FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 9, September, 2025, pp 359 – 365
DOI: https://doi.org/10.33003/fjs-2025-0909-3824

ALTERNATIVE TWO-PHASE VARIANCE ESTIMATORS IN THE PRESENCE OF RANDOM NONRESPONSE

*[1]Isah Muhammad, [2]Mujtaba Suleiman and [1]Mannir Abdu

[1]Department of Statistics, Binyaminu Usman Polytechnic, Hadejia, Nigeria.
[2]Department of Health Information Management, Federal Polytechnic, Daura, Nigeria.

*Corresponding authors' email: isahsta@gmail.com

ABSTRACT

Variance estimators are utilized to estimate the variability of population under study, and this variation estimates can aid in devising better policies. In this study, a two-phase population variance estimator under two-phase sampling is suggested. The properties of the estimator such as bias and mean square error were derived, and they were compared theoretically with some existing estimators. The efficiency conditions of the modified estimators under two realistic situations of random non-response were derived theoretically. The performances of the estimators were assessed using the criterion of mean square error and percentage relative efficiency. The empirical results using real data sets revealed that the proposed estimators performed better than the existing variance estimators considered. Thus, the proposed estimators in this study can be used to estimate variations that exist in real-world problems when there is random non-response.

Keywords: Variance, Auxiliary variable, Double-sampling, Mean square error, Bias, Efficiency

INTRODUCTION

Estimation of population variance of the study variable is an important issue and has been discussed by many experts engaged in survey statistics. For instance, the determinants of variation in the economy are required for a nation to develop and implement suitable policies for the nation stability, in agriculture, the production variation of crops is required for further planning or in manufacturing industries and pharmaceutical laboratories, the variability of their products is necessary for their quality control (Muhammad *et al.*, 2022 and Muhammad & Oyeyemi, 2025a).

The use of auxiliary information enhances the accuracy and efficiency of survey estimates under various sampling schemes. Recent studies have explored various methods for utilizing auxiliary variables in survey sampling, including the use of extreme values and ranks of auxiliary variables (Zakari *et al.*, 2020; Muhammad *et al.*, 2021; Zakari & Muhammad, 2022; Muhammad *et al.*, 2022; Muhammad *et al.*, 2023; Zakari *et al.*, 2023; Oyeyemi *et al.*, 2023; Zakari & Muhammad, 2023; Audu *et al.*, 2023; Muhammad & Oyeyemi, 2025b).

A Two-phase sampling scheme can be utilized in obtaining an improved estimator when the information on the population variance of the auxiliary variable is not available. This sampling scheme is used to obtain the information about the auxiliary variable cheaply from a larger sample at the first phase and a relatively small sample at the second phase (Neyman, 1938). Studies by many authors have extended the application of two-phase sampling under various strategies, including Cochran (1977), Shabbir and Gupta (2007), Singh *et al.* (1988), Mishra *et al.* (2019) and Muhammad (2023) among others. However, the existing ratio-type estimators are inefficient when there exist a negative correlation between the study and auxiliary variables while the existing product-type estimators are inefficient when there exist a positive correlation between the study and auxiliary variables and this may yield inaccurate results. Therefore, in order to address this problem, this study proposed variance estimator under two-phase sampling that will be more flexible and efficient.

Various estimators under different strategies have been widely developed by many researchers. For instance, a conventional unbiased estimator of variance is developed as

$$t_0 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \quad (1)$$

The expression for the variance of conventional unbiased estimator of variance (t_0) is given as

$$Var(t_0) = \lambda S_y^4 \beta_{2y}^* \quad (2)$$

Isaki (1983) proposed usual ratio estimator of the population variance using auxiliary information under two-phase sampling as

$$t_1 = s_y^2 \left(\frac{s_x^{'2}}{s_x^2}\right) \quad (3)$$

The expression for the mean square error (MSE) equation of the estimator in (3) is given as

$$MSE(t_1) \cong S_y^4\left[\lambda\beta_{2y}^* + \left(\lambda - \lambda^{'}\right)\left(\beta_{2x}^* - 2\lambda_{22}^*\right)\right] \quad (4)$$

Isaki (1983) proposed usual regression estimator of the population variance as

$$t_2 = s_y^2 + b_{(s_y^2, s_x^2)}\left(s_x^{'2} - s_x^2\right) \quad (5)$$

where $\rho_{(s_y^2, s_x^2)} = \lambda_{22}^* / \sqrt{\beta_{2y}^* \beta_{2x}^*}$ is the sample regression coefficient. The expression for the mean square error (MSE) equation of the estimator in (5) is given as

$$MSE(t_2) \quad {}^{4}{}^{*} y_{2y} \frac{S_y^4(\lambda - \lambda^{'})\lambda_{22}^*}{\beta_{2x}^*} \, _{min} \quad (6)$$

Singh *et al.*, (1988) proposed usual difference estimator under two-phase sampling as

$$t_3 = k_1 s_y^2 + k_2(s_x^{'2} - s_x^2) \quad (7)$$

where k_1 and k_2 are unknown constants, whose values are to be determined. The optimum values of k_1 and k_2 along with the minimum mean square error (MSE) equation of the estimator in (7), up to the first order of approximation are given, respectively as

$$k_1^{opt} = \frac{\beta_{2x}^*}{\beta_{2x}^* + \beta_{2y}^*\beta_{2x}^* - \lambda_{22}^{*2}}$$

$$k_2^{opt} = \frac{S_x^2 \lambda_{22}^*}{S_y^2(\beta_{2x}^* + \beta_{2y}^*\beta_{2x}^* - \lambda_{22}^{*2})}$$

$$MSE(t_3) \frac{MSE(t_2)_{min}}{1 + \frac{MSE(t_2)_{min}}{S_y^4}} \, _{min} \quad (8)$$

Shabbir and Gupta (2007) proposed a regression cum exponential variance estimator in two-phase sampling as

$$t_4 = \left[k_3 s_y^2 + k_4(s'^2_x - s_x^2)\right] exp\left(\frac{s_x^{'2} - s_x^2}{s_x^{'2} + s_x^2}\right) \quad (9)$$

where k_3 and k_4 are unknown constants The optimum values of k_3 and k_4 along with the minimum mean square error are given, respectively as

$k_3^{opt} = \frac{\beta_{2x}^*}{8}\left(\frac{8-\beta_{2x}^*}{\beta_{2x}^*+\beta_{2y}^*\beta_{2x}^*-\lambda_{22}^{*2}}\right)$

$$k_4^{opt} = \frac{S_y^2}{8S_x^2}\left(\frac{-4\beta_{2y}^* + \beta_{2x}^* + 8\lambda_{22}^* - \lambda_{22}^*\beta_{2x}^* + 4\beta_{2y}^*\beta_{2x}^* - 4\lambda_{22}^{*2}}{(\beta_{2x}^* + \beta_{2y}^*\beta_{2x}^* - \lambda_{22}^{*2})}\right)$$

$$MSE(t_4) \frac{MSE(t_2)_{min}}{1+\frac{MSE(t_2)_{min}}{S_y^4} - \frac{(\lambda-\lambda')\beta_{2x}^*\left[MSE(t_2)\frac{(\lambda-\lambda')S_y^4\beta_{2x}^*}{16}_{min}[]\right]}{4\left[1+\frac{MSE(t_2)_{min}}{S_y^4}[]\right]}}_{min} \quad (10)$$

Mishra *et al.* (2019) proposed four class of estimators for estimating population variance under double sampling scheme using log type transformation as

$$Pl_1 = s_y^2 + w_0 \log\left(\frac{s_x^2}{s_x^{'2}}\right) \quad (11)$$

$$Pl_2 = s_y^2(w_1 + 1) + w_2 \log\left(\frac{s_x^2}{s_x^{'2}}\right) \quad (12)$$

$$Pl_3 = \left[s_y^2(w_3 + 1) + w_4 \log\left(\frac{s_x^2}{s_x^{'2}}\right)\right] exp\left\{\frac{s_x^{'2}-s_x^2}{s_x^{'2}+s_x^2}\right\} \quad (13)$$

$$Pl_4 = s_y^2(w_5 + 1) + w_6 \log\left(\frac{s_x^2}{s_x^{'2}}\right) exp\left\{\frac{s_x^{'2}-s_x^2}{s_x^{'2}+s_x^2}\right\} \quad (14)$$

where $w_0, w_1, w_2, w_3, w_4, w_5$ and w_6 are unknown constants. The optimum values of the parameters involved along with the expression for the minimum mean square error (MSE) equation of the estimators in (11-14), up to the first order of approximation are given, respectively as

$$MSE(Pl_1) \quad {}^4_y\left[\lambda\beta_{2y}^* - (\lambda-\lambda')\frac{\lambda_{22}^*}{\beta_{2x}^*}\right]_{min} \quad (15)$$

$$MSE(Pl_2) \frac{BC^2+(A-2C)D^2}{D^2-AB}_{min} \quad (16)$$

$$MSE(Pl_3)1 \frac{B_1C_1^2+A_1D_1^2-2C_1D_1^2}{E_1^2-A_1B_1}_{min} \quad (17)$$

$$MSE(Pl_4)3 \frac{B_3C_3^2+A_3D_3^2-2C_3D_3^2}{E_3^2-A_3B_3}_{min} \quad (18)$$

where:

$w_0 = \left[\frac{-S_y^2\lambda_{22}^*}{\beta_{2x}^*}\right]$, $w_1^{opt} = \frac{(A-C)D}{D^2-AB}$, $w_2^{opt} = \frac{CB-D^2}{D^2-AB}$,
$w_3^{opt} = \frac{C_1B_1-D_1E_1}{E_1^2-A_1B_1}$, $w_4^{opt} = \frac{A_1D_1-C_1E_1}{E_1^2-A_1B_1}$, $w_5^{opt} = \frac{C_3B_3-D_3E_3}{E_3^2-A_3B_3}$,
$w_6^{opt} = \frac{A_3D_3-C_3E_3}{E_3^2-A_3B_3}$, $A = S_y^4(1+\lambda\beta_{2y}^*)$, $B = (\lambda-\lambda')\beta_{2x}^*$,
$C = S_y^4\lambda\beta_{2y}^*$, $D = S_y^2(\lambda-\lambda')\lambda_{22}^*$, $A_1 = S_y^4(1+\lambda\beta_{2y}^*)$,
$B_1 = (\lambda-\lambda')\beta_{2x}^*$, $C_1 = S_y^4\lambda\beta_{2y}^*$,
$D_1 = S_y^2(\lambda-\lambda')\left(\lambda_{22}^* - \frac{\beta_{2x}^*}{2}\right)$, $E_1 = S_y^2(\lambda-\lambda')(\lambda_{22}^* - \beta_{2x}^*)$,
$F_1 = S_y^4\left[\lambda\beta_{2y}^* + (\lambda-\lambda')\left(\frac{\beta_{2x}^*}{4} - \lambda_{22}^*\right)\right]$,
$A_3 = S_y^4(1+\lambda\beta_{2y}^*)$, $B_3 = (\lambda-\lambda')\beta_{2x}^*$, $C_3 = S_y^4\lambda\beta_{2y}^*$,
$D_3 = S_y^2(\lambda-\lambda')\lambda_{22}^*$, $E_3 = S_y^2(\lambda-\lambda')\left(\lambda_{22}^* - \frac{\beta_{2x}^*}{2}\right)$,
$E_1 = S_y^4(\lambda-\lambda')(\lambda_{22}^* - \beta_{2x}^*)$ and
$F_1 = S_y^4\left[\lambda\beta_{2y}^* + (\lambda-\lambda')\left(\frac{\beta_{2x}^*}{4} - \lambda_{22}^*\right)\right]$

Methodology

Two-phase Sampling Procedure

The following scenario is adopted in proposing the current methodology; consider a finite population $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_N)$ of size *N*, from which a sample of size *n* units from ζ is selected by using simple random sample without replacement *(SRSWOR)*. Let (y_i, x_i) be the value of the study variable Y and the auxiliary variable X on ith unit $\zeta_i, i = 1, \ldots N$. Let $\overline{Y}$ and $\overline{X}$ be population means of the study variable Y and the auxiliary variable X, respectively. We assume that the population mean $\overline{X}$ and the population variance S_x^2 of the auxiliary variable are known.

Let $s_y^{*2} = \sum_{i=1}^{n-r}(y_i - \bar{y}^*)^2/(n-r-1)$ and $s_x^{*2} = \sum_{i=1}^{n-r}(x_i - \bar{x}^*)^2/(n-r-1)$ be the sample variance of study and auxiliary variables, respectively on the basis of the responding part of sample, $S_y^2 = \sum_{i=1}^{N}(y_i - \overline{Y})^2/(N-1)$ and $S_x^2 = \sum_{i=1}^{N}(x_i - \overline{X})^2/(N-1)$ be the population variance of the study and auxiliary variables, respectively. Under the double sampling scheme, the first phase sample $S'(S' \subset \zeta)$ of a fixed size n is drawn to measure only on the auxiliary variable x in order to formulate a good estimate of population mean $\overline{X}$.

Random Non-Response

Random non-response as defined by Singh & Joarder (1998), in sample S_n of size n drawn at the first phase, let r_1 sampling units lack complete information due to random non-response, where r_1 can take values from the set $\{0,1,\ldots,(n-2)\}$. Similarly, in the second phase sample S_m of size m_h, let complete information be unavailable for r_2 sampling units, where r_2 lies in the range $\{0,1,2,\ldots,(m-2)\}$ It is assumed that $r_j \geq 0, j = 1,2$ and $r_1 \leq (n-2)$, $r_2 \leq (m-2)$. Non-response can take $(n-2)$ and $(m-2)$ possible values in the samples S_n and S_m, respectively. Let these respective probabilities be denoted by p_1 and p_2. The total number of ways in which $r_j (j = 1, 2)$ non-responses can be obtained are ${}^{n-2}C_{r_1}$ and ${}^{m-2}C_{r_2}$, respectively. Then r_1 and r_2 are discrete random variables having the respective probability distributions as given below:

$$P(r_1) = \frac{n-r_1}{nq_1+2p_1}\,{}^{n-2}C_{r_1}p_1^{r_1}q_1^{n-r_1-2}; \qquad r_1 = 0,1,2,\ldots,(n-2)$$

and

$$P(r_2) = \frac{m-r_2}{nq_2+2p_2}\,{}^{m-2}C_{r_2}p_2^{r_2}q_2^{m-r_2-2}; \qquad r_2 = 0,1,2,\ldots,(m-2)$$

where $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

Proposed Estimator

In this section, a new estimator for estimating the finite population variance under two-phase sampling is proposed when random non-response occurs on both study and auxiliary variables as:

$$\hat{T}_{prop_i}^{**} = \left\{k_1 s_y^{*2}\left[\frac{1}{2}\left(\frac{s_x^{'*2}}{s_x^{*2}} + \frac{s_x^{*2}}{s_x^{'*2}}\right)\right]^{\delta} + k_2(s_x^{'*2} - s_x^{*2})\right\} exp\left\{\frac{s_x^{'*2}-s_x^{*2}}{s_x^{'*2}+s_x^{*2}}\right\} \quad (19)$$

where, $\hat{T}_{prop_i}^{**}$ is the proposed estimator, β_1 and β_2 are real parameters to be determined such that the mean square error of the proposed estimator $\hat{T}_{prop_i}^{**}$ is minimum and δ is a driving parameter suitably chosen. To obtain the bias and MSE equations for the proposed estimator, we define the following notations:

$s_y^{*2} = S_y^2(1+e_0)$, $s_x^{*2} = S_x^2(1+e_1)$ and $s_x^{'*2} = S_x^2(1+e_2)$

Properties of the Proposed Estimator

To derive the properties of the modified estimators, we consider the following existing results as defined in Isaki (1983).

$E(e_0) = E(e_1) = E(e_2) = 0$

$E(e_0^2) = f_2^*(\lambda_{40} - 1) = f_2^*C_0^2$, $E(e_1^2) = f_2^*(\lambda_{04} - 1) = f_2^*C_1^2$, $E(e_2^2) = f_1^*(\lambda_{04} - 1) = f_1^*C_1^2$, $E(e_0e_1) = f_2^*(\lambda_{22} - 1) = f_2^*\rho_{01}$, $E(e_0e_2) = f_1^*(\lambda_{22} - 1) = f_1^*\rho_{01}$, $E(e_1e_2) = f_1^*(\lambda_{04} - 1) = f_1^*C_1^2$

where $f_1^* = \left(\frac{1}{nq_1+2p_1} - \frac{1}{N}\right)$ and $f_2^* = \left(\frac{1}{mq_2+2p_2} - \frac{1}{N}\right)$

Expressing the estimator $\hat{T}_{prop_i}^{**}$ in terms of e_i $(i = 0, 1, 2)$ we can write (19) as

$$\hat{T}^{**}_{prop_i} = \begin{Bmatrix} k_1 S_y^2(1+e_0)\frac{1}{2^\delta}\left[(1+e_2)(1+e_1)^{-1} + (1+e_1)(1+e_2)^{-1}\right]^\delta \\ -k_2\left[S_x^2(1+e_2) - S_x^2(1+e_1)\right] \end{Bmatrix} exp\left\{\frac{S_x^2(1+e_2) - S_x^2(1+e_1)}{S_x^2(1+e_2) + S_x^2(1+e_1)}\right\} \quad (20)$$

Expanding the RHS of (20) to the first order of approximation, we get:

$$\hat{T}^{**}_{prop_i} = \left\{k_1 S_y^2(1+e_0)\left[1 + \frac{\delta e_1^2}{2} + \frac{\delta e_2^2}{2} - \delta e_1 e_2\right] - k_2 S_x^2(e_2 - e_1)\right\} exp\left\{1 + \frac{e_2}{2} - \frac{e_1}{2} - \frac{e_2^2}{8} + \frac{3e_1^2}{8} - \frac{e_1 e_2}{4}\right\} \quad (21)$$

Expanding the RHS of (21) to the first order of approximation, neglecting the terms of e's greater than two and taking out the common terms gives:

$$\hat{T}^{**}_{prop_i} - S_y^2 = \begin{Bmatrix} (k_1 - 1)S_y^2 + k_1 S_y^2\left[e_0 - \frac{e_1}{2} + \frac{e_2}{2} + A_i e_1^2 + B_i e_2^2 - \frac{e_0 e_1}{2} + \frac{e_0 e_2}{2} - D_i e_1 e_2\right] \\ +k_2 S_x^2\left[e_1 - e_2 - \frac{e_1^2}{2} - \frac{e_2^2}{2} + e_1 e_2\right] \end{Bmatrix} \quad (22)$$

where $A_i = \left[\frac{4\delta+3}{8}\right]$, $B_i = \left[\frac{4\delta-1}{8}\right]$ and $D_i = \left[\frac{8\delta+1}{8}\right]$

Taking expectation of (22), the bias of the proposed generalized estimator is obtained as:

$$Bias\left(\hat{T}^{**}_{prop_i}\right) == \begin{Bmatrix} (k_1 - 1)S_y^2 + k_1 S_y^2\left[\left(f_2^* A_i + f_1^* B_i - f_1^* D_i\right)C_1^2 + \frac{1}{2}\left(f_1^* - f_2^*\right)\rho_{01}\right] \\ +k_2 S_x^2 \frac{1}{2}\left(f_1^* - f_2^*\right)C_1^2 \end{Bmatrix} \quad (23)$$

Squaring equation (23) to the first order of approximation, neglecting the terms of e's greater than two and taking out the common terms gives:

$$\left(\hat{T}^{**}_{prop_i} - S_y^2\right)^2 = \begin{Bmatrix} (k_1 - 1)^2 S_y^4 + k_1^2 S_y^4 \begin{bmatrix} 2e_0 - e_1 + e_2 + e_0^2 + \frac{(8A_i+1)e_1^2}{4} + \frac{(8B_i+1)e_2^2}{4} \\ -2e_0 e_1 + 2e_0 e_2 - \frac{(4D_i+1)e_1 e_2}{2} \end{bmatrix} \\ +k_2^2 S_x^4(e_1^2 + e_2^2 - 2e_1 e_2) - 2k_1 k_2 S_y^2 S_x^2(e_1 - e_2 - e_1^2 - e_2^2 + e_0 e_1 - e_0 e_2 + 2e_1 e_2) \\ -2k_1 S_y^4\left(e_0 - \frac{e_1}{2} + \frac{e_2}{2} + A_i e_1^2 + B_i e_2^2 - \frac{e_0 e_1}{2} + \frac{e_0 e_2}{2} - D_i e_1 e_2\right) \\ -2k_2 S_y^2 S_x^2\left(e_1 - e_2 - \frac{e_1^2}{2} - \frac{e_2^2}{2} + e_1 e_2\right) \end{Bmatrix} \quad (24)$$

Taking expectation of (24), the mean square error of the proposed estimator, up to the first order of approximation, is given as:

$$MSE\left(\hat{T}^{**}_{prop_i}\right) = \begin{Bmatrix} (k_1 - 1)^2 S_y^4 + k_1^2 S_y^4\left[f_2^* C_0^2 + \frac{1}{4}\left([8B_i - 8D_i - 1]f_2^* + [8A_i + 1]f_1^*\right)C_1^2 + 2\left(f_1^* - f_2^*\right)\rho_{01}\right] \\ +k_2^2 S_x^4\left(f_2^* - f_1^*\right)C_1^2 + 2k_1 k_2 S_y^2 S_x^2\left[\left(f_1^* - f_2^*\right)\left(C_1^2 - \rho_{01}\right)\right] \\ -2k_1 S_y^4\left[\left(A_i f_2^* + (B_i - D_i)f_1^*\right)C_1^2 + \frac{1}{2}\left(f_1^* - f_2^*\right)\rho_{01}\right] - 2k_2 S_y^2 S_x^2 \frac{\left(f_1^* - f_2^*\right)}{2} C_1^2 \end{Bmatrix} \quad (25)$$

The MSE equation of the proposed estimator is subsequently reduced from equation (25) as:

$$MSE\left(\hat{T}^{**}_{prop_i}\right) = (k_1 - 1)^2 S_y^4 + k_1^2 S_y^4 J_i + k_2^2 S_x^4 E_i + 2k_1 k_2 S_y^2 S_x^2 F_i - 2k_1 S_y^4 G_i - 2k_2 S_y^2 S_x^2 H_i \quad (26)$$

where:

$J_i = \left[f_2^* C_0^2 + \frac{1}{4}\left([8B_i - 8D_i - 1]f_2^* + [8A_i + 1]f_1^*\right)C_1^2 + 2\left(f_1^* - f_2^*\right)\rho_{01}\right]$

$E_i = \left(f_1^* - f_2^*\right)C_1^2$

$F_i = \left[\left(f_1^* - f_2^*\right)\left(C_1^2 - \rho_{01}\right)\right]$

$G_i = \left[\left(A_i f_2^* + (B_i - D_i)f_1^*\right)C_1^2 + \frac{1}{2}\left(f_1^* - f_2^*\right)\rho_{01}\right]$

$H_i = \frac{1}{2}\left(f_1^* - f_2^*\right)C_1^2$

We obtain the optimum values of k_1 and k_2, by differentiating (26) partially with respect to k_1 and k_2, and equating to zero, respectively as:

$$k^{**}_{1(opt)} = \frac{(1+G_i)E_i - F_i H_i}{(1+J_i)E_i - F_i^2}$$

and

$$k^{**}_{2(opt)} = \frac{S_y^2}{S_x^2}\left[\frac{(1+J_i)H_i - (1+G_i)F_i}{(1+J_i)E_i - F_i^2}\right]$$

We obtain the minimum mean square error of the proposed estimator $\hat{T}^{**}_{prop_i}$ by substituting the optimum values of k_1 and k_2 into Equation (27), as:

$$MSE\left(\hat{T}^{**}_{prop_i}\right) \quad {}^4_y\left\{1 - \frac{\begin{bmatrix}(1+J_i)\left((G_i+1)^2 E_i^2 - F_i^2 H_i^2\right) + 2(1+G_i)\left(F_i^2 - J_i E_i - E_i\right)F_i H_i \\ +(1+2J_i)E_i H_i^2 - (1+G_i)^2 E_i F_i^2\end{bmatrix}}{\left[(1+J_i)E_i - F_i^2\right]^2}\right\}_{min} \quad (27)$$

Special Classes:

For $\delta = 1$, the proposed estimator in equation (19) becomes

$$\hat{T}^{**}_{prop_1} = \left\{k_1 s_y^2\left[\frac{1}{2}\left(\frac{s_x^{'*2}}{s_x^{*2}} + \frac{s_x^{*2}}{s_x^{'*2}}\right)\right] + k_2\left(s_x^{'*2} - s_x^{*2}\right)\right\} exp\left\{\frac{s_x^{'*2} - s_x^{*2}}{s_x^{'*2} + s_x^{*2}}\right\} \quad (28)$$

The optimum values of k_1 and k_2, of the proposed estimator in equation (28) are obtain as

$$k^{**}_{1(opt)} = \frac{(1+G_1)E_1 - F_1 H_1}{(1+J_1)E_1 - F_1^2}$$

and

$$k^{**}_{2(opt)} = \frac{S_y^2}{S_x^2}\left[\frac{(1+J_1)H_1 - (1+G_1)F_1}{(1+J_1)E_1 - F_1^2}\right]$$

The minimum mean squared error of the proposed estimator in equation (28) is obtained as

$$MSE\left(\hat{T}^{**}_{prop_1}\right) \quad {}^{4}_{y}\left\{1-\frac{\left[\begin{array}{l}(1+J_1)((G_1+1)^2E_1^2-F_1^2H_1^2)+2(1+G_1)(F_1^2-J_1E_1-E_1)F_1H_1\\+(1+2J_1)E_1H_1^2-(1+G_1)^2E_1F_1^2\end{array}\right]}{[(1+J_1)E_1-F_1^2]^2}\right\}_{min} \quad (29)$$

When $\delta = 2$, the proposed estimator in equation (19) becomes

$$\hat{T}^{**}_{prop_2} = \left\{k_1 s_y^2\left[\frac{1}{2}\left(\frac{s_x^{'*2}}{s_x^{*2}}+\frac{s_x^{*2}}{s_x^{'*2}}\right)\right]^2 + k_2\left(s_x^{'*2}-s_x^{*2}\right)\right\} exp\left\{\frac{s_x^{'*2}-s_x^{*2}}{s_x^{'*2}+s_x^{*2}}\right\} \quad (30)$$

The optimum values of k_1 and k_2, of the proposed estimator in equation (30) are obtain as:

$$k^{**}_{1(opt)} = \frac{(1+G_2)E_2-F_2H_2}{(1+J_2)E_2-F_2^2}$$

and

$$k^{**}_{2(opt)} = \frac{S_y^2}{S_x^2}\left[\frac{(1+J_2)H_2-(1+G_2)F_2}{(1+J_2)E_2-F_2^2}\right]$$

The minimum mean squared error of the proposed estimator in equation (30) is obtain as

$$MSE\left(\hat{T}^{**}_{prop_2}\right) \quad {}^{4}_{y}\left\{1-\frac{\left[\begin{array}{l}(1+J_2)((G_2+1)^2E_2^2-F_2^2H_2^2)+2(1+G_2)(F_2^2-J_2E_2-E_2)F_2H_2\\+(1+2J_2)E_2H_2^2-(1+G_2)^2E_2F_2^2\end{array}\right]}{[(1+J_2)E_2-F_2^2]^2}\right\}_{min} \quad (31)$$

Efficiency Comparisons

In this section, the conditions under which the proposed estimators are more efficient than existing estimators considered are presented.

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

$Var(\hat{s}_y^2) - MSE(\hat{S}^2_{d_i})_{min}$, if

$$S_y^4\lambda(\beta_{2y}-1) - S_y^4\left(1-\frac{K}{M}\right) > 0$$

$$\lambda(\beta_{2y}-1) - \left(1-\frac{K}{M}\right) > 0 \quad (32)$$

where:

$K = (1+D_2)((G_2+1)^2E_2^2 - F_2^2H_2^2) + 2(1+G_2)(F_2^2 - D_2E_2 - E_2)F_2H_2 + (1+2D_2)E_2H_2^2 - (1+G_2)^2E_2F_2^2$ $M = [(1+D_2)E_2 - F_2^2]^2$

ii. Comparing the proposed estimator's MSE with that usual ratio variance estimator defined by Isaki (1983), we have:

$MSE(t_1) - MSE(\hat{S}^2_{d_i})_{min}$, if

$$S_y^4[\lambda\beta^*_{2y} + (\lambda-\lambda^{'})(\beta^*_{2x} - 2\lambda^*_{22})] - S_y^4\left(1-\frac{K}{M}\right) > 0$$

$$[\lambda\beta^*_{2y} + (\lambda-\lambda^{'})(\beta^*_{2x} - 2\lambda^*_{22})] - \left(1-\frac{K}{M}\right) > 0 \quad (33)$$

iii. Comparing the proposed estimator's MSE with that usual regression variance estimator defined by Isaki (1983), we have:

$MSE(t_2)\hat{S}^2_{d_{i_{min_{min}}}}$, if

$$S_y^4\left[\lambda\beta^*_{2y} - \frac{(\lambda-\lambda^{'})\lambda^*_{22}}{\beta^*_{2x}}\right] - S_y^4\left(1-\frac{K}{M}\right) > 0$$

$$\left[\lambda\beta^*_{2y} - \frac{(\lambda-\lambda^{'})\lambda^*_{22}}{\beta^*_{2x}}\right] - \left(1-\frac{K}{M}\right) > 0 \quad (34)$$

iv. Comparing the proposed estimator's MSE with that usual difference variance estimator defined by Singh *et al* (1988), we have:

$MSE(t_3)\hat{S}^2_{d_{i_{min_{min}}}}$, if

$$S_y^4\left[\frac{MSE(t_2)_{min}}{1+\frac{MSE(t_2)_{min}}{S_y^4}} - S_y^4\left(1-\frac{K}{M}\right) > 0\right]$$

$$\left[\frac{MSE(t_2)_{min}}{1+\frac{MSE(t_2)_{min}}{S_y^4}} - \left(1-\frac{K}{M}\right) > 0\right] \quad (35)$$

v. Comparing the proposed estimator's MSE with that usual difference variance estimator defined by Shabbir and Gupta (2007), we have:

$MSE(t_3)\hat{S}^2_{d_{i_{min_{min}}}}$, if

$$\left[\frac{MSE(t_2)_{min}}{1+\frac{MSE(t_2)_{min}}{S_y^4} - \frac{(\lambda-\lambda^{'})\beta^*_{2x}\left[MSE(t_2)\frac{(\lambda-\lambda^{'})S_y^4\beta^*_{2x}}{16}{}_{min}[]\right]}{4\left[1+\frac{MSE(t_2)_{min}}{S_y^4}[]\right]}} - \left(1-\frac{K}{M}\right) > 0\right] \quad (36)$$

Since condition (58) is satisfied, the proposed estimator is more flexible and efficient than the variance estimator defined by Shabbir and Gupta (2007).

Data Source and Description

Four real-life data sets were used to illustrate the efficacious performances of the proposed estimators. The sources of the data sets, the nature of the variables y, x, z and the values of the various parameters are given as follows:

Dataset I: Cochran (1977)

y: Number of placebo children.

x: Number of paralytic polio cases in the placebo group.

z: Number of paralytic polio cases in the „not inoculated" group.

$N = 34,\quad n = 20,\quad m = 12,\quad S_y^2 = 23154.85561,\quad S_x^2 = 28123.21925$ $C_0^2 = 2.32188$, $C_0^2 = 1.82685$, $C_x = 1.2333$, $\rho_{01} = 0.6661$, ρ = 02 0.5657, ρ = 12 0.6005, λ = 030 1.5224 and λ = 210 1.4083.

Dataset II: Murthy (1967)

y: Area under wheat in 1964.

x: Area under wheat in 1963.

z: Cultivated area in 1961.

N= 34, n = 20, m= 12, C0 = 1.6510, C1 = 1.3828, Cx = 0.7205, ρ = 01 0.9218, ρ = 020.8914, ρ = 12 0.9346, λ = 030 0.9345 and λ = 210 1.0196.

Dataset III: Sukhatme & Sukhatme (1970)

y: Area under wheat in 1937.

x: Area under wheat in 1936.

z:Total cultivated area in 1931.

N= 34, n = 20, m= 12, C0 = 1.5959, C1 = 1.5105, Cx = 0.7678, ρ = 01 0.6251, ρ = 02 0.8007, ρ = 12 0.5342, λ = 030 1.0982 and λ = 210 0.8886.

Dataset IV: Murthy (1967)

y: Output.

x: Fixed Capital

z: Number of workers.

N= 80, n = 60, m= 40, C0 = 1.1255, C1 = 1.6065, Cx = 0.9485, ρ = 01 0.7319, ρ = 02 0.7940, ρ = 12 0.9716, λ = 030 1.2761 and λ = 210 0.5461.

RESULTS AND DISCUSSION

Comparison of Estimators

The empirical results of the proposed and some existing estimators were computed and presented in tables below.

Table 1: Various Estimators' MSE and PRE Values with Regard to s_y^2

Estimators	Dataset I		Dataset II	
	MSE	PRE	MSE	PRE
Sample variance (t_0)	9286.713	100	300106473570	100
Isaki (1983) Classical Ratio (t_1)	10506.05	88.394	29256901796.03	1025.763
Isaki (1983) Classical Regression (t_2)	7554.903	122.923	27722181291.40	1082.550
Singh *et al.,* (1988) (t_3)	6926.661	134.072	27429779412.11	1094.090
Shabbir and Gupta (2007) (t_4)	6720.736	138.180	26492917739.54	1132.780
Mishra *et al.,* (2019) Estimator (Pl_1)	7554.903	241.1409	27722181291.40	1082.550
Mishra *et al.,* (2019) Estimator (Pl_2)	6926.661	122.923	27429779412.11	1094.090
Mishra *et al.,* (2019) Estimator (Pl_3)	6684.022	134.072	27672258497.21	1084.503
Mishra *et al.,* (2019) Estimator (Pl_4)	6569.176	141.368	19355064814.53	1550.532
Proposed Estimator ($\hat{T}^{**}_{prop_1}$)	3218.826	288.512	15684959477	1913.339
Proposed Estimator ($\hat{T}^{**}_{prop_2}$)	1888.781	491.678	7586821411	3955.629

Table 1 Presented the Mean Square Error (MSE) and Percentage Relative Efficiency (PRE) Values of the Proposed and some Existing Estimators Considered using Datasets 1 and 2. The Proposed Estimator Performed Better with Minimum MSE and Higher PRE Values.

Table 2: Various Estimators' MSE and PRE Values with Regard to s_y^2

Estimators	Dataset III		Dataset IV	
	MSE	PRE	MSE	PRE
Sample variance (t_0)	558.2446	100	209772707	100
Isaki (1983) Classical Ratio (t_1)	190.8131	292.561	28490643.789	736.2863
Isaki (1983) Classical Regression (t_2)	107.935	517.205	18485825.132	1134.776
Singh *et al.,* (1988) (t_3)	107.369	519.932	18156904.286	1155.333
Shabbir and Gupta (2007) (t_4)	105.521	529.036	17455922.388	1201.728
Mishra *et al.,* (2019) Estimator (Pl_1)	107.935	517.205	18485825.132	1134.776
Mishra *et al.,* (2019) Estimator (Pl_2)	107.369	519.932	18156904.286	1155.333
Mishra *et al.,* (2019) Estimator (Pl_3)	107.058	521.422	18319672.456	1145.068
Mishra *et al.,* (2019) Estimator (Pl_4)	99.1400	563.085	11136040.182	1883.728
Proposed Estimator ($\hat{T}^{**}_{prop_1}$)	54.2591	1028.85	4740125	4425.468
Proposed Estimator ($\hat{T}^{**}_{prop_2}$)	13.0590	4274.79	1890131	11098.32

Table 2 Displayed the Estimators' mean Square Error (MSE) and Percentage Relative Efficiency (PRE) Values Using Datasets 1 and 2. The Result Implied that the Proposed Estimator Performed Better with Minimum MSE and Higher PRE Values.

Discussion

A new two-phase variance estimator in the presence of random non-response was developed in this study. The performance of the proposed and some existing estimators based on the criteria of mean square error and percentage relative efficiency were assessed using four real-life datasets. The results obtained from dataset I revealed that the proposed classes of estimator; I and II (3218.826 and 288.512;

1888.781 and 491.678) have minimum MSE and higher PRE values, respectively, compared to the sample variance (9286.713 and 100); Isaki (1983) classical ratio (10506.05 and 88.394); Isaki (1983) classical regression (7554.903 and 122.923); Singh *et al.,* (1988) difference-type (6926.661 and 134.072); Shabbir and Gupta (2007) ratio-regression-type estimator (6720.736 and 138.180) and Mishra *et al.,* (2019) ratio estimators Pl_1, Pl_2, Pl_3 and Pl_4 (7554.903 and 122.923), (6926.661 and 134.072), (6684.022 and 134.072) and (6569.176 and 141.368), respectively. The results obtained from dataset II revealed that, the proposed estimators I and II (15684959477 and 1913.339; 7586821411 and 3955.629) have minimum MSE and higher PRE values, respectively, compared to the sample variance (300106473570 and 100); Isaki (1983) classical ratio (29256901796.03 and 1025.763); Isaki (1983) classical regression (27722181291.40 and 1082.550); Singh *et al.,* (1988) difference-type (27429779412.11 and 1094.090); Shabbir and Gupta (2007) ratio-regression-type estimator (26492917739.54 and 1132.780) and Mishra *et al.,* (2019) ratio estimators Pl_1, Pl_2, Pl_3 and Pl_4 (27722181291.40 and 1082.550), (27429779412.11 and 1094.090), (27672258497.21 and 1084.503) and (19355064814.53 and 1550.532), respectively. The results obtained from dataset III similarly revealed that, the proposed estimators I and II (54.2591 and 1028.85; 13.0590 and 4274.79) possessed minimum MSE and higher PRE values, respectively, compared to the sample variance (558.2446 and 100); Isaki (1983) classical ratio (190.8131 and 292.561); Isaki (1983) classical regression (107.935 and 517.205); Singh *et al.,* (1988) difference-type (107.369 and 519.932); Shabbir and Gupta (2007) ratio-regression-type estimator (105.521 and 529.036) and Mishra *et al.,* (2019) ratio estimators Pl_1, Pl_2, Pl_3 and Pl_4 (107.935 and 517.205), (107.369 and 519.932), (107.058 and 521.422) and (99.1400 and 563.085), respectively. The results obtained from dataset IV further revealed that, the proposed estimators I and II (4740125 and 1028.85; 1890131 and 11098.32) possessed minimum MSE and higher PRE values, respectively, compared to the sample variance (209772707 and 100); Isaki (1983) classical ratio (28490643.789 and 736.2863); Isaki (1983) classical regression (18485825.132 and 1134.776); Singh *et al.,* (1988) difference-type (18156904.286 and 1155.333); Shabbir and Gupta (2007) ratio-regression-type estimator (17455922.388 and 1201.728) and Mishra *et al.,* (2019) ratio estimators Pl_1, Pl_2, Pl_3 and Pl_4 (18485825.132 and 1134.776), (18156904.286 and 1155.333), (18319672.456 and 1145.068) and (11136040.182 and 1883.728), respectively. Therefore, based on the criteria of mean square error and percentage relative efficiency, the proposed estimators are more efficient and better.

CONCLUSION

A two-phase estimator with two special classes for estimating finite population variance is proposed in this study. Expressions for bias and MSE of the modified estimators were derived. The theoretical efficiency conditions in which the modified estimators over some of existing estimators were derived. Evidence from the empirical results revealed that the modified estimators performed better than some existing estimators considered based on the criterion of mean square error and percentage relative efficiency. Thus, the study contributed by developing new variance estimators that will provide accurate and reliable estimates of variation for various phenomenon when there is random non-response. Extension of the modified estimator in this study to capture measurement error can be considered in future research.

REFERENCES

Audu, A., Zakari, Y., Yunusa, M. A., Ishaq, O. O., Manu, F., & Muhammad, I. (2023). Regression-Cum-Ratio mean imputation class of estimators using non-conventional robust measures. *Caliphate Journal of Science & Technology (CaJoST),* 3, 246-254.

Cochran, W. G. (1977). *Sampling techniques*. John Wiley and Sons Isaki, C. T. (1983). Variance estimation using auxiliary information. *Journal of American Statistics Association*, 78, 117–123.

Mishra, P., Singh, R. & Khare, S. (2019). Estimation of population variance in log-product type estimators under double sampling scheme. *Journal of Reliability and Statistical Studies,* 12(2), 11-20.

Muhammad, I. (2023). Generalized Ratio-Product cum Regression Variance Estimator in Two-Phase Sampling. *CBN Journal of Applied Statistics*, 14(2): 73-101. DOI: https://doi.org/10.33429/Cjas.14223.4/5

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.

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.

Muhammad, I., Zakari, Y., Abdu, M., Iliyasu, R., Suleiman, M., Manzo, S., Muhammad, A. & Zakar, A. A. (2023). Enhanced ratio-type estimator for finite population mean using auxiliary variable in simple random sampling. *NIPES Journal of Science and Technology Research*, 5(1): 242-252.

Muhammad, I & Oyeyemi, G. M. (2025a). Generalized Variance Estimator using Two Auxiliary Variables under Stratified Random Sampling. *Communication in Physical Sciences*, 12(2), 386-405.

Muhammad, I & Oyeyemi, G. M. (2025b). Generalized Almost Unbiased Estimator of Finite Population Variance under Stratified Random Sampling. *Asian Journal of Probability and Statistics*, 27(7): 159-178.

Murthy, M. N. (1967). *Sampling Theory and Methods*. Statistical Publication Society, Calcutta, OH, US.
Neyman, J. (1938). Contribution to the theory of sampling human populations. *Journal of the American Statistical Association*, 33(201), 101-116.

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.

Shabbir, J. & Gupta, S. (2007). On improvement in variance estimation using auxiliary information. *Communication in Statistical-Theory and Methods*, 36, 2177–2185.

Singh, S. and Joarder, A. H. (1998). Estimation of finite population variance using random nonresponse in survey sampling, *Metrika*, 98, 241–249.

Singh, H. P, Upadhyaya, L. N. & Namjoshi, U. D. (1988). Estimation of finite population variance. *Curr Science*, 57, 1331–1334.

Sukhatme, P. V. & Sukhatme, B. V. (1970). *Sampling Theory of Surveys with Applications*. Ames, IA: Iowa State University Press.

Zakari, Y & Muhammad, I. (2022). Improved product estimator using measurable auxiliary character. Proceedings of the 58th Annual National Conference of Mathematical Association of Nigeria. Published by the *Mathematical Association of Nigeria (MAN)*.

Zakari, Y. & Muhammad, I. (2023). Modified estimator of finite population variance under stratified random sampling. Engineering Proceedings of the 4th International Electronic Conference on Applied Sciences of Science Forum. Published by *Multidisciplinary Digital Publishing Institute (MDPI)*.

Zakari, Y., Muhammad, I., & Sani, N. M. (2020). Alternative ratio-product type estimator in simple random sampling. *Communication in Physical Sciences*, 5(4), 418-426.

Zakari, Y., Yunusa, M. A., Rashida, A., Muhammad, I., Mohammed, A. S. & Audu, A. (2023). The Efficiency of a Ratio Product Estimator in the Estimation of the Finite Population Coefficient of Variation. Engineering Proceedings of the 4th International Electronic Conference on Applied Sciences of Science Forum. Published by *Multidisciplinary Digital Publishing Institute (MDPI)*.