ON ALTERNATIVE ESTIMATION METHOD FOR FINITE POPULATION TOTAL UNDER SIMPLE RANDOM SAMPLING SCHEME

Adam Taiye Mudi[1*] and Ismail Adekunle Kolawole[1]

[1]Department of Mathematics and Statistics, Kaduna Polytechnic, Kaduna, Nigeria
**Corresponding author E-mail: adamu110202@gmail.com*

ABSTRACT

This paper proposes an alternative estimator for finite population total under Simple Random Sampling Scheme. It suggests a modified composite estimator that combines the Simple Random Sampling estimator with the Ratio Estimator using stratification. The Mean Square Error or Variance of the suggested estimator are expressed mathematically. The empirical comparison was made through real data with some existing estimators to obtain a better estimate under a certain population circumstance. Results indicate that stratification combined with composite weighting reduces both variance and bias, achieving a proportional reduction in variance (PRV) of 35.8%. Hence the proposed estimator performs better.

Keywords: *Composite Estimator***,** *Ratio Estimator, Simple Random Estimator, , Auxiliary Measurement*

1. INTRODUCTION

Composite estimation is a statistical estimation procedure that combines data from several sources, for example, from different surveys or databases or from different periods of time in the same longitudinal survey. It is a method that combines multiple estimators to improve the overall estimation accuracy. It is created by taking a weighted average of two or more individual estimators, each of which may have different strengths and weaknesses. By combining these estimators, a composite estimator can provide more reliable and accurate estimates than any individual estimator alone (see Lee *et al* 2016). This approach is commonly used among others in survey sampling and other statistical estimation processes to minimize bias and reduce the variance of the estimates. The ratio estimator, which incorporates the information on the auxiliary measurements into estimation, is one of the most commonly-used estimators in surveys. When the correlation between the auxiliary measurement and the studied measurement is high, the ratio estimator can outperform the simple random estimator with respect to precision (or variance) and if this correlation is low, the ratio estimator can be less precise than the simple random estimator (Cochran 1977). The ratio estimator is biased, but its bias decreases to 0 as the sample size increases to 1 (Cochran 1977). Because the bias of the ratio estimator is of order 1/n, we commonly focus our attention on variance when comparing the ratio estimator with the simple random estimator (Scheaffer *et al*. 2012). Some discussions on use of the composite estimation technique to improve the precision of existing estimators in small area estimation can be obtained in (Lui and Cumberland 1991; Schaible 1978, Royall 1970) or to extend the ratio estimator to multivariate ratio-type estimators have been presented elsewhere (Sukhatme *et al*. 1984).

A recent discussion on use of the composite estimator with weights proportional to the reciprocal of the variance of its component estimator for two-stage cluster sampling has also been presented by Lee *et al*. (2016). However, the weight suggested by Lee *et al*. (2016) is optimal only when its two component estimators are independent. Also Lui (2020) suggests that the idea of the composite estimator to reduce the bias and variance (or the mean-squared-error (MSE) of the ratio estimator may be applied with no additional efforts of collecting extra data. This paper modifies the work of Lui (2020) by proposing a composite estimator that combines the unbiased stratified simple random estimator with the stratified (combined) ratio estimator to achieve an optimal estimator in terms of reduction in vatiances and the biases of the component estimators.

2. AIM AND OBJECTIVES

The primary aim of this study is to enhance the composite estimator proposed by Lui (2020) by incorporating stratification into its component estimators. Specifically, the study seeks to:

1. Develop a stratified modification of the Lui (2020) composite estimator;
2. Derive expressions for the bias and mean square error (MSE) of the proposed estimator, and
3. Empirically compare the performance of the proposed estimator with existing estimators using real data.

3. METHODOLOGY

Suppose that a population consists of $[(X_i, Y_i) | X_i > 0, Y_i > 0, i = 1, 2, ..., N]$, , where X_i and Y_i are measurements on subject i and N is the population size. The following notations are defined below

$X = \sum_{i=1}^{N} X_i$ and $Y = \sum_{i=1}^{N} Y_i$; population totals of measurements X_i and Y_i,

$\bar{X} = \sum_{i=1}^{N} X_i / \mathrm{N}$ and $\bar{Y} = \sum_{i=1}^{N} Y_i / N$; population means of measurements X_i and Y_i,

$R = Y / X = \bar{Y} / \bar{X}$; population total or mean ratio

$S_x^2 = \sum_{i=1}^{N} (X_i - \bar{X})^2 / (N-1)$; population variance of measurement X_i

$S_y^2 = \sum_{i=1}^{N} (\mathrm{Y}_i - \bar{Y})^2 / (N-1)$; population variance of measurement Y_i

$S_{xy}^2 = \sum_{i=1}^{N} (X_i - \bar{X})(\mathrm{Y}_i - \bar{Y}) / (N-1)$; population covariance between X_i and Y_i,

$\rho = S_{xy} / S_x S_y$; simple correlation coefficient between X_i and Y_i

$CV_x = S_x / \bar{X}$ and $CV_y = S_y / \bar{\mathrm{Y}}$; coefficients of variation for X_i and Y_i

Suppose we employ the use of Simple Random Sampling (SRS) scheme and obtain n samples. The following notations shall also be employed

$\bar{x} = \sum_{i=1}^{n} x_i / \mathrm{n}$ and $\bar{y} = \sum_{i=1}^{n} y_i / \mathrm{n}$; sample means of X_i and Y_i

$\hat{R} = \bar{y} / \bar{\mathrm{x}}$; ratio of the two sample means.

$s_x^2 = \sum_{i=1}^{n} (\mathrm{x}_i - \bar{x})^2 / (\mathrm{n}-1)$; sample variance of X_i

$s_y^2 = \sum_{i=1}^{n} (\mathrm{y}_i - \bar{y})^2 / (\mathrm{n}-1)$; sample variances Y_i

$s_{xy}^2 = \sum_{i=1}^{n} (\mathrm{x}_i - x)(\mathrm{y}_i - \bar{y}) / (\mathrm{n}-1)$ sample covariance between X_i and Y_i

$\hat{\rho} = s_{xy} / \mathrm{s}_x\, s_y$; sample correlation coefficient between X_i and Y_i

$cv_x = s_x / \bar{\mathrm{x}}$ and $cv_y = s_y / \bar{\mathrm{y}}$; coefficients of variation for X_i and Y_i

3.1 Assumptions.

- Considering ratio estimator implies that X_i and Y_i are positively correlated, and
- The correlation between X_i and Y_i is greater than zero ($\rho > 0$)
- The value of the auxiliary variable X is known

Here, our parameter of interest is the population total. The conventional estimator of the population total using Simple Random Estimator and its variance as given by (Cochran,1977) are

$$\hat{Y} = N\bar{y} \tag{1}$$

$$V(\hat{Y}) = N^2 (1 - f) s_y^2 / n \tag{2}$$

where $f = n / N$

The conventional variance estimator for $\hat{Y}_R$ its variance, covariance and bias are approximately given by (Cochran 1977) and (Lui 2020) as

$$\hat{Y}_R = \hat{R}X \tag{3}$$

$$V(\hat{Y}_R) = N^2 (1 - f)(S_y^2 + R^2 S_x^2 - 2RS_{xy}) / n \tag{4}$$

$$Cov(\hat{\mathrm{Y}}, \hat{Y}_R) = N^2 (1 - f)\bar{Y}^2 (CV_y^2 + CV_x^2 - 2\rho CV_y CV_x) / n \tag{5}$$

$$B(\hat{Y}_R) = (1 - f) N\bar{Y} (\mathrm{Cv}_x^2 - \rho Cv_x Cv_y) \tag{6}$$

Considering the use of composite estimator to estimate the population total Y and its variance. According to (Cochran 1977) as given by (Lui 2020), the estimators are

$$\hat{Y}_{comp} = w\hat{Y} + (1 - w)\hat{Y}_R \tag{7}$$

$$V(\hat{Y}_{comp}) = \frac{V(\hat{Y})V(\hat{Y}_R) - Cov(\hat{Y}, \hat{Y}_R)^2}{V(\hat{Y}) + V(\hat{Y}_R) - 2Cov(\hat{Y}, \hat{Y}_R)} \tag{8}$$

Where w is the optimal weight minimizing the variance in (3.8) given by (Lui 2020) as

$$w = \frac{V(\hat{Y}_R) - Cov(\hat{Y}, \hat{Y}_R)}{V(\hat{Y}) + V(\hat{Y}_R) - 2Cov(\hat{Y}, \hat{Y}_R)} \tag{9}$$

Based on equations (2) and (8), Lui (2020) showed that the proportional reduction of variance (PRV) by use of the composite estimator $\hat{Y}_c$ instead of its component parts $\hat{Y}$ and $\hat{Y}_R$ can be estimated respectively by

$$PRV = [V(\hat{Y}) - V(\hat{Y}_{comp})] / \mathrm{V}(\hat{\mathrm{Y}}) \tag{10}$$

and

$$PRV_R = [V(\hat{Y}) - V(\hat{Y}_{comp})] / \mathrm{V}(\hat{\mathrm{Y}}_R) \tag{11}$$

4. PROPOSED MODIFICATION

Consider a population $\cup = (\cup_1, \cup_2, ..., \cup_N)$ of size N based on studied variable y and the auxiliary variable x . Let $\cup$ be partitioned into L mutually exclusive and exhaustive strata with stratum h employing N_h units, $h = 1, 2, ..., L,$ where $\sum_{h=1}^{L} N_h = N$ Let a sample of size n_h be measured using simple random sampling without replacement SRSWOR from the stratum h such that $\sum_{h=1}^{L} n_h = n$ and let (x_{hi}, y_{hi}) are the observed values of the variables $(x, y),$ respectively, on the unit i of the stratum h The notations considered throughout this paper are defined below.

$N;$ size of population,

$N_h;$ size of population in stratum $h,$

$n;$ size of sample,

$n_h;$ size of sample in stratum $h,$

$W_h = N_h / N;$ weight of stratum $h,$

$\bar{y}_h = \sum_{i=1}^{n_h} y_{hi} / n_h;$ sample mean of variable y in stratum $h,$

$\bar{y}_{st} = \sum_{h-1}^{L} W_h \bar{y}_h;$ sample mean of variable y ,

$\bar{x}_h = \sum_{i=1}^{n_h} x_{hi} / n_h;$ sample mean of variable x in stratum $h,$

$\bar{x}_{st} = \sum_{h-1}^{L} W_h \bar{x}_h;$ sample mean of variable x ,

$\bar{Y}_h = \sum_{i=1}^{N_h} y_{hi} / N_h;$ population mean of variable y in stratum $h,$

$\bar{Y}_{st} = \sum_{h=1}^{L} W_h \bar{Y}_h;$ population mean of variable y

$\bar{X}_h = \sum_{i=1}^{N_h} x_{hi} / N_h;$ population mean of variable x in stratum $h,$

$\bar{X}_{st} = \sum_{h=1}^{L} W_h \bar{X}_h;$ population mean of variable x

$R = \bar{Y} / \bar{X}$; population ratio

$R_h = \bar{Y}_h / \bar{X}_h$; population ratio in stratum h,

$S_{yh}^2 = (N_h - 1)^{-1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$; population variance of variable y in stratum h,

$S_{xh}^2 = (N_h - 1)^{-1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2$; population variance of variable x in stratum h,

$S_{xyh}^2 = (N_h - 1)^{-1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)(y_{hi} - \bar{Y}_h)$; population covariance between variables x and y in stratum h,

$\rho_{xyh} = S_{xyh} / S_{xh} S_{yh}$; population correlation coefficient between variables x and y in stratum h,

Cv_{yh}; population coefficient of variation for variable y in stratum h,

Cv_{xh}; population coefficient of variation for variable x in stratum h,

$\hat{R}_h = \bar{y}_h / \bar{x}_h$; sample ratio estimator in stratum h,

$\hat{R}_c = \sum_{h=1}^{L} N_h \bar{y}_h / N_h \bar{x}_h$; combine ratio estimator,

$\hat{Y}_{st} = \sum_{h=1}^{L} N_h \bar{y}_h$; stratified estimator for total (simple random estimator)

$s_{yh}^2 = (n_h - 1)^{-1} \sum_{i=1}^{N_h} (y_{hi} - \bar{y}_h)^2$; sample variance of variable y in stratum h,

$s_{xh}^2 = (n_h - 1)^{-1} \sum_{i=1}^{N_h} (x_{hi} - \bar{x}_h)^2$; sample variance of variable x in stratum h,

$s_{xyh}^2 = (n_h - 1)^{-1} \sum_{i=1}^{N_h} (x_{hi} - \tilde{x}_h)(y_{hi} - \bar{y}_h)$; sample covariance between variables x and y in stratum h,

$\hat{\rho}_{xyh} = s_{xyh} / s_{xh} s_{yh}$; sample correlation coefficient between variables x and y in stratum h,

c_{yh}; sample coefficient of variation for variable y in stratum h,

c_{xh}; sample coefficient of variation for variable x in stratum h,

The proposed estimator $\hat{Y}$ of the population total Y is a linear weighted estimator, when the study variable Y is linearly and positively correlated with the auxiliary variable X

Since our focus is the estimation of population total Y, the stratified unbiased estimator for Y and its variance given by (Cochran 1977) are

$$\hat{Y}_{str} = \sum_{h=1}^{L} N_h \hat{\bar{Y}}_h \tag{12}$$

$$V(\hat{Y}_{str}) = \sum_{h=1}^{L} N_h{}^2 (1 - f_h) s_{yh}^2 / n_h \tag{13}$$

Where $f_h = n_h / N_h$ is the sampling fraction in stratum h

Also, the stratified ratio estimator (combined) of the population total, and its variance are given as

$$\hat{Y}_{Rc} = \hat{R}_c X \tag{14}$$

$$V(\hat{Y}_{Rc}) = \sum_{h=1}^{L} N_h{}^2(1-f_h)(s^2_{yh} + \hat{R}^2 s^2_{xh} - 2\hat{R}S_{xyh})/n_h \quad (15)$$
$$= \sum_{h=1}^{L} N_h{}^2(1-f_h)\,\bar{Y}^2(cv^2_{xh} + cv^2_{yh} - 2\rho_{xyh}c_{xh}c_{yh})/n_h$$

Following Lui (2020) the bias and covariance are derived as

$$B(\hat{Y}_{Rc}) = \sum_{h=1}^{L} N_h{}^2(1-f_h)\,\bar{Y}^2(cv^2_x - \rho_{xyh}cv_{xh}cv_{yh}) \quad (16)$$

$$Cov(\hat{Y}_{str}, \hat{Y}_{Rc}) = \sum_{h=1}^{L} N_h{}^2(1-f_h)\,\bar{Y}^2(cv^2_y - \rho_{xyh}cv_x cv_y)/n_h \quad (17)$$

Considering the use of composite estimator to estimate the population total Y and its variance, the modified version can be given as

$$\hat{Y}_{mcomp} = k\hat{Y}_{str} + (1-k)\hat{Y}_{Rc} \quad (18)$$

Where k is the weight to be determined. The bias and variance, MSE and the optimum value of k of $\hat{Y}_{com}$ are derived respectively below

Let Y be the finite-population total of interest. Suppose we have two stratified estimators of Y (12) and (14), where (12) is an unbiased estimator and (14) is a biased estimator.

The Bias is

$$\begin{aligned} \text{Bias}\{\hat{Y}_{mcomp}(k)\} &= E\left[\hat{Y}_{mconp}(k)\right] - Y \\ &= kE(\hat{Y}_{str}) + (1-k)E(\hat{Y}R_{Rc}) - Y \\ &= kY + (1-k)(Y + \text{Bias}(\hat{Y}_{str})) - Y \\ &= (1-k)\,\text{Bias}(\hat{Y}_{Rc}) \end{aligned} \quad (19)$$

And using standard formula for variance of a linear combination as contained in we have,

$$V(\hat{Y}_{mcomp}) = k^2V(\hat{Y}_{str}) + (1-k)^2V(\hat{Y}_{Rc}) + 2k(1-k)Cov(\hat{Y}, \hat{Y}_{Rc}) \quad (20)$$

And the Mean squared error (MSE) becomes

$$\begin{aligned} \text{MSE}\{\hat{Y}_{mcomp}(k)\} &= \text{Var}\{\hat{Y}_{mcomp}(k)\} + (\text{Bias}\{\hat{Y}_{mcomp}(k)\})^2 \\ &= k^2Var(\hat{Y}_{str}) + (1-k)^2Var(\hat{Y}_{Rc}) + 2k(1-k)Cov(\hat{Y}_{str}, \hat{Y}_{Rc}) \\ &+ (1-k)^2[\text{Bias}\{\hat{Y}_{Rc}\}]^2 \end{aligned} \quad (21)$$

Differentiating (21) with respect to k and set the derivation to zero gives the optimal weight as

$$k^* = \frac{\text{Var}(\hat{Y}_{Rc}) - \text{Cov}(\hat{Y}_{str}, \hat{Y}_{Rc}) + \text{Bias}(\hat{Y}_{Rc})^2}{\text{Var}(\hat{Y}_{str}) + \text{Var}(\hat{Y}_{Rc}) - 2\text{Cov}(\hat{Y}_{str}, \hat{Y}_{Rc}) + \text{Bias}(\hat{Y}_{Rc})^2} \quad (22)$$

Substituting (22) in (19) and (20), the bias and variance of $\hat{Y}_{mcomp}$ becomes

$$\text{Bias}(\hat{Y}_{MCOMP}) = (1-k^*)\text{Bias}(\hat{Y}_{Rst}) \tag{23}$$

$$Var(\hat{Y}_{MCOMP}) = (k^*)^2 Var(\hat{Y}_{st}) + (1-k^*)^2 Var(\hat{Y}_{Rst}) + 2k^*(1-k^*)Cov(\hat{Y}_{st}, \hat{Y}_{Rst}). \tag{24}$$

Substituting the values from eqs (13), (15),(16) and (17) into eqns (19),(20),(21) and (22) we get k_{opt}, bias and the variance of $\hat{Y}_{com}$, to the first degree of approximations, as follows

$$k_{opt} = 1 - \sum_{h=1}^{L} \rho_{xyh} cv_{yh} / cv_{xh} \tag{25}$$

$$\text{B}(\hat{Y}_{mcomp}) = \sum_{h=1}^{L} [(\rho_{xyh} cv_{yh} / cv_{xh})(N_h{}^2 (1-f_h)\bar{\text{Y}}^2 (cv_x^2 - \rho_{xyh} cv_{xh} cv_{yh})] \tag{26}$$

$$V(\hat{Y}_{mcomp}) = \sum_{h=1}^{L} N_h^2 (1-f_h) \bar{Y}^2 cv_y^2 (1-\rho_{xyh}^2) / n_h \tag{27}$$

5. EFFICIENCY COMPARISON

We compare the composite estimator without stratification $\hat{Y}_{comp}$, with proposed estimator $\hat{Y}_{mcomp}$

$$V(\hat{Y}_{mcomp}) < V(\hat{Y}_{comp})$$

$$\Rightarrow \sum_{h=1}^{L} N_h^2 (1-f_h) \bar{Y}^2 cv_{yh}^2 (1-\rho_{xyh}^2) / n_h < N^2 (1-f) \bar{Y}^2 cv_y^2 (1-\rho^2) / n \tag{28}$$

$$\Rightarrow (\sum_{h=1}^{L} N_h^2 (1-f_h) cv_{yh}^2 (1-\rho_{xyh}^2) / n_h < N^2 (1-f) cv_y^2 (1-\rho^2) / n)$$

Let $\varphi_1 = \sum_{h=1}^{L} N_h^2 (1-f_h) cv_{yh}^2 (1-\rho_{xyh}^2) / n_h$ and $\varphi_2 = N^2 (1-f) cv_y^2 (1-\rho^2) / n)$

The estimator $\hat{Y}_{mcomp}$ will be more efficient than $\hat{Y}_{comp}$ if

$\varphi_1 < \varphi_2$

with the proportional reduction as of variance using the modified composite $\hat{Y}_{mcomp}$ instead of the Liu(2020) composite estimator $\hat{Y}_{comp}$ as

$$PRV_{SRSC} = [V(\hat{Y}_{comp}) - V(\hat{Y}_{mcomp})] / \text{V}(\hat{\text{Y}}_{comp}) \tag{29}$$

also

$$\left|\text{B}(\hat{Y}_{mcomp})\right| < \left|B(\hat{Y}_{comp})\right| \tag{30}$$

$$\Rightarrow \sum_{h=1}^{L} (1-f_h) N_h \bar{Y} (\text{Cv}_{xh}^2 - \rho_{xyh} Cv_{x_h} Cv_{y_h}) < (1-f) N\bar{Y} (\text{Cv}_x^2 - \rho Cv_x Cv_y)$$

Let $\phi_1 = \sum_{h=1}^{L} (1-f_h) N_h (\text{Cv}_{xh}^2 - \rho_{xyh} Cv_{x_h} Cv_{y_h})$ and $\phi_2 = (1-f)\,\text{N}(\text{Cv}_x^2 - \rho Cv_x Cv_y)$

The estimator $\hat{Y}_{mcomp}$ will be more efficient than $\hat{Y}_{comp}$ if

$$\phi_1 < \phi_2$$

6. DATA PRESENTATION AND ANALYSIS

The data used is the population census data of 1991 (X) and 2006 (Y) obtained from National Population Commission of Nigeria (web), National Bureau of Statistics (web). Kaduna State. Kaduna state consists of $N = 23$ number of Local Government Areas (LGAs) , using the senatorial zones as strata, simple random sample of three local governments were selected from each Senatorial zone using a table of random number making a sample of $n = 9$ local government areas.
The performance of each estimator was assessed in terms of:

1. ***Bias***: $\text{Bias}(\hat{Y}_{MCOMP}) = E[\hat{Y}_{MCOMP}] - Y$
2. ***Variance***: $\text{Var}(\hat{Y}_{MCOMP}) = E[(\hat{Y}_{MCOMP} - E[\hat{Y}_{MCOMP}])^2]$
3. ***Mean Square Error (MSE)***: $\text{MSE}(\hat{Y}_{MCOMP}) = \text{Var}(\hat{Y}_{MCOMP}) + \text{Bias}^2(\hat{Y}_{MCOMP})$
4. ***Relative Efficiency (RE)*** **:** $$RE(\hat{Y}_{MCOMP}, \hat{Y}_j) = \frac{MSE(\hat{Y}_j)}{MSE(\hat{Y}_{MCOMP})},$$

where $\hat{Y}_j$ represents a comparator estimator. An $RE > 1$ indicates superior performance of the proposed estimator.
The R package is used in the analysis and gives the following summary statistics and results.

$$N = 23, N_1 = 8, N_2 = 7, N_3 = 8, n_1 = n_2 = n_3 = 3, X = 3,935,618$$

$$\bar{x}_1 = 207457.3, \bar{x}_2 = 102599.3, \bar{x}_3 = 201337, \bar{y}_1 = 327239.3, \bar{y}_2 = 216530, \bar{y}_3 = 307423.7$$

$$s_{y1}^2 = 8321970092, s_{y2}^2 = 7960675993, s_{y3}^2 = 783267606$$

$$s_{x1}^2 = 7813554282, s_{x2}^2 = 1373869880, s_{x3}^2 = 635709700$$

$$s_{x1y1} = 8060737929, s_{x2y2} = 3056076331, s_{x3y3} = 108707090$$

Table 1: Comparison of Estimators in terms of Mean Square Error (MSE), Variance (V), Bias and Relative Efficiency (RE)

Estimator	Value	Bias $(\times 10^5)$	Variance $(\times 10^{10})$	MSE $(\times 10^{10})$	Relative Efficiency (RE)
Str	6,482,305	0.00	19.43	19.43	6.7
Ratio	6,550,669	141.00	10.46	12.45	4.29
Comp	6,541,754	90.40	4.01	4.82	1.66
Strc	6,568,975	93.40	8.01	8.89	3.07
Mcomp	6,534,707	56.50	2.58	2.90	1.00
Srs	6,525,813	43.00	24.62	24.80	8.55

7. FINDINGS AND CONCLUSION

When Table 1 is examined, variance and MSE values of modified composite estimator (mcomp), is seen smaller than variances and MSE of the other estimators. This is followed by the Simple Random Composite (Comp) estimator.

Using the **non-stratified composite (Comp)** as the baseline, the **Proportional Reduction in Variance (PRV)** for Mcomp is:

$$PRV = \frac{Var(comp) - Var(mcomp)}{Var(comp)} = \frac{4.01 - 2.58}{4.01} = 0.358 \approx 35.8\%$$

This demonstrates that incorporating stratification into the composite estimator reduces variance by more than one-third relative to the non-stratified composite. The Relative Efficiency (RE) also confirms this finding: mcomp is the most efficient (RE = 1.00 relative to itself, and lower RE for others indicates greater variance), reinforcing the advantage of stratification. The absolute bias for the modified composite estimator has a lesser value as compared to absolute biases of the stratified combine ratio (Strc), conventional ratio (Ratio), and the Lui (2020) (Comp) estimators.

8. CONCLUSION

In conclusion, it is seen that the proposed estimator outperformed the others in terms of variance, MSE, and bias, with the proportional reduction in variance (PRV) due to stratification as 0.3582. This implies that there is a moderate gain in efficiency by use of the modified version as compared with the Liu 2020 estimator, that is, stratification has contributed to the variance and bias reduction.

References

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). John Wiley & Sons.

Lee, S. E., Lee, P. R., & Shin, K. H. (2016). A composite estimator for stratified two-stage cluster sampling. *Communications for Statistical Applications and Methods, 23*(1), 47–55. https://doi.org/10.5351/CSAM.2016.23.1.047

Lui, J. K. (2020). Notes on use of the composite estimator: An improvement of the ratio estimator. *Journal of Official Statistics, 36*(1), 137–149.

Lui, J. K., & Cumberland, W. G. (1991). A model-based approach: Composite estimator for small area estimation. *Journal of Official Statistics, 7*(1), 69–76.

Royall, R. M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika, 57*(2), 377–387. https://doi.org/10.1093/biomet/

Schaible, W. L. (1978). Choosing weights for composite estimators for small area statistics. *Proceedings of the Section on Survey Research Methods* (pp. 741–746). American Statistical Association.

Scheaffer, R. L., Mendenhall, W., Ott, R. L., & Gerow, K. G. (2012). *Elementary survey sampling* (7th ed.). Cengage Learning.