*Proceeding Paper*

# Modified Estimator of Finite Population Variance under Stratified Random Sampling

**Yahaya Zakari[1]* and Isah Muhammad[2]**

[1]Department of Statistics, Ahmadu Bello University, Zaria, Nigeria.
[2]Department of Statistics, Binyaminu Usman Polytechnic, Hadejia, Nigeria.
* Correspondence: Author: Email: yzakari@abu.edu.ng

**Abstract:** This paper proposes a generalized estimator of finite population variance using the auxiliary information under stratified random sampling. The expressions for bias and mean square error equations of the proposed estimator are derived up to the first degree of approximation. The theoretical efficiency conditions under which the proposed estimator is better than some existing estimators are obtained. The performances of the existing and proposed estimators were assessed using three real datasets based on the criteria of minimum mean square error and supreme percentage relative efficiency. Evidence from the study showed that the proposed estimator performed better and was more efficient than some existing estimators considered.

***Keywords:*** Auxiliary variable; Mean square error; Bias; Efficiency

## 1. Introduction

Researchers have been considering the use of auxiliary information in various forms to construct more accurate estimators for population parameters in experimental surveys. This approach has been shown to significantly improve the accuracy of the population mean estimation, and it has been explored by several researchers, including [1], [2] and [3]. The field of survey sampling has seen a significant amount of research devoted to developing estimators of population variance to measure variations that exist in real-world scenarios. These variations can arise in various industries, such as manufacturing, pharmaceuticals, agriculture, and biological experiments. Researchers like [4], [5], [6], [7], [8], [9], [10], [11], [12], [13] and [14] have worked extensively in this area. In real-life situations, population units of variables of interest can have diverse features, which makes the population units heterogeneous. In such cases, estimators of finite population variance, based on simple random sampling, cannot be applied. Therefore, the stratification approach is used to provide precise results when population units are heterogeneous ([5]; [15]; [16]; [17]).

Most ratio-based and product-based estimators are not efficient in estimating population characteristics when there is a negative correlation between the study and auxiliary variables for ratio-based estimators or a positive correlation for product-based estimators. To address this issue under stratified random sampling, a new generalized ratio-product cum regression-type estimator is proposed. The generalized ratio-product cum regression-type estimator proposed in this study is flexible in providing better efficient estimates than existing estimators when the data exhibits either a negative or positive correlation between the study and auxiliary variables.

Following the introduction is section 2 which contains the notations and literature review while section 3 presents the methodology of the study. Section 4 presents the efficiency conditions of the proposed methodology. Section 5 discusses the results while the conclusion and recommendations are presented in Section 6.

## 2. Notations and Some Existing Estimators

Consider a finite population $U = \{U_1, , , , , U_N\}$ of size $N$ which is divided into $L$ non-overlapping group or strata, with each stratum containing $N_h (h = 1,2, \dots, L)$ units, whose union gives N units, i.e. $\sum_{h=1}^{G} N_h = N$. A sample using simple random sample of size $n_h$ is drawn without replacement *(SRSWOR)* from the strata population $N_h$ such that $\sum_{h=1}^{G} n_h = n$. Let $(y_{hi}, x_{hi})$ be the value of the study variable Y and the auxiliary variable X on $ith$ unit $U_i, i = 1, \dots N$. Let $\overline{Y}_h$, and $\overline{X}_h$ be population means corresponding to the sample means $\overline{y}_h$ and $\overline{x}_h$, respectively, in each stratum. Let $s_{y(h)}^2$ and $s_{x(h)}^2$ be the sample mean squares for the study and auxiliary variables for the $hth$ strata, respectively, $S_{y(h)}^2$ and $S_{x(h)}^2$ be the population mean squares for the study and auxiliary variables for the $hth$ strata, respectively. $W_h = \frac{N_h}{N}$ be the proportion of the population units falling in the $hth$ stratum. The correlation coefficient between the study variable and the auxiliary variable for the $hth$ strata be $\rho_{yx(h)}$. Also, let $C_{y(h)} = S_{y(h)}/\overline{Y}_h$ and $C_{x(h)} = S_{x(h)}/\overline{X}_h$ be the stratum coefficients of variation of the study variable Y and the auxiliary variable X. $\beta_{2(y)} = \psi_{40}$ and $\beta_{2(x)} = \psi_{04}$ be the coefficient of kurtosis of y and x for the $hth$ strata, respectively.

$\beta_{2(y)} = \psi_{40(h)} = \frac{\mu_{40(h)}}{\mu_{20(h)}^2}$, $\beta_{2(x)} = \psi_{04(h)} = \frac{\mu_{04(h)}}{\mu_{02(h)}^2}$, $\psi_{22(h)} = \frac{\mu_{22(h)}}{\mu_{20(h)}\mu_{02(h)}}$, $\overline{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} Y_{hi}$,

$\overline{X}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} X_{hi}$, $\overline{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$, $\overline{x}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$, $s_{yh}^2 = \frac{1}{(n_h-1)}\sum_{i=1}^{n_h}\left(y_{hi} - \overline{y}_h\right)^2$, $s_{xh}^2 = \frac{1}{(n_h-1)}\sum_{i=1}^{n_h}(x_{hi} - \overline{x}_h)^2$ , $S_{Y_h}^2 = \frac{1}{(N_h-1)}\sum_{i=1}^{N_h}\left(Y_{hi} - \overline{Y}_h\right)^2$ , $S_{X_h}^2 = \frac{1}{(N_h-1)}\sum_{i=1}^{N_h}\left(X_{hi} - \overline{X}_h\right)^2$, $\gamma_h = \frac{1}{n_h} - \frac{1}{N_h}$, $\mu_{rs(h)} = \frac{1}{N_h}\sum_{j=1}^{N_h}\left(Y_{hj} - \overline{Y}_h\right)^r\left(X_{hj} - \overline{X}_h\right)^s$.

The usual unbiased population variance estimator under stratified random sampling can be used to estimate the population variance and is given as:

$$s_{y_{st}}^2 = \sum_{h=1}^{L} W_h^2 \gamma_h s_{y_h}^2 \tag{1}$$

$$Var\left(s_{y_{st}}^2\right) = \sum_{h=1}^{L} W_h^4 \gamma_h^3 S_{y_h}^4 \left(\psi_{40(h)} - 1\right) \tag{2}$$

The usual ratio stratified estimator for estimating population variance proposed by [5] is given as:

$$T_{r(st)} = \sum_{h=1}^{G} W_h^2 \gamma_h \left( s_{y_h}^2 \frac{S_{X_h}^2}{s_{x_h}^2} \right) \tag{3}$$

$$Bias(T_{r(st)}) = \sum_{h=1}^{G} W_h^2 \gamma_h^2 S_{y_h}^2 \left[\left(\psi_{40(h)} - 1\right) - \left(\theta_{22(h)} - 1\right)\right] \tag{4}$$

$$MSE(T_{r(st)}) \cong \sum_{h=1}^{G} W_h^4 \gamma_h^3 S_{y_h}^4 \left[\left(\psi_{40(h)} - 1\right) + \left(\psi_{04(h)} - 1\right) - 2\left(\theta_{22(h)} - 1\right)\right] \tag{5}$$

The usual regression stratified estimator for estimating population variance proposed by [5] is given as:

$$S_{reg(st)}^2 = \sum_{h=1}^{L} W_h^2 \gamma_h \left[ s_{y(h)}^2 + b_h \left( S_{x(h)}^2 - s_{x(h)}^2 \right) \right] \tag{6}$$

$$Var(S_{reg(st)}^2) = \sum_{h=1}^{L} W_h^4 \gamma_h^3 S_{y_h}^4 \left(\psi_{40(h)} - 1\right)\left(1 - \rho_h^2\right) \tag{7}$$

Also, [15] proposed a difference-type estimator of population variance under stratified random sampling as follows:

$$S_{D(st)}^2 = \sum_{h=1}^{G} W_h^2 \gamma_h \left[ k_{1h} s_{yh}^2 + k_{2h} \left( S_{xh}^2 - s_{xh}^2 \right) \right] \tag{8}$$

$$Bias\left(S^2_{D(st)}\right)_{\min} \cong \sum_{h=1}^{G} W_h^2 \gamma_h^2 S_{y_h}^2 \left\{ \frac{\left(\psi_{04(h)}-1\right)}{\gamma_h\left[\left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)^2\right]} + \left(\psi_{04(h)}-1\right)-1 \right\} \quad (9)$$

$$MSE\left(S^2_{D(st)}\right)_{\min} \cong \sum_{h=1}^{G} W_h^4 \gamma_h^3 S_{y_h}^4 \left\{ \left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right) - \frac{\left(\theta_{22(h)}-1\right)^2}{\gamma_h\left[\left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)^2\right]} + \left(\psi_{04(h)}-1\right) \right\} \quad (10)$$

where

$$k_{1h}^{(opt)} = \frac{\left(\psi_{04(h)}-1\right)}{\gamma_h\left[\left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)^2\right]+\left(\psi_{04(h)}-1\right)} \quad (11)$$

$$k_{2h}^{(opt)} = \frac{S_{y_h}^2\left(\theta_{22(h)}-1\right)}{S_{x_h}^2\left\{\gamma_h\left[\left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)^2\right]+\left(\psi_{04(h)}-1\right)\right\}} \quad (12)$$

Ratio-type and product-type estimators of finite population variance under stratified random sampling proposed by [17] is as follows:

$$S^2_{SR(st)} = \sum_{h=1}^{L} W_h^2 \gamma_h s_{y_h}^2 \exp\left(\frac{S_{x_h}^2 - s_{x_h}^2}{S_{x_h}^2 + s_{x_h}^2}\right) \quad (13)$$

$$S^2_{SP(st)} = \sum_{h=1}^{L} W_h^2 \gamma_h s_{y_h}^2 \exp\left(\frac{s_{x_h}^2 - S_{x_h}^2}{S_{x_h}^2 + s_{x_h}^2}\right) \quad (14)$$

$$Bias\left(S^2_{SR(st)}\right) = \sum_{h=1}^{L} W_h^2 \gamma_h^2 S_{y_h}^2 \left[\frac{3}{8}\left(\psi_{04(h)}-1\right)-\frac{1}{2}\left(\theta_{22(h)}-1\right)\right] \quad (15)$$

$$Bias\left(S^2_{SP(st)}\right) = \sum_{h=1}^{L} W_h^2 \gamma_h^2 S_{y_h}^2 \left[\frac{1}{2}\left(\theta_{22(h)}-1\right)-\frac{1}{8}\left(\psi_{04(h)}-1\right)\right] \quad (16)$$

$$MSE\left(S^2_{SR(st)}\right) = \sum_{h=1}^{L} W_h^4 \gamma_h^3 S_{y_h}^4 \left[\left(\psi_{40(h)}-1\right)+\frac{1}{4}\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)\right] \quad (17)$$

$$MSE\left(S^2_{SP(st)}\right) = \sum_{h=1}^{L} W_h^4 \gamma_h^3 S_{y_h}^4 \left[\left(\psi_{40(h)}-1\right)+\frac{1}{4}\left(\psi_{04(h)}-1\right)+\left(\theta_{22(h)}-1\right)\right] \quad (18)$$

An improved population variance estimator under stratified random sampling proposed by [16] is given as follows:

$$S^2_{SG(st)} = \sum_{h=1}^{G} W_h^2 \gamma_h \left[\omega_{1h} s_{yh}^2 + \omega_{2h}\left(S_{xh}^2 - s_{xh}^2\right)\right] \exp\left(\frac{S_{x_h}^2 - s_{x_h}^2}{S_{x_h}^2 + s_{x_h}^2}\right) \quad (19)$$

$$Bias\left(S^2_{SG(st)}\right) \cong \sum_{h=1}^{G} W_h^2 \gamma_h^2 \begin{bmatrix} S_{y_h}^2 + \omega_{1h} S_{y_h}^2 \left\{1+\frac{3}{8}\left(\psi_{40(h)}-1\right)-\frac{1}{2}\left(\theta_{22(h)}-1\right)\right\} \\ +\frac{1}{2}\omega_{2h} S_{x_h}^2 \gamma_h \left(\psi_{04(h)}-1\right) \end{bmatrix} \quad (20)$$

$$MSE\left(S^2_{SG(st)}\right)_{\min} \cong \sum_{h=1}^{G} W_h^4 \gamma_h^3 S_{y_h}^4 \left\{ \frac{\begin{array}{c} 64\left[\left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)^2\right]-\gamma_h\left(\psi_{04(h)}-1\right)^3 \\ -16\gamma_h\left(\psi_{04(h)}-1\right)\left[\left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)^2\right] \end{array}}{64\left\{\gamma_h\left[\left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)^2\right]+\left(\psi_{40(h)}-1\right)\right\}} + \left(\psi_{04(h)}-1\right) \right\} \quad (21)$$

where $\omega_{1h}$ and $\omega_{2h}$ are suitable chosen constants having optimum values

$$\omega_{1h}^{(opt)} = \frac{\gamma_h\left(\psi_{04(h)}-1\right)}{8}\left[\frac{8-\gamma_h\left(\psi_{04(h)}-1\right)}{\gamma_h\left[\left(\psi_{04(h)}-1\right)+\left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)^2\right]}\right] \quad (22)$$

$$\omega_{2h}^{(opt)} = \frac{S_{y_h}^2}{8S_{x_h}^2}\left[\frac{\left(\psi_{40(h)}-1\right)\left(4\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)^2-3\right)+4\left(\theta_{22(h)}-1\right)^2}{\left[\left(\psi_{04(h)}-1\right)+\left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)^2\right]}\right] \quad (23)$$

**3. The Proposed Estimator**

Motivated by [14] estimators and having studied some estimators under stratified random sampling, the following generalized estimator is proposed to estimate the population variance of the variable of interest by utilizing the auxiliary information.

$$G_{(st)} = \sum_{h=1}^{L} W_h^2 \lambda_h \left\{ \tau_{1(h)} s_{y(h)}^2 \left[\frac{1}{2}\left(\frac{S_{x(h)}^2}{s_{x(h)}^2}+\frac{s_{x(h)}^2}{S_{x(h)}^2}\right)\right] + \tau_{2(h)}\left(S_{x(h)}^2 - s_{x(h)}^2\right)\right\} \exp\left\{\frac{S_{x(h)}^2 - s_{x(h)}^2}{S_{x(h)}^2 + s_{x(h)}^2}\right\} \quad (24)$$

where $\tau_{i(h)}$ $(i = 1,2)$ are unknown constants whose values to be determined later, and $\sigma$ is suitably chosen constant.

In order to find the bias and *MSE* of the estimators, we define the following relative error terms:

$$e_{0(h)} = \frac{s_{y(h)}^2 - S_{y(h)}^2}{S_{y(h)}^2} \text{ and } e_{1(h)} = \frac{s_{x(h)}^2 - S_{x(h)}^2}{S_{x(h)}^2} \text{ where } s_{y(h)}^2 = S_{y(h)}^2\left(1+e_{0(h)}\right) \text{ and } s_{x(h)}^2 = S_{x(h)}^2\left(1+e_{1(h)}\right)$$

such that

$$E\left(e_{0(h)}\right) = E\left(e_{1(h)}\right) = 0,\ E\left(e_{0(h)}^2\right) = \gamma_h\left(\psi_{40(h)}-1\right),\ E\left(e_{1(h)}^2\right) = \gamma_h\left(\psi_{04(h)}-1\right) \text{ and } E\left(e_{0(h)}e_{1(h)}\right) = \gamma_h\left(\psi_{22(h)}-1\right).$$

Expressing the estimator $G_{i(st)}$ in terms of $e_i$ $(i = 0,\ 1)$ and by taking the expectation, the bias of the proposed generalized estimator, up to the first order of approximation, is given as:

$$Bias\left(G_{(st)}\right) = \sum_{h=1}^{G} \frac{W_h^2}{\lambda_h^{-1}}\left[\begin{array}{l} \left(\tau_{1(h)}-1\right)S_{y(h)}^2 + \tau_{1(h)}S_{y(h)}^2\left(\frac{7}{8}\lambda_h\left(\psi_{04(h)}-1\right)-\frac{1}{2}\lambda_h\left(\psi_{22(h)}-1\right)\right) \\ +\frac{\tau_{2(h)}}{2}S_{x(h)}^2\lambda_h\left(\psi_{04(h)}-1\right) \end{array}\right] \quad (25)$$

The mean square error of the proposed estimator, up to the first order of approximation, is given as:

$$MSE\left(G_{(st)}\right)=\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}\left\{\begin{array}{l}\left(\tau_{1(h)}-1\right)^2 S_{y(h)}^4+\tau_{1(h)}^2 S_{y(h)}^4\lambda_h\left[\left(\psi_{40(h)}-1\right)-2\left(\psi_{22(h)}-1\right)+2\left(\psi_{04(h)}-1\right)\right]\\+\tau_{2(h)}^2 S_{x(h)}^4\lambda_h\left(\psi_{04(h)}-1\right)-\tau_{1(h)}S_{y(h)}^4\lambda_h\left(\frac{7}{4}\left(\psi_{04(h)}-1\right)-\left(\psi_{22(h)}-1\right)\right)\\-2\tau_{1(h)}\tau_{2(h)}S_{y(h)}^2 S_{x(h)}^2\lambda_h\left(\left(\psi_{22(h)}-1\right)-\left(\psi_{04(h)}-1\right)\right)-\tau_{2(h)}S_{y(h)}^2 S_{x(h)}^2\lambda_h\left(\psi_{04(h)}-1\right)\end{array}\right\} \quad (26)$$

Minimizing equation (26) with respect to $\tau_{1(h)}$ and $\tau_{2(h)}$, we obtained the optimum values of $\tau_{1(h)}$ and $\tau_{2(h)}$ as:

$$\tau_{1(h)}^{(opt)}=\frac{1+\frac{3}{8}\lambda_h\left(\psi_{04(h)}-1\right)}{1+\lambda_h\left(\psi_{04(h)}-1\right)+\lambda_h\left(\left(\psi_{40(h)}-1\right)-\frac{\left(\psi_{22(h)}-1\right)^2}{\left(\psi_{04(h)}-1\right)}\right)} \quad (27)$$

and

$$\tau_{2(h)}^{(opt)}=\frac{S_{y(h)}^2}{S_{x(h)}^2}\left\{\frac{1}{2}-\tau_{1(h)}^{(opt)}\left(1-\frac{\left(\psi_{22(h)}-1\right)}{\left(\psi_{04(h)}-1\right)}\right)\right\} \quad (28)$$

Substituting the optimum values of $\tau_{1(h)}$ and $\tau_{2(h)}$ into equation (26), we obtained the minimum mean square error of $G_{i(st)}$ as:

$$MSE\left(G_{(st)}\right)_{\min}=\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\left\{\left(1-\frac{\lambda_h\left(\psi_{04(h)}-1\right)}{4}\right)-\frac{\left[1+\frac{3}{8}\lambda_h\left(\psi_{04(h)}-1\right)\right]^2}{\left[1+\lambda_h\left(\psi_{04(h)}-1\right)+\lambda_h\left(\left(\psi_{40(h)}-1\right)-\frac{\left(\psi_{22(h)}-1\right)^2}{\left(\psi_{04(h)}-1\right)}\right)\right]}\right\} \quad (29)$$

**4. Efficiency Comparisons**

In this section, we obtain the theoretical efficiency conditions for the proposed estimator by comparing the mean square error (MSE) equation of the proposed estimator with the MSE equations of some existing estimators. Thus, conditions under which the proposed estimator is more efficient are given below:

- Comparing the proposed estimator's MSE with that of the usual variance estimator,
  we have:

$$Var(\hat{s}_{yst}^2)-MSE\left(G_{(st)}\right)_{\min}>0 \quad (30)$$

Substituting the variances of the usual variance estimator and proposed estimator into equation (30), it gives:

$$\sum_{h=1}^{L}\frac{W_h^4}{n_h^3}S_{y_h}^4\left(\psi_{40(h)}-1\right)-\sum_{h=1}^{G}\frac{W_h^4}{n_h^2}S_{y(h)}^4\left\{B_{(h)}-\frac{C_{(h)}^2}{D_{(h)}}\right\}>0 \quad (31)$$

Further simplifying equation (31), we have:

$$\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\frac{C_{(h)}^2}{D_{(h)}}>\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4 B_{(h)}-\sum_{h=1}^{L}\frac{W_h^4}{n_h^3}S_{y_h}^4\left(\psi_{40(h)}-1\right) \quad (32)$$

where $B_{(h)}=\left(1-\frac{\gamma_h\left(\psi_{04(h)}-1\right)}{4}\right)$, $C_{(h)}=1+\frac{3}{8}\lambda_h\left(\psi_{04(h)}-1\right)$ and $D_{(h)}=1+\lambda_h\left(\psi_{04(h)}-1\right)+\lambda_h\left(\left(\psi_{40(h)}-1\right)-\frac{\left(\theta_{22(h)}-1\right)^2}{\left(\psi_{04(h)}-1\right)}\right)$

Since condition (32) is satisfied, the proposed estimators are more efficient than sample variance.

- Comparing the proposed estimator's MSE with that of the usual ratio estimator defined by [5], we have:

$$MSE\left(T_{r(st)}\right)-MSE\left(G_{(st)}\right)_{\min}>0 \quad (33)$$

Substituting the MSE expressions of the usual ratio estimator and proposed estimator into equation (33), it gives:

$$\sum_{h=1}^{G}W_h^4\gamma_h^3S_{y_h}^4\left[\begin{array}{l}\left(\psi_{40(h)}-1\right)+\left(\psi_{04(h)}-1\right)\\-2\left(\theta_{22(h)}-1\right)\end{array}\right]-\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\left\{B_{(h)}-\frac{C_{(h)}^2}{D_{(h)}}\right\}>0 \quad (34)$$

Further simplifying equation (34), we have:

$$\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\frac{C_{(h)}^2}{D_{(h)}}>\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4B_{(h)}-\sum_{h=1}^{G}W_h^4\gamma_h^3S_{y_h}^4\left[\begin{array}{l}\left(\psi_{40(h)}-1\right)+\left(\psi_{04(h)}-1\right)\\-2\left(\theta_{22(h)}-1\right)\end{array}\right] \quad (35)$$

Since condition (35) is satisfied, the proposed estimators are more efficient than [5] stratified regression estimator.

- Comparing the proposed estimator's MSE with that of the usual stratified regression estimator, we have:

$$Var(S^2_{reg(st)})-MSE\left(G_{(st)}\right)_{\min}>0 \quad (36)$$

Substituting the MSE expressions of the usual ratio estimator and proposed estimator into equation (36), it gives:

$$\sum_{h=1}^{L}\frac{W_h^4}{n_h^3}S_{y_h}^4\left(\psi_{40(h)}-1\right)\left(1-\rho_h^2\right)-\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\left\{B_{(h)}-\frac{C_{(h)}^2}{D_{(h)}}\right\}>0 \quad (37)$$

Further simplifying equation (37), we have:

$$\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\frac{C_{(h)}^2}{D_{(h)}}>\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4B_{(h)}-\sum_{h=1}^{L}\frac{W_h^4S_{y_h}^4}{n_h^3}\left(\psi_{40(h)}-1\right)\left(1-\rho_h^2\right) \quad (38)$$

Since condition (38) is satisfied, the proposed estimators are more efficient than [5] stratified regression estimator.

- Comparing the proposed estimator's MSE with that of the [15] stratified estimator, we have:

$$MSE\left(S^2_{D(st)}\right)_{\min}-MSE\left(G_{i(st)}\right)_{\min}>0 \quad (39)$$

**Substituting** the MSE expressions of the [15] estimator and proposed estimator into equation (39); it gives:

$$\sum_{h=1}^{G}W_h^4\gamma_h^3S_{y_h}^4\left\{\begin{array}{l}\left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)+\left(\psi_{04(h)}-1\right)\\-\dfrac{\left(\theta_{22(h)}-1\right)^2}{\gamma_h\left[\left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)^2\right]}\end{array}\right\}-\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\left\{B_{(h)}-\frac{C_{(h)}^2}{D_{(h)}}\right\}>0 \quad (40)$$

Further simplifying equation (40), we have:

$$\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\frac{C_{(h)}^2}{D_{(h)}} > \sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4 B_{(h)} - \sum_{h=1}^{G}W_h^4\gamma_h^3 S_{y_h}^4 \left\{ \begin{array}{l} \left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)+\left(\psi_{04(h)}-1\right) \\ -\dfrac{\left(\theta_{22(h)}-1\right)^2}{\gamma_h\left[\left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h}\right.\right.} \end{array}\right.$$

Since condition (41) is satisfied, the proposed estimators are more efficient than [15] stratified estimator.

- Comparing the proposed estimator's MSE with that of the [17] ratio-type and product-type estimators, respectively we have:

$$MSE\left(S_{SR(st)}^2\right) - MSE\left(G_{(st)}\right)_{\min} > 0 \tag{42}$$

Substituting the MSE expressions of the [17] ratio-type estimator and proposed estimator into equation (42), it gives:

$$\sum_{h=1}^{L}W_h^4\gamma_h^3 S_{y_h}^4\left[\left(\psi_{40(h)}-1\right)+\frac{1}{4}\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)\right] - \sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\left\{B_{(h)}-\frac{C_{(h)}^2}{D_{(h)}}\right\} > 0 \tag{43}$$

Further simplifying equation (43), we have:

$$\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\frac{C_{(h)}^2}{D_{(h)}} > \sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4 B_{(h)} - \sum_{h=1}^{L}W_h^4\gamma_h^3 S_{y_h}^4\left[\begin{array}{l}\left(\psi_{40(h)}-1\right)+\frac{1}{4}\left(\psi_{04(h)}-1\right) \\ -\left(\theta_{22(h)}-1\right)\end{array}\right] \tag{44}$$

Similarly, for the [17] product-type estimator, we have:

$$MSE\left(S_{SP(st)}^2\right) - MSE\left(G_{(st)}\right)_{\min} > 0 \tag{45}$$

Substituting the MSE expressions of the [17] product-type estimator and proposed estimator into equation (45), it gives:

$$\sum_{h=1}^{L}W_h^4\gamma_h^3 S_{y_h}^4\left[\left(\psi_{40(h)}-1\right)+\frac{1}{4}\left(\psi_{04(h)}-1\right)+\left(\theta_{22(h)}-1\right)\right] - \sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\left\{B_{(h)}-\frac{C_{(h)}^2}{D_{(h)}}\right\} > 0 \tag{46}$$

Further simplifying equation (46), we have:

$$\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\frac{C_{(h)}^2}{D_{(h)}} > \sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4 B_{(h)} - \sum_{h=1}^{L}W_h^4\gamma_h^3 S_{y_h}^4\left[\begin{array}{l}\left(\psi_{40(h)}-1\right)+\frac{1}{4}\left(\psi_{04(h)}-1\right) \\ +\left(\theta_{22(h)}-1\right)\end{array}\right] \tag{47}$$

Since conditions (44) and (47) are satisfied, the proposed estimator is more efficient than [17] ratio-type and product-type stratified estimators, respectively.

- Comparing the proposed estimator's MSE with that of the [16] variance estimator, we have:

$$MSE\left(S_{SG(st)}^2\right) - MSE\left(G_{(st)}\right)_{\min} > 0 \tag{48}$$

Substituting the MSE expressions of the [16] estimator and proposed estimator into equation (48), it gives:

$$\sum_{h=1}^{G}W_h^4\gamma_h^3 S_{y_h}^4\left\{\frac{64K-\gamma_h\left(\psi_{04(h)}-1\right)^3-16\gamma_h\left(\psi_{04(h)}-1\right)K}{64\gamma_h K+\left(\psi_{40(h)}-1\right)}+\left(\psi_{04(h)}-1\right)\right\} - \sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\left\{B_{(h)}-\frac{C_{(h)}^2}{D_{(h)}}\right\} > 0 \tag{49}$$

Further simplifying equation (49), we have:

$$\sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4\frac{C_{(h)}^2}{D_{(h)}} > \sum_{h=1}^{G}\frac{W_h^4}{\lambda_h^{-2}}S_{y(h)}^4 B_{(h)} - \sum_{h=1}^{G}W_h^4\gamma_h^3 S_{y_h}^4\left\{\frac{\begin{array}{c}64K-\gamma_h\left(\psi_{04(h)}-1\right)^3\\-16\gamma_h\left(\psi_{04(h)}-1\right)K\end{array}}{64\gamma_h K+\left(\psi_{40(h)}-1\right)}+\left(\psi_{04(h)}-1\right)\right\} \quad (50)$$ where

$$K=\left[\left(\psi_{40(h)}-1\right)\left(\psi_{04(h)}-1\right)-\left(\theta_{22(h)}-1\right)^2\right]$$

Since condition (50) are satisfied, the proposed estimator is more efficient than [16] variance estimator.

## 5. Empirical Results

To observe the performance of our proposed estimator with respect to other considered estimators, we use the following data sets for numerical comparisons, which were earlier used by many authors in the literature.

**Table 1.** Source and Description of Data Sets.

| Data set | Source | y | X |
|---|---|---|---|
| 1 | [19] | Leaf area for the newly developed strain of wheat | Weight of leaves |
| 2 | [7] | Level of apple production | Rate of apple trees |
| 3 | [18] | Area under wheat in the region in 1974 | Area under wheat in the region in 1973 |

**Table 2.** Parameters of the Dataset 1.

| Parameters | $N_h$ | $n_h$ | $C_{xh}$ | $S_{y_h}^2$ | $S_{x_h}^2$ | $\beta_{2(y_h)}$ | $\beta_{2(x_h)}$ | $\lambda_{22(h)}$ |
|---|---|---|---|---|---|---|---|---|
| Stratum 1 | 12 | 4 | 0.1118928 | 6.0664603 | 11.5715804 | 1.9394547 | 2.2748233 | 1.9123464 |
| Stratum 2 | 13 | 5 | 0.0733753 | 5.2915229 | 8.139014 | 2.9819269 | 3.436904 | 2.970998 |
| Stratum 3 | 14 | 5 | 0.1193784 | 6.4961301 | 12.4494617 | 2.3448986 | 2.8955496 | 2.5134376 |

**Table 3.** Parameters of the Dataset 2.

| Parameters | $N_h$ | $n_h$ | $S_{y_h}^2$ | $S_{x_h}^2$ | $\beta_{2(y_h)}$ | $\beta_{2(x_h)}$ | $\lambda_{22(h)}$ |
|---|---|---|---|---|---|---|---|
| Stratum 1 | 19 | 10 | 583977.5 | 11.5715804 | 3.28 | 1.45 | 1.46 |
| Stratum 2 | 32 | 16 | 456563.3 | 8.139014 | 1.56 | 3.09 | 1.74 |
| Stratum 3 | 14 | 7 | 195208.8 | 12.4494617 | 1.62 | 1.62 | 1.80 |
| Stratum 4 | 15 | 8 | 437923.5 | 10.334267 | 2.22 | 1.90 | 2.02 |

**Table 4.** Parameters of the Dataset 3.

| Parameters | $N_h$ | $n_h$ | $C_{xh}$ | $S_{y_h}^2$ | $S_{x_h}^2$ | $\beta_{2(y_h)}$ | $\beta_{2(x_h)}$ | $\lambda_{22(h)}$ |
|---|---|---|---|---|---|---|---|---|
| Stratum 1 | 9 | 3 | 0.1118928 | 31978.25 | 11.5715804 | 2.9286 | 2.07 | 2.38 |
| Stratum 2 | 10 | 3 | 0.0733753 | 37629.39 | 8.139014 | 1.511 | 1.42 | 1.42 |
| Stratum 3 | 15 | 4 | 0.11937840 | 6893.067 | 12.4494617 | 2.20 | 2.42 | 2.28 |

**Table 5.** Bias Estimates of the Existing Estimators and the Proposed Estimator.

| | Dataset 1 | Dataset 2 | Dataset 3 |
|---|---|---|---|
| Sample Variance | 0 | 0 | 0 |
| [5] Ratio Estimator | 0.01521998 | 83.79038 | -27.989 |
| [5] Regression | 0 | 0 | 0 |
| [15] Estimator | 1.338503 | 21539.92 | 16941.3 |
| [17] Ratio | -0.00092405 | 10.84317 | -46.103 |
| [17] Product | 0.01799213 | 51.26086 | 110.319 |
| [16] | 0.1906794 | 76.34123 | 154.855 |
| Proposed Estimator | -0.00200567 | 3.20301640 | 2.1320047 |

**Table 6. MSE and PRE Values of Proposed and Existing Estimators with Respect to $\hat{S}_y^2$**

| | Dataset 1 | | Dataset 3 | | Dataset 3 | |
|---|---|---|---|---|---|---|
| | MSE | PRE | MSE | PRE | MSE | PRE |
| Sample Variance | 0.004715021 | 100.00 | 419607.1 | 100.00 | 188354.1 | 100.00 |
| [5] Ratio Estimator | 0.0009897653 | 476.38 | 220271.5 | 190.50 | 21269.09 | 885.58 |
| [5] Regression | 0.0008908265 | 529.29 | 254877.6 | 164.63 | 20959.36 | 898.66 |
| [15] Estimator | 0.0036798567 | 128.13 | 1204329 | 34.84 | 885096.4 | 21.28 |
| [17] Ratio | 0.0010637556 | 443.24 | 298307.4 | 140.66 | 62669.71 | 300.55 |
| [17] Product | 0.008128177 | 58.01 | 762244.1 | 55.05 | 292602.2 | 64.37 |
| [16] | 0.002745902 | 171.71 | 254968 | 164.57 | 11841.29 | 1590.66 |
| Proposed Estimator | 0.000173869 | 2711.82 | 220094.1 | 190.65 | 9616.144 | 1958.73 |

Table 5 and 6 showed the values of bias, mean square error (MSE) and percentage relative efficiency (PRE) of the proposed estimators and some existing estimators considered using three datasets. Based on the results obtained from the three datasets, it is observed that the proposed estimator has the minimum mean square error and the supreme percentage relative efficiency values, respectively, compared to the sample variance estimator, [5] variance ratio and regression estimators, [15] variance estimator, [17] ratio-type and product-type exponential variance estimators and [16] variance estimator. Therefore, based on the criteria of mean square error and percentage relative efficiency, the proposed estimator performed better and is more efficient than the existing estimators considered.

## 6. Conclusion

In this study, a new generalized estimator of finite population variance using the auxiliary information in stratified random sampling is proposed. The expressions for bias and mean square error equations of the proposed estimator are derived up to first degree of approximation. The theoretical efficiency conditions under which the proposed estimator is better than some existing estimators are obtained. The proposed estimator is thus empirically compared with some of the existing estimators in the literature using real datasets. Evidence from the results revealed that the proposed estimator performs better than the existing estimators considered when compared using mean square error and percentage relative efficiency.

## References


1. Zakari, Y., Muhammad, I., & Sani, N. M. (2020a). Alternative ratio-product type estimator in simple random sampling. *Communication in Physical Sciences,* 5(4), 418-426.
2. Muhammad, I., Zakari, Y., & Audu, A. (2021). An alternative class of ratio-regression-type estimator under two-phase sampling scheme. *CBN Journal of Applied Statistics*, 12(2), 1-26.
3. Oyeyemi, G. M., Muhammad, I., & Kareem, A. O. (2023). Combined Exponential-Type Estimators for Finite Population Mean in Two-Phase Sampling. *Asian Journal of Probability and Statistics*, 21 (2), 44-58.
4. Das, A. K., & Tripathi, T. P. (1978). Use of auxiliary information in estimating the finite population variance. *Sankhya C*, 40, 139–148.

5. Isaki, C. T. (1983). Variance estimation using auxiliary information. *Journal of American Statistics Association*, 78, 117–123.
6. Ahmed, M. S., Walid, A. D., & Ahmed, A. O. H. (2003). Some estimators for finite population variance under two-phase sampling. *Stat Transit*, 6 (1), 143–150.
7. Kadilar, C., & Cingi, H. (2006). Ratio estimators for the population variance in simple and stratified random sampling. *Applied Mathematics and Computation,* 173 (2), 1047–1059.
8. Upadhyaya, L. N., Singh, H. P., & Singh, S. (2004). A class of estimators for estimating the variance of the ratio estimator. *Journal of Japan Statistical Society,* 34 (1), 47–63.
9. Swain, A. K. P. C. (2015). Generalized estimator of finite population variance. *Journal of Statistical Theory and Application*, 14(1), 45–51.
10. Sumramani, J., & Kumarapandiyan, G. (2015). Generalized modified ratio type estimator for estimation of population variance. *Sri-Lankan Journal of Applied Statistics,* 16 (1), 69–90.
11. Singh, R., & Malik, S. (2014). Improved estimation of population variance using information on auxiliary attribute in simple random sampling. *Applied Mathematics and Computation,* 235:43–49.
12. Yadav, S. K., Kadilar, C., & Shabbir, J. (2015). Improved family of estimators of population variance in simple random sampling. *Journal of Statistical Theory Practice*, 9 (2), 219–226.
13. Zakari, Y., Muili, J.O., Tela, M.N., Danchadi, N.S., & Audu, A. (2020b). Use of Unknown Weight to Enhance Ratio-Type Estimator in Simple Random Sampling. *Lapai Journal of Applied and Natural Sciences*, 5(1), 74-81.
14. Muhammad, I., Zakari, Y., & Audu, A. (2022). Generalized estimators for finite population variance using measurable and affordable auxiliary character. *Asian Research Journal of Mathematics*, 18(1), 14-30.
15. Rao, T. J. (1991). On certail methods of improving ratio and regression estimators. *Communication in Statistical-Theory and Methods*, 20 (10), 3325–3340.
16. Shabbir, J., & Gupta, S. (2010). On improvement in variance estimation using auxiliary information. *Communication in Statistical-Theory and Methods*, 36, 2177–2185.
17. Singh, H. P., Upadhyaya, L. N., & Namjoshi, U. D. (2009). Estimation of finite population variance. *Curr Science*, 57, 1331–1334.
18. Ahmad, S., Hussain, S., Shabbir, J., Zahid, E., Aamir, M., & Onyango, R. (2022). Improved Estimation of Finite Population Variance Using Dual Supplementary Information under Stratified Random Sampling. *Mathematical Problems in Engineering*, Article ID 3813952, 12 pages. DOI: 10.35629/5252-0204309319.
19. Koyuncu, N. (2013). Improved Estimators of Finite Population Variance in Stratified Random Sampling. *World Applied Sciences Journal,* 23(12), 130-137.