\alpha} - \bar{x}\right]\right]$

(22)

$$T_P = K(T_{l3}) \qquad (23)$$

$$= K\left[\bar{y} + b_{yx}\left[\bar{x}'\left(\frac{\bar{Z}}{\bar{z}'}\right)^{\alpha} - \bar{x}\right]\right] \qquad (24)$$

$$= K\left[\bar{Y}(1+\in_0) + \beta\bar{X}\left[\in_2 - \alpha \in_4 - \in_1 - \alpha \in_2\in_4 + \frac{\alpha(\alpha+1)}{2}\in_4^2\right]\right]$$

$$T_P - Y \approx K\left[\bar{Y}(1+\in_0) + \beta\bar{X}\left[\in_2 - \alpha \in_4 - \in_1 - \alpha \in_2\in_4 + \frac{\alpha(\alpha+1)}{2}\in_4^2\right]\right] - Y$$

(25)

$$\approx K\bar{Y}\in_0 + \beta\bar{X}K\left[\in_2 - \alpha \in_4 - \in_1 - \alpha \in_2\in_4 + \frac{\alpha(\alpha+1)}{2}\in_4^2\right] + \bar{Y}(K-1)$$

$$E(T_P - \bar{Y}) = K\left[Bias(T_{l3})\right] + \bar{Y}(K-1)$$

(26)

$$(T_P-\bar{Y})^2 = K^2\bar{Y}^2\epsilon_0^2 + 2\beta\bar{Y}\bar{X}K^2[\epsilon_0\epsilon_2+\alpha\epsilon_0\epsilon_4-\epsilon_0\epsilon_1]+2\bar{Y}^2\epsilon_0 K(K-1)+$$
$$\beta^2\bar{X}^2K^2\left(\epsilon_2^2-2\alpha\epsilon_2\epsilon_4-2\alpha\epsilon_2\epsilon_4-2\epsilon_1\epsilon_2+\alpha^2\epsilon_4^2+2\alpha\epsilon_1\epsilon_4+\epsilon_1^2\right)+ \quad (27)$$
$$2\beta\bar{X}\bar{Y}K(K-1)\left[\epsilon_2+\alpha\epsilon_4-\epsilon_1-\alpha\epsilon_2\epsilon_4+\frac{\alpha(\alpha+1)}{2}\epsilon_4^2\right]+\bar{Y}^2(K-1)^2$$

$$MSE(T_P)=E(T_P-\bar{Y})^2=K^2MSE(T_{l3})+2K(K-1)\bar{Y}\left(Bias(T_{l3})\right)+\bar{Y}^2(K-1)^2 \quad (28)$$

The derivative of MSE was taken with respect to K and equate to zero to find the optimum value of K

$$\frac{\partial MSE(T_P)}{\partial K}=2KMSE(T_{l3})+4K\bar{Y}(BiasT_{l3})-2\bar{Y}(BiasT_{l3})+2\bar{Y}^2(K-1)=0$$

(29)

$$K=\frac{\bar{Y}Bias(T_{l3})+\bar{Y}^2}{MSE(T_{l3})+2\bar{Y}Bias(T_{l3})+\bar{Y}^2} \quad (30)$$

Where $0<K<1$

Mean square error of the proposed estimator

$$E(T_P-\bar{Y})^2=K^2\left[\begin{array}{l}\bar{Y}^2\left(\frac{1}{n}-\frac{1}{N}\right)C_y^2-\bar{Y}^2\left(\frac{1}{n}-\frac{1}{m}\right)\rho_{xy}^2C_y^2+\left(\frac{1}{m}-\frac{1}{n}\right)\bar{Y}^2+\\ \left[\frac{\alpha^2\rho_{xy}^2C_y^2C_z^2}{C_x^2}-\frac{2\alpha\rho_{xy}\rho_{yz}C_y^2C_z}{C_x}\right]\end{array}\right]+$$
$$2K(K-1)\bar{Y}\left(\beta\bar{X}\left[-\alpha\left(\frac{1}{m}-\frac{1}{N}\right)\rho_{xz}C_xC_z+\frac{\alpha(\alpha+1)}{2}\left(\frac{1}{m}-\frac{1}{N}\right)C_z^2\right]\right)+\bar{Y}^2(K-1)^2 \quad (31)$$

NUMERICAL ILLUSTRATON

Summary of Parameters

Table 1: Descriptive statistics on tobacco production

Parameters	Y	X	Z
N	106	106	106
m	80	80	80
n	50	50	50
Mean (1st phase)	1.5386	19743.78	552239.60
Mean (2nd Phase)	1.5462	19948.94	20757.1
Population mean	1.551038	22169.73	50184.13
Standard deviation	0.7950129	57916.08	253183.6
Covariance	0.5125684	2.612395	5.045093

Corr(Y,X) = -0.1777226

Corr(Y,Z) = -0.1435275

Corr(X,Z) = 0.9688092

$\beta(Y,X)=-2.824$

The table above shows the parameters of the first data set used in the double sampling analysis, the mean of the first and second phase for each of the characters used in the analysis. The correlation between each character shows that it is poorly negatively correlated aside from the character Z and X which is highly positively correlated.

Table 4: MSE of the existing and proposed estimator for data on tobacco production

α	Existing Khare et al. (2013)	Proposed
0.0	0.00803	0.00540
0.4	0.00794	
0.8	0.00802	
1.2	0.00828	
1.6	0.00872	
2.0	0.00933	
Bias	2435.242945	2035.245

The table above shows the MSE of the proposed and existing estimators. As explained earlier the existing mean square error was derived using iterative procedures thereby resulting in different values of 'alpha', these values were used to obtain the range of MSEs, it was also plotted in figure 10. The proposed method introduced another constant which will not use iterative procedure and obtained a minimum MSE which is more minimal than the minimum of the existing estimator. The minimum MSE is 0.00794, which was obtained for the existing estimator when the value of alpha is 0.4. The proposed MSE is 0.00540. It can be seen that the proposed MSE is more minimum, thereby the proposed estimator is more efficient.

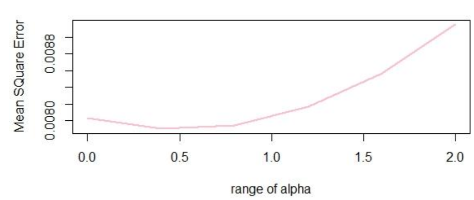

Figure 10: plot on range of 'alpha' for existing estimator on tobacco data

Table 5: Descriptive statistics on number of pupils in primary six, five and one

Parameters	Y	X	Z
N	116	116	116
m	80	80	80
n	50	50	50
Mean (1st phase)	88.725	77.6625	70.05
Mean (2nd phase)	75.9	69.4	65.18
Population mean	88.09483	76.02586	69.57759
Standard deviation	68.653	54.29812	39.09212

Corr(Y,Z) = 0.6127134
Corr(X,Z) = 0.7205506
$\beta(Y, X) = 1.074983$

Table 6: MSE of the existing and proposed estimator for data on number of pupils

α	Existing Khare et al. (2013)	Proposed
0.0	20.732	6.523
0.4	11.077	
0.8	7.487	
1.2	9.961	
1.6	18.499	
2.0	33.103	
Bias	0.99903	0.7738

The minimum MSE is 7.487, which was obtained for the existing when the value of alpha is 0.8. The proposed MSE is 6.523. It can be seen that the proposed MSE is more minimum, thereby the proposed estimator is more efficient. The values of alpha were also plotted for the MSE.

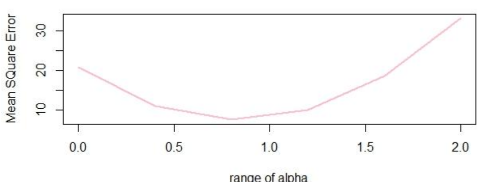

Figure 11: plot on range of 'alpha' for existing estimator on number of pupils

CONCLUSION

A generalized chain ratio in regression estimator for population mean has been proposed and its properties have been studied. A comparative study of the proposed estimator has been made with relevant estimators. From the theoretical discussion and numerical examples, we infer that the proposed estimator is more efficient than the existing estimator by Khare *et al.* (2013), as the mean squared errors of the proposed is minimal than the mean squared errors of the existing estimator by Khare *et al.* (2013). Also the constant "K" that was introduced eliminated the iterative procedures used to obtain "alpha" which was a constant introduced by Khare et al (2013). The efficiency of the proposed procedure was also compared with the existing works, and the condition which must always be met was given. The proposed estimator since does not require iterative procedures saves time.

REFERENCES

Chand, L. (1975): Some ratio-type estimators based on two or more auxiliary variables.

Cochran, W.G. 1977. Sampling Techniques (3rd Edition), New York, Wiley.

Holt, D. and Smith, T.M.F. 1979. Post-stratification, *J.R. Stat. Soc., A*, 142, 33-36.

Khare B.B, Srvastava U, Kumar K. 2013. A generalized chain ratio in regression estimator for population mean using two auxiliary characters in sample survey. *J Sci Res* 57: 147–153.

Kiregyera, B. 1980). A chain ratio type estimator infinite population double sampling. Unpublished Ph.D. dissertation, Iowa State University, Ames, Iowa.

Kiregyera, B. (1984). Regression type estimators using two auxiliary variables and the model of double sampling from finite population. 31:215{226}.

Srivastava, S. Rani, Khare, B.B. and Srivastava, S.R. (1990). A generalized chain ratio estimator for mean of finite population. *Jour. Ind. Soc. Agri. Statist*., 42(1): 108-117.