International Journal of Mathematics, Statistics and Computer Science
Vol. 1 No. 4, pp 33-39

An alternative stratified composite estimator for finite population total under simple random sampling

[1,*] Adam Taiye Mudi
[2]Ismail Adekunle Kolawole
[1,2] Department of Mathematics and Statistics,
Kaduna Polytechnic, Kaduna, Nigeria

*adamu110202@gmail.com

Abstract: This paper proposes a stratified composite estimator for the finite population total under a simple random sampling (SRS) scheme. The estimator combines the stratified unbiased SRS estimator with the stratified ratio estimator through an optimal linear combination that minimizes the mean squared error (MSE). Bias, variance, and MSE of the proposed estimator are derived mathematically. An empirical comparison with classical estimators, including simple random, ratio, and non-stratified composite estimators, is conducted using Nigerian census data. Monte Carlo simulations further validate efficiency gains under controlled population scenarios. Results indicate that stratification combined with composite weighting reduces both variance and bias, achieving a proportional reduction in variance (PRV) of 35.8% and 33.3% respectively for real and simulated data .over non-stratified composite.

Keywords: *Composite Estimator; Ratio Estimator; Simple Random Sampling; Stratification; Auxiliary Variable; Finite Population.*

1. INTRODUCTION

Composite estimation is a widely used technique in survey methodology that combines multiple estimators to improve overall accuracy by exploiting the strengths of each component. In the context of finite population sampling, unbiased estimators provide theoretical guarantees of accuracy, while biased estimators such as the ratio estimator gain efficiency by using auxiliary information. Composite estimators balance these competing properties to reduce mean squared error (MSE).

The ratio estimator, in particular, enhances precision when the study and auxiliary variables are strongly correlated [1, 2]. However, it is inherently biased, though its bias is of order $O(1/n)$, often making variance reduction the central criterion of comparison [3]. Composite estimators have been employed in small area estimation [4,5], variance estimation [6], and repeated surveys such as the Current Population Survey [7,8].

Recent developments have extended composite estimation to more complex designs. In [9], a proposal of a non-stratified composite estimator combining unbiased SRS and ratio estimators was canvased. While [10] extended the composite estimation to stratified two-stage cluster sampling. A modified estimator of population variance under stratified sampling was developed in [11] and [12] reported on new calibration-based composite classes while [13] highlighted its application to environmental monitoring surveys.

Building on [9], this paper introduces stratification into the component estimators, as stratification improves representativeness and reduces sampling variability based on [14,15]. Specifically, the objectives are threefold: first, to propose a stratified composite estimator that extends [9] non-stratified version through the incorporation of stratification; second, to derive explicit expressions for the bias, variance, and mean squared error (MSE) of the proposed

Citation: *Mudi TA and Kolawole IA.* An alternative stratified composite estimator for finite population total under simple random sampling. *IJMSCS, 2025. 1(4): p. 33-39.*

estimator; and third, to empirically assess its performance relative to classical alternatives using both census data and simulated populations. The overall aim is to demonstrate that stratification, when combined with composite weighting, yields an estimator with lower MSE and higher efficiency than both classical and non-stratified composite estimators.

2. METHODOLOGY

This section introduces the survey design, the classical estimators, and the proposed stratified composite estimator. All derivations are presented formally, with proof and derivations.

2.1 Sampling Frame

Consider a finite population $U = \{1, 2, \ldots, N\}$ of size N. Let Y denote the study variable of interest and X denote an auxiliary variable correlated with Y. The population is partitioned into L mutually exclusive strata such that:

$$N = \sum_{h=1}^{L} N_h, \quad h = 1, 2, \ldots, L \tag{1}$$

2.1.1 Stratified Simple Random Estimator

The unbiased stratified sample estimator of the population total is:

$$\hat{Y}_{st} = \sum_{h=1}^{L} N_h \overline{Y}_h, \tag{2}$$

with variance:

$$\mathrm{Var}(\hat{Y}_{st}) = \sum_{h=1}^{L} N_h^2 \frac{S_{Yh}^2}{n_h}\left(1 - \frac{n_h}{N_h}\right) \tag{3}$$

where S_{Yh}^2 is the within-stratum variance of the study variable Y

2.1.2 Stratified Ratio Estimator

When auxiliary information is available, the stratified ratio estimator is

$$\hat{Y}_{Rst} = \sum_{h=1}^{L} N_h R_h \overline{X}_h \tag{4}$$

with approximate variance

$$\mathrm{Var}(\hat{Y}_{Rst}) \approx \sum_{h=1}^{L} N_h^2 \frac{S_{e,h}^2}{n_h}\left(1 - \frac{n_h}{N_h}\right) \tag{5}$$

where the residual is defined as

$$\left.\begin{aligned} e_{hi} &= Y_{hi} - R_h X_{hi} \\ S_{e,h}^2 &= \frac{1}{N_h - 1}\sum_{i=1}^{N_h}(e_{hi} - \bar{e}_h)^2 \end{aligned}\right\} \tag{6}$$

3. PROPOSED MODIFIED COMPOSITE ESTIMATOR

The proposed estimator is a weighted linear combination of (2) and (4)

$$\hat{Y}_{MCOMP} = k\hat{Y}_{st} + (1-k)\hat{Y}_{Rst} \tag{7}$$

where $k \in [0,1]$ is chosen to minimize the mean squared error (MSE). The Bias, Variance and MSE of the proposed composite estimator are derived as follows

Let Y be the finite-population total of interest. Suppose we have two stratified estimators of Y (2) and (4), where (2) is an unbiased estimator and (4) is a biased estimator.The Bias using [9] is

$$\begin{aligned}\text{Bias}\{\hat{Y}_{MCOMP}(k)\} &= E\left[\hat{Y}_{MCOMP}(k)\right] - Y \\ &= kE(\hat{Y}_{st}) + (1-k)E(\hat{Y}R_{Rst}) - Y \\ &= kY + (1-k)(Y + \text{Bias}(\hat{Y}_{Rst})) - Y\end{aligned}$$

$$\text{Bias}\{\hat{Y}_{MCOMP}(k)\} = (1-k)\,\text{Bias}(\hat{Y}_{Rst}) \tag{8}$$

The Variance, using standard formula for variance of a linear combination as contained in [9] we have,

$$\text{Var}\{\hat{Y}_{MCOMP}(k)\} = k^2\text{Var}(\hat{Y}_{st}) + (1-k)^2\,\text{Var}(\hat{Y}_{Rst}) + 2k(1-k)\,\text{cov}(\hat{Y}_{st}, \hat{Y}_{Rst})$$

$$\hat{Y}_{MCOMP}(k) = k^2 V_A + (1-k)^2 V_B + 2k(1-k)C_{AB} \tag{9}$$

And the Mean squared error (MSE) becomes

$$\text{MSE}\{\hat{Y}_{MCOMP}(k)\} = \text{Var}\{\hat{Y}_{MCOMP}(k)\} + (\text{Bias}\{\hat{Y}_{MCOMP}(k)\})^2$$

$$\begin{aligned}\text{MSE}\{\hat{Y}_{MCOMP}(k)\} = {} & k^2 Var(\hat{\text{Y}}_{st}) + (1-k)^2 Var(\hat{\text{Y}}_{Rst}) \\ & + 2k(1-k)Cov(\hat{\text{Y}}_{st}, \hat{\text{Y}}_{Rst}) + (1-k)^2[\text{Bias}\{\hat{Y}_{Rst}\}]^2\end{aligned} \tag{10}$$

Differentiating (10) with respect to k and set the derivation to zero gives the optimal weight as

$$k^* = \frac{\text{Var}(\hat{Y}_{Rst}) - \text{Cov}(\hat{Y}_{st}, \hat{Y}_{Rst}) + \text{Bias}(\hat{Y}_{Rst})^2}{\text{Var}(\hat{Y}_{st}) + \text{Var}(\hat{Y}_{Rst}) - 2\text{Cov}(\hat{Y}_{st}, \hat{Y}_{Rst}) + \text{Bias}(\hat{Y}_{Rst})^2} \tag{11}$$

Substituting (11) in (8) and (9), the bias and variance of $\hat{Y}_{MCOMP}$ becomes

$$\text{Bias}(\hat{Y}_{MCOMP}) = (1-k^*)\text{Bias}(\hat{Y}_{Rst}) \tag{12}$$

$$Var(\hat{Y}_{MCOMP}) = (k^*)^2 Var(\hat{Y}_{st}) + (1-k^*)^2 Var(\hat{Y}_{Rst}) + 2k^*(1-k^*)Cov(\hat{Y}_{st}, \hat{Y}_{Rst}). \tag{13}$$

4. DATA PRESENTATION

This study employed both empirical census data and Monte Carlo simulated data to evaluate the performance of the proposed stratified composite estimator.

4.1 Census Data

The population census data of 1991 and 2006 were obtained from the Kaduna State Bureau of Statistics. The empirical focus was on Kaduna State, which is administratively divided into Local Government Areas (LGAs).

For the sampling design, the state was stratified by senatorial zones to ensure geographic representativeness and to capture socio-political diversity across the state. Within each stratum, a simple random sample of three LGAs was selected using a random number table, resulting in a total of nine LGAs. The choice of three LGAs per zone provided a balance between statistical precision and operational feasibility, yielding a sample that adequately represents the population while maintaining manageable field work and computational requirements. The data is given below

Table 1: Census Data

LOCAL GOVERNMENT AREA	1991 CENSUS (X)	2006 CENSUS (Y)
KUBAU	280704	414700
MAKARFI	146574	216600
ZARIA	284318	406990
BIRNI GWARI	143072	258581
GIWA	171061	292384
IGABI	171061	292384
KAURA	101455	174626
KAURU	116284	221276
ZANGO KATAF	140224	318991

Source: *Kaduna State Bureau of Statistics*

4.2 Simulated Data

To complement the census analysis, a Monte Carlo simulation study was conducted to test the robustness of the estimators under controlled conditions. Artificial finite populations were generated with known parameters (totals, means, and auxiliary variables) and stratified into three strata to mimic the structure of Kaduna State. Each stratum contained multiple clusters, from which simple random samples were repeatedly drawn.. Synthetic finite populations of size $N = 1000$ were generated with Study variable Y and auxiliary variable X with correlation $\rho = 0.80$. Assuming Normal Population distribution: $(Y \sim N(100, 20^2)$ Stratifying each population into $L = 3$ strata with proportional sizes 0.3, 0.5, and 0.2. Within each stratum, a simple random sample of size $n_h = 20$ was drawn. The process was repeated $R = 1000$ times to obtain stable estimates.

5. DATA ANALYSIS

Both Census and Monte Carlo Simulated data were analyzed and the performances of the proposed and the considered estimators were assessed and compared.

5.1 Performance Assessment

The performance of each estimator was assessed in terms of

B*ias* $\text{Bias}(\hat{Y}) = E[\hat{Y}] - Y$

Variance $\text{Var}(\hat{Y}) = E[(\hat{Y} - E[\hat{Y}])^2]$

Mean Square Error (MSE): $\text{MSE}(\hat{\theta}) = \text{Var}(\hat{Y}) + \text{Bias}^2(\hat{Y})$

Relative Efficiency (RE) **:** [9] $RE(\hat{Y}_{MCOMP}, \hat{Y}_j) = \dfrac{MSE(\hat{Y}_j)}{MSE(\hat{Y}_{MCOMP})},$

where $\hat{Y}_j$ represents a comparator estimator and $RE > 1$ indicates superior performance of the proposed estimator.

5.2 Efficiency Comparison

The proposed stratified composite estimator (MCOMP) was compared against classical alternatives: the Simple Random Estimator (SRS), the Ratio Estimator (RATIO), the Stratified Unbiased Estimator (STR), the Stratified Ratio Estimator (SR), and the Non-stratified Composite Estimator (COMP). This comparison quantified the relative precision of the proposed estimator in both empirical and simulated contexts.

5.3 Proportional Reduction in Variance (PRV)

To further highlight gains in precision, the Proportional Reduction in Variance (PRV) was computed relative to the non-stratified composite estimator according to [9] as:

$$PRV = \frac{\text{Var}(\hat{Y}_{COMP}) - \text{Var}(\hat{Y}_{MCOMP})}{\text{Var}(\hat{Y}_{COMP})}. \tag{14}$$

A higher PRV value indicates greater variance reduction and improved efficiency of the proposed estimator relative to the baseline.

The analysis was done using R Statistical package and yielded the following results.

Table 2: Empirical Results

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

Table 3: Simulated Result

Estimator	Value	Bias ($\times 10^5$)	Variance ($\times 10^{10}$)	MSE ($\times 10^{10}$)	Relative Efficiency (RE)
STR	100,210	0.00	5.20	5.20	3.00
RATIO	100,540	2.10	3.90	4.34	2.49
COMP	100,480	1.50	2.40	2.62	1.51
STRC	100,590	1.80	2.90	3.22	1.85
MCOMP	100,470	1.20	1.60	1.74	1.00
SRS	100,210	0.00	5.20	5.20	3.00

6. DISCUSSION OF RESULTS

From Table 2 the MCOMP estimator achieved the lowest variance (2.58×10^{10}) and MSE (2.90×10^{10}) among all estimators. The absolute bias (56.50×10^3) was also smaller than that of STRC (93.40×10^3) and RATIO (141.00×10^3), indicating that the proposed estimator balances bias reduction with variance minimization.

Using the non-stratified composite (COMP) as the baseline, the Proportional Reduction in Variance (PRV) for MCOMP as contained in [9] is:

$$\mathrm{PRV} = \frac{\mathrm{Var(COMP)} - \mathrm{Var(MCOMP)}}{\mathrm{Var(COMP)}} = \frac{4.01 - 2.58}{4.01} = 0.358 \approx 35.8\%$$

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.

Similarly, Table 3 shows that the proposed estimator consistently outperforms the alternatives under controlled simulation scenarios. MCOMP achieved the lowest variance (1.60×10^o) and MSE (1.74×10^o). The absolute bias (1.20×10^o) is minimal, further confirming its robustness. Using COMP as baseline, the PRV for MCOMP is:

$$\mathrm{PRV} = \frac{\mathrm{Var(COMP)} - \mathrm{Var(MCOMP)}}{\mathrm{Var(COMP)}} = \frac{2.40 - 1.60}{2.40} = 0.333 \approx 33.3\%$$

These results highlight that the proposed estimator achieves significant variance reduction and efficiency gains under both real-world and simulated scenarios, especially when auxiliary information and stratification are leveraged.

7. CONCLUSION

The study demonstrates that the stratified composite estimator provides significant improvements in precision for estimating finite population totals. By optimally combining stratified simple random sampling and ratio estimators, the method achieves lower mean squared error, reduced variance, and controlled bias relative to both classical and non-stratified composite estimators. The results from both empirical and simulated data confirm its robustness, efficiency, and practical value in survey applications.

These findings suggest that the estimator has strong potential for adoption in official surveys and research where auxiliary information is available. Stratification, when thoughtfully applied to reflect population heterogeneity or administrative divisions such as senatorial zones and local government areas, can substantially enhance efficiency without requiring larger samples. For new or specialized surveys, simulation studies can be used to validate anticipated efficiency gains before implementation. The practical implications extend to national statistical agencies and policy institutions, where improved efficiency translates into reduced costs and enhanced data quality in demographic, labour force, and health surveys.

REFERENCES

[1] Cochran WG. *Sampling Techniques*. 3rd ed. New York: Wiley; 1977.

[2] Kish L. *Survey Sampling*. New York: Wiley; 1965.

[3] Scheaffer RL, Mendenhall W, Ott RL, Gerow KG. *Elementary Survey Sampling*. 7th ed. Boston: Cengage Learning; 2012.

[4] Royall RM. On finite population sampling theory under certain linear regression models. *Biometrika*. 1970;57(2):377–387. https://doi.org/10.1093/biomet/57.2.377

[5] Lui KJ, Cumberland WG. Composite estimation for small area statistics. *J Am Stat Assoc*. 1991;86(413):39–48. https://doi.org/10.1080/01621459.1991.10475001

[6] Wolter KM. *Introduction to Variance Estimation*. New York: Springer-Verlag; 1985.

[7] McIllece J. Evaluation of composite estimators in the Current Population Survey. *Survey Methodol*. 2022;48(2):215–230.

[8] Hungnes H. Composite estimation methods for repeated surveys: An application to the Current Population Survey. *J Off Stat*. 2025;41(1):55–74. https://doi.org/10.2478/jos-2025-0003

[9] Lui KJ. Composite estimation for finite population totals. *Commun Stat Theory Methods*. 2020;49(18):4578–4594. https://doi.org/10.1080/03610926.2019.1577975

[10] Lee H, Lee J, Shin D. Composite estimation in stratified two-stage cluster sampling. *Survey Methodol*. 2016;42(1):67–82.

[11] Zakari M, Muhammad S. Modified estimator of population variance under stratified sampling. *Pak J Stat*. 2023;39(3):201–218.

[12] Merkouris T, Rueda M, Estevao V. Calibration-based composite estimation methods. *J Off Stat*. 2024;40(2):233–254. https://doi.org/10.2478/jos-2024-0009

[13] Garman B, Thompson M, Gregoire TG. Composite estimation for environmental monitoring surveys. *Environ Ecol Stat*. 2024;31(2):245–266. https://doi.org/10.1007/s10651-024-00567-3

[14] Okafor FC. *Sample Survey Theory with Applications*. Nsukka: Afro-Orbis; 2002.

[15] Arnab R. *Survey Sampling Theory and Applications*. London: Academic Press; 2017.

Citation: *Mudi TA and Kolawole IA.* An alternative stratified composite estimator for finite population total under simple random sampling. *IJMSCS, 2025. 1(4): p. 33-39.*