International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

Efficiency Comparison of Modified Regression-Cum-Dual Mean Imputation Schemes for Estimating Population Mean Under Two-Phase Simple Random Sampling Using Simulation Studies

B. B. Bayedo[1], K. E. Lasisi[2], A. Ahmed[3], A. A. Issa[4]
[1,2,3,4]*Department of Statistics, Abubakar Tafawa Balewa University, Bauchi, Nigeria.*

Abstract-- **Auxiliary information has been widely used in survey sampling to improve the precision of estimators of population parameters, particularly when observations are missing. This study proposes modified regression-cum-dual mean imputation schemes for estimating the population mean under two-phase simple random sampling. The proposed estimators integrate the advantages of regression estimation and dual mean imputation to enhance estimation efficiency while minimizing bias and mean squared error (MSE). The theoretical properties of the estimators, including expressions for bias and MSE, are derived under first-order approximation. Their performance is evaluated through extensive Monte Carlo simulation studies conducted across different population settings that represent varying relationships between the study and auxiliary variables. The proposed estimators are compared with the conventional sample mean estimator, the regression imputation estimator, the dual mean imputation estimator, and other existing regression-cum-imputation estimators reported in the literature. Simulation results indicate that the proposed estimators consistently produce lower bias and MSE and achieve higher Percentage Relative Efficiency (PRE) than the competing estimators across the simulation scenarios considered. The findings demonstrate the robustness and superiority of the modified regression-cum-dual mean imputation schemes for handling missing data in two-phase simple random sampling. Consequently, the proposed estimators provide a reliable and efficient alternative for estimating the population means in practical survey sampling applications.**

Keywords-- **Auxiliary information; Bias; Mean squared error; Missing data; Population means; Regression-cum-dual mean imputation; Simulation study; Two-phase simple random sampling.**

I. INTRODUCTION

Sampling theory is the field of statistics that is concerned with the collection, analysis and interpretation of data gathered from sampling the population under study.

The application of sampling theory is concerned not only with the proper selection of observations from the auxiliary variable. Many prominent authors developed several modified and improved ratio, regression, and exponential-type estimators using population information on the auxiliary variable x. However, the information about the population mean of the auxiliary variable is not always available. In the environment mentioned above, the most popular sampling scheme is the two-phase sampling scheme which was first established by Neyman (1938) to accumulate information on sampling. It is customarily acquired when the accumulation of information on a study variable is very costly but relatively cheaper to accumulate information on auxiliary variables that are correlated with the study variables. Due to these reasons, the two-phase sampling becomes a powerful and cost-effective scheme for obtaining the authentic estimate in the one-phase sample for the unknown parameters of the auxiliary variable. Authors like Kumar and Bahl (2006), Singh and Vishwakarma (2007), Singh (2011), Ozgul and Cingi (2014), Kalita *et a*l. (2016), Noor-Al-Amin *et al*. (2016), Bazad and Bazad (2019), Bhushan and Gupta (2019), Adamu *et al* (2019) and Bhushan *et al.* (2023) have extensively worked under two-phase sampling. However, the applicability of the aforementioned estimators depends on the full availability of sample information for both the study and auxiliary variables. Reduction in the sizes of sample information due to non-response decreases the efficiency of these estimators. In literature, Sukhatme (1962), Choudhury and Singh (2015), Audu and Adewara (2017a,b), Adamu *et al.* (2019) investigated the classical ratio estimator in two-phase sampling. Following Srivenkataraman (1980), Kumar and Bahl (2006) envisaged a class of dual-to-exponential-type ratio estimators using two phases. Singh and Vishwakarma (2007) investigated Bahl and Tuteja's (1991) exponential ratio and product estimators of population mean under two-phase sampling.

Singh (2011), Ozgul and Cingi (2014) developed a class of exponential regression cum ratio estimator in two-phase sampling. Kalita *et al.* (2016) proposed exponential ratio-cum-exponential dual-to-ratio estimators based on two-phase sampling. Following Kumar and Bahl (2006) and Kalita *et al.* (2016), Bazad and Bazad (2019) developed some classes of dual-to-ratio exponential-type estimators. Bhushan and Gupta (2019) provided some log-type estimators of population mean using two-phase sampling. Bhushan *et al.* (2021a) suggested some efficient classes of estimators. Bhushan *et al.* (2021b) developed some efficient classes of estimators under two-phase sampling. Recently, Bhushan and Kumar (2023) proposed a new, efficient class of estimators for the population mean using two-phase sampling.

II. Literature Review

2.1 Some Existing Estimators under SRSWOR

The contributions of Cochran (1942), Deming (1956), Hurvitz (1952) and others helped lay the foundation for modern sampling theory. The work done during these periods made important contributions to modern sampling theory by suggesting methods of utilizing the auxiliary information for the purpose of estimating the population mean in order to increase the precision of the estimates.

To discuss some of the developed estimators in literature based on auxiliary variables, the following descriptions about the population and sample units were considered.

Let $\Omega = (1\ 2, 3 \dots N)$be a population of size N andY, X be two real valued functions having values$(Y_i,\ X_i)\in\mathbb{R}^+$ on the i^{th} unit of $U(1 = i = N)$..

Let $\bar{Y}$ and $\bar{X}$ be the population means of Y and X respectively with C_y and C_x as coefficients of variation of Y and X.

Let a pair (y_i, x_i) of simple random sample of size n be drawn without replacement from the population $\Omega = (1\ 2, 3 \dots N)$ and $\bar{y}$, $\bar{x}$ be sample means based on the sample drawn.

The usual sample mean $\bar{y}$, traditional ratio estimator $\bar{y}_r$ (Cochran, 1942), product estimator $\bar{y}_p$ (Murthy, 1964) and Dual to ratio estimator $\bar{y}_{st}$ (Strivenkataramana, 1980) with their respective bias and variance/mean square error are defined as

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad (2.1)$$

$$\operatorname{var}(\bar{y}) = \frac{1-f}{n}\bar{Y}^2 C_y^2 \qquad (2.2)$$

$$\bar{y}_r = \frac{\bar{y}}{\bar{x}}\bar{X} \qquad (2.3)$$

$$Bias(\bar{y}_r) = \bar{Y}\frac{1-f}{n}\left[C_x^2 - \rho_{xy}C_xC_y\right] \qquad (2.4)$$

$$MSE(\bar{y}_r) = \bar{Y}^2\frac{1-f}{n}\left[C_y^2 + C_x^2 - 2\rho_{xy}C_xC_y\right] \qquad (2.5)$$

$$\bar{y}_p = \frac{\bar{y}}{\bar{X}}\bar{x} \qquad (2.6)$$

$$Bias(\bar{y}_p) = \bar{Y}\frac{1-f}{n}\left[C_x^2 + \rho_{xy}C_xC_y\right] \qquad (2.7)$$

$$MSE(\bar{y}_p) = \bar{Y}^2\frac{1-f}{n}\left[C_y^2 + C_x^2 + 2\rho_{xy}C_xC_y\right] \qquad (2.8)$$

$$\bar{y}_{st} = \bar{y}\frac{N\bar{X} - n\bar{x}}{(N-n)\bar{X}} \qquad (2.9)$$

$$Bias(\bar{y}_{st}) = \bar{Y}\frac{1-f}{n}\left[\left(\frac{n}{N-n}\right)^2 C_x^2 - \frac{n}{N-n}\rho_{xy}C_xC_y\right] \qquad (2.10)$$

$$MSE(\bar{y}_{st}) = \bar{Y}^2\frac{1-f}{n}\left[C_y^2 + \left(\frac{n}{N-n}\right)^2 C_x^2 - 2\frac{n}{N-n}\rho_{xy}C_xC_y\right] \qquad (2.11)$$

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

Where $f = \frac{n}{N}$, $f^* = \frac{n}{n_1}$, $f_1 = \frac{n_1}{N}$, $f^{**} = \frac{1-f}{n} + \frac{1-f_1}{n_1}$, $C_{yx} = \rho_{xy} C_y / C_x$, $C_{yz} = \rho_{yz} C_y / C_z, C_{xz} = \rho_{xz} C_x / C_z$.

The traditional ratio estimator $\bar{y}_r$ and the dual to ratio estimator $\bar{y}_{st}$ have higher efficiencies when the correlation between the study and auxiliary variables is strong and positive, while the product estimator $\bar{y}_p$ has higher efficiency when the correlation between the study and auxiliary variables is strong and negative.

For estimating the population mean $\bar{Y}$ of the study variate y, Singh and Espejo (2003) considered a linear combination of ratio and product estimators' type given by

$$\bar{y}_{RP} = \bar{y}\left[k\frac{\bar{X}}{\bar{x}} + (1-k)\frac{\bar{x}}{\bar{X}}\right] \qquad (2.12)$$

and obtained that the estimator attained optimality when $k = \frac{(1+C)}{2}$ with $C = C_y/C_x$, where Cy is the coefficient of variation of y and Cx is the coefficient of variation of x.

Singh and Espejo (2007), modified that when the population mean $\bar{X}$ of x is not known, a first-phase sample of size n_1 should be drawn from the population on which only the x-characteristic is measured in order to furnish a good estimate of $\bar{X}$. Then a second-phase sample of size n is drawn on which both the variables y and x are measured. Let $\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i$ denote the sample mean of x based on the first-phase sample of size n_1, Singh and Espejo (2007) considered a ratio–product type estimator in the two-phase sampling given by

$$\bar{y}_{RP}^{(d)} = \bar{y}\left[k\frac{\bar{x}_1}{\bar{x}} + (1-k)\frac{\bar{x}}{\bar{x}_1}\right] \qquad (2.13)$$

Choudhury and Singh (2012) consider the work of Singh and Espejo (2007) and modified a class chain ratio-product type estimator $\bar{y}_{RP}^{(dc)}$ for estimating population mean $\bar{Y}$ using two auxiliary characters under two conditions in Singh and Espejo (2007). The modified estimator, its bias and MSE are respectively given as

$$\bar{y}_{RP}^{(dc)} = \bar{y}\left[k\frac{\bar{x}_1}{\bar{x}}\frac{\bar{Z}}{\bar{z}_1} + (1-k)\frac{\bar{x}}{\bar{x}_1}\frac{\bar{z}_1}{\bar{Z}}\right] \qquad (2.14)$$

Subramani and Prabavathy (2014) modified two estimators of population mean based on median of study variable using auxiliary information as:

$$\bar{y}_{SP1} = \bar{y}\left[\frac{M_d M + \bar{X}}{M_d m + \bar{X}}\right] \qquad (2.35)$$

` $$\bar{y}_{SP2} = \bar{y}\left[\frac{\bar{X}M + M_d}{\bar{X}m + M_d}\right] \qquad (2.36)$$

where M, M_d and m are the median of the study variable, the median of the auxiliary variable, and the sample median of the study variable, respectively.
The MSEs of modified estimators are given below;

$$MSE(\bar{y}_{SPi}) = \mathrm{var}(\bar{y}) + R_{m}^2\phi_i^2\,\mathrm{var}(m) - 2R_{m}\phi_i\,\mathrm{cov}(\bar{y},m) \quad i = 1,2 \qquad (2.37)$$

where $R_m = \bar{Y}/M$ $\phi_1 = MM_d / (MM_d + \bar{X})$ and $\phi_2 = \bar{X}M / (\bar{X}M + M_d)$.

Singh (2015) modified an improved class of ratio type estimator for finite population mean using unknown weight and power transformation strategies. The particular cases of this estimator are Walsh (1970) estimator at $k = 1$ and $\alpha = w$, *Ray et al.,* (1979) estimator at $k = -1$ and $\alpha = -\gamma$, Srivenkataramana and Tracy (1979) at $k = -1$ and $\alpha = 1 - f$, Srivenkataramana and Tracy (1979) at $k = -1$ and $\alpha = -n/(N-n)$ and Srivenkataramana (1980) at $k = -1$ and $\alpha = 1 - w$. The modified estimator and its properties are given below;

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

$$\bar{y}_{RVK} = \bar{y}\left[\frac{\sqrt{\bar{X}}}{\alpha\sqrt{\bar{x}} + (1-\alpha)\sqrt{\bar{X}}}\right]^{K} \quad (2.38)$$

$$Bias(\bar{y}_{RVK}) = \bar{Y}\frac{1-f}{n}\left[\left(\frac{k(k+1)}{4}\alpha^2 + \frac{1}{8}k\alpha\right)C_x^2 - \frac{1}{2}k\alpha\rho_{xy}C_xC_y\right]$$
(2.39)

$$MSE(\bar{y}_{RVK}) = \bar{Y}^2\frac{1-f}{n}\left[C_y^2 + \frac{1}{4}k^2\alpha^2C_x^2 - k\alpha\rho_{xy}C_xC_y\right]$$
(2.40)

In his study, the modified estimators attain the optimality when $\alpha k = 2\rho_{xy}C_y / C_x$. The empirical study in his work revealed that the modified estimator is more efficient than the estimators of Walsh (1979), Ray *et al.,* (1979), Srivenkataramana and Tracy (1979), Srivenkataramana (1980).

III. Materials And Methods

3.1 Robust Outlier-Free Measure to be used in the Study

The study will utilize robust, outlier-resistant parameter estimation techniques to minimize or eliminate the influence of extreme values in the sample data. These methods include:

i. Gini's Mean Difference method
ii. Downton's Method
iii. The Method of Probability-Weighted Moments

3.1.1 Gini's Mean Method

Let Δ_{xi}, $= X_{(i+1)} - X_{(i)}$, where $X_{(i)}$ is the order statistics so that $\Delta_{xi}, i = 1,\ldots,n-1$ is the distance between adjacent observations. Then Gini's mean method proposed by Nair (1936) is given as

$$G = \frac{4}{n-1}\sum_{i=1}^{n}\left(\frac{2i-n-1}{2n}\right)X_{(i)} \quad (3.1)$$

3.1.2 Downton's Method

Let $X_1, X_2, \ldots, X_n$, be a random sample from a normal distribution with mean μ and variance σ^2; that is, $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$. Let $X_{(1)} \le X_{(2)} \le \ldots X_{(n)}$ denote the corresponding order statistics.

The Downton estimator, D proposed by Downton (1966), is given as

$$D = C\sum_{i=1}^{n}\frac{(2i-n-1)X_{(i)}}{n(n-1)} \quad (3.2)$$

Where $C = \sqrt{\pi} = 1.772453851$ for a normal distribution and does not depend on the sample size n.

3.1.3 Method of probability weighted moments

Let $X_1, X_2, \ldots, X_n$, be a random sample from a normal distribution with mean μ and variance σ^2; that is, $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$.

Let $X_{(1)} \le X_{(2)} \le \ldots X_{(n)}$ denote the corresponding order statistics. The PWMs is defined by Greenwood *et al.* (1979) and is given as

$$S_{pw} = \frac{\sqrt{\pi}}{n^2}\sum_{i=1}^{n}(2i-n-1)X_{(i)} \quad (3.3)$$

3.2 Proposed Imputation Schemes

Having studied the work of Audu *et al.,* (2021b) and Musa et al. (2023), based on the identified weaknesses listed in chapter one, the following two imputation schemes are proposed.

Let Ω be a population with N units, $\Phi \subset \Omega$ with cardinality $|\Phi| = r$ and Φ^c be complement of Φ. Let S_1 be a set of preliminary sample observations of size n_1 $(n_1 < n)$ from which $\bar{x}_1$, the unbiased estimate of $\bar{X}$, the population mean of X, is estimated. Let S_2 be a set of sample of size n_2 $(n_2 < n_1)$ of which only r $(r < n_2)$ elements respond.

$$y_{.i} = \begin{cases} y_i & if \ \mathrm{i} \in R \\ \left\{ n\left(\bar{y}_r + \hat{\beta}_{rg}\left(\bar{x}_1^* - \bar{x}_r\right)\right)\exp\left(\dfrac{H_s\left(1 - \dfrac{\bar{x}_r}{\bar{x}_1^*}\right)}{1 + \left(\left(\phi_j\bar{x}_r + \phi_k\right)/\left(\phi_j\bar{x}_1^* + \phi_k\right)\right)}\right) - r\bar{y}_r \right\}\dfrac{x_i}{\sum_{i \in A^i} x_i} & if \ \mathrm{i} \in R^c \end{cases}$$
(3.5)

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

$$y_{.i} = \begin{cases} y_i & \text{if } i \in R \\ \left\{ n\bar{y}_r \left(\dfrac{\phi_j \bar{x}_1^* + \phi_k}{\phi_j \bar{x}_r + \phi_k} \right) \exp\left(\dfrac{H_s \left(1 - \dfrac{\bar{x}_r}{\bar{x}_1^*} \right)}{1 + \left(\left(\phi_j \bar{x}_r + \phi_k \right) / \left(\phi_j \bar{x}_1^* + \phi_k \right) \right)} \right) - r\bar{y}_r \right\} \dfrac{x_i}{\sum_{i \in A^i} x_i} & \text{if } i \in R^c \end{cases}$$

(3.6)

$$y_{.i} = \begin{cases} y_i & \text{if } i \in R \\ \left\{ n\left(\bar{y}_r + \hat{\beta}_{rg} \left(\bar{x}_1^* - \bar{x}_r \right) \right) \left(\dfrac{\phi_j \bar{x}_1^* + \phi_k}{\phi_j \bar{x}_r + \phi_k} \right) \exp\left(\dfrac{H_s \left(1 - \dfrac{\bar{x}_r}{\bar{x}_1^*} \right)}{1 + \left(\left(\phi_j \bar{x}_r + \phi_k \right) / \left(\phi_j \bar{x}_1^* + \phi_k \right) \right)} \right) - r\bar{y}_r \right\} \dfrac{x_i}{\sum_{i \in A^i} x_i} & \text{if } i \in R^c \end{cases}$$

(3.7)

where $\bar{x}_1^* = \dfrac{n_1 \bar{x}_1 - n\bar{x}}{n_1 - n}, H_s = \phi_j \bar{x}_1^* / \left(\phi_j \bar{x}_1^* + \phi_k \right),$

$\Gamma = \{ (G \times n), (D \times n), (S_{pw} \times n) \}, \quad \phi_j, \phi_k \in \Gamma,$

$a = 1,2,3,4,5,6$. G, D and S_{pw} are the Gini's mean difference, Downtown's method and Probability weighted moments, which are the nonconventional robust measures free from the influence of outliers.

3.2.1 Assumption Of The Proposed Imputation Schemes

The assumptions of proposed imputation schemes are:

i. The size of the studied population is finite, $(N < \infty)$

ii. Characteristic of interest (Y) and auxiliary characteristics (X) are measurable

iii. Correlation between X and Y is strong.

$\bar{y}_r \neq 0, \ \bar{x}_r \neq 0, \ \bar{x}_1 \neq 0$

IV. Data Analysis And Discussion Of Result

To empirically validate the performance of the proposed modified regression-cum-dual mean imputation estimators, four simulated population datasets representing different sampling conditions were generated. The simulation study was designed to evaluate the robustness and efficiency of the proposed estimators under two-phase simple random sampling by considering varying population characteristics and relationships between the study and auxiliary variables.

The proposed estimators were compared with selected existing estimators using three standard performance measures: Bias, Mean Squared Error (MSE), and Percentage Relative Efficiency (PRE).

Bias was used to assess the accuracy of each estimator by measuring the extent to which the estimated population mean deviated from the true population mean. The MSE, which incorporates both the variance and squared bias of an estimator, served as a comprehensive measure of estimation precision. Percentage Relative Efficiency (PRE) was computed to compare the efficiency of the proposed estimators relative to the existing estimators, with higher PRE values indicating superior performance.

The combined use of these evaluation criteria provides a comprehensive framework for assessing estimator performance. The simulation results offer empirical evidence of the effectiveness, robustness, and potential superiority of the proposed modified regression-cum-dual mean imputation schemes for estimating the population mean under two-phase simple random sampling.

4.1 Populations used for Simulation Study for positive relationship

Table 1: Biases, MSEs and PREs of Existing Estimators and Proposed Estimators of the proposed schemes 1, 2 and 3 using N=2000, n_1=800, n=300, r=200 (Case 1 & 2)

Estimators	Case 1			Case 2		
	Biases	MSEs	PREs	Biases	MSEs	PREs
Sample mean T0	0.00974	0.029	100.0000	0.00016	0.0298	100
Lee et al. (1994) T1	-0.00107	0.0102	284.2900	-0.00471	0.01019	292.3500
Singh & Horn (2000) T2	-0.00265	0.00994	291.8400	-0.00543	0.01088	273.8100
Kadilar & Cingi (2008) T3	0.01267	0.02634	110.1000	0.02825	0.04539	65.6500
Singh (2009) T4	0.00133	0.01002	289.5600	0.00058	0.01134	262.8300
Audu et al. (2021b) T5	-0.02768	0.01163	249.4300	-0.04152	0.01544	192.9900
Musa et al. (2023)						
T6 *(a=Cx,b=Sx)*	-1.08465	1.25058	2.3200	-1.09581	1.26833	2.3500
T6 *(a=Cx,b=Skw)*	-1.18521	1.47997	1.9600	-1.19667	1.50113	1.9800
T6 *(a=Cx,b=Kurt)*	-1.20351	1.52412	1.9000	-1.2151	1.5463	1.9300
T6 *(a=Sx,b=Cx)*	-1.2069	1.5324	1.8900	-1.21852	1.55478	1.9200
T6 *(a=Sx,b=Skw)*	-1.20694	1.53248	1.8900	-1.21855	1.55486	1.9200
T6 *(a=Sx,b=Kurt)*	-1.20916	1.53792	1.8900	-1.22079	1.56043	1.9100
T6 *(a=Skw,b=Sx)*	-1.08394	1.24905	2.3200	-1.09511	1.26679	2.3500
T6 *(a=Skw,b=Cx)*	-1.1847	1.47874	1.9600	-1.19615	1.49987	1.9900
T6 *(a=Skw,b=Kurt)*	-1.20344	1.52395	1.9000	-1.21502	1.54612	1.9300
T6 *(a=Kurt,b=Cx)*	-1.13035	1.35207	2.1400	-1.14157	1.37098	2.1700
T6 *(a=Kurt,b=Sx)*	-1.00982	1.09422	2.6500	-1.0211	1.11102	2.6800
T6 *(a=Kurt,b=Skw)*	-1.13097	1.35346	2.1400	-1.14219	1.3724	2.1700
Estimators of Proposed Scheme 1						
Tp11 (G,D)	-0.00293	0.00959	302.3200	-0.00521	0.00993	300.0000
Tp12 (G,S)	-0.00311	0.00965	300.6500	-0.005	0.01024	290.9600
Tp13 (D,G)	-0.00323	0.00972	298.4500	-0.00479	0.01053	282.8800
Tp14 (D,S)	-0.00328	0.00976	297.2700	-0.00468	0.01067	279.2400
Tp15 (S,G)	-0.00303	0.00962	301.6000	-0.0051	0.01009	295.4400
Tp16 (S,D)	-0.0029	0.00959	302.4800	-0.00524	0.00989	301.3200
Estimators of Proposed Scheme 2						
Tp21 (G,D)	0.01144	0.03097	93.6300	0.00188	0.03217	92.6300
Tp22 (G,S)	0.01264	0.0324	89.5200	0.00312	0.0339	87.9000
Tp23 (D,G)	0.01363	0.0336	86.3000	0.00416	0.03538	84.2300
Tp24 (D,S)	0.01407	0.03413	84.9600	0.00461	0.03603	82.6900
Tp25 (S,G)	0.01206	0.0317	91.4700	0.00252	0.03305	90.1500
Tp26 (S,D)	0.01126	0.03075	94.2900	0.0017	0.0319	93.3900
Estimators of Proposed Scheme 3						
Tp31 (G,D)	-0.00219	0.00965	300.5000	-0.00558	0.00934	319.0300
Tp32 (G,S)	-0.0019	0.00974	297.6400	-0.00564	0.00924	322.4900
Tp33 (D,G)	-0.00165	0.00985	294.5000	-0.00566	0.0092	323.9800
Tp34 (D,S)	-0.00154	0.0099	292.9400	-0.00566	0.00919	324.2300
Tp35 (S,G)	-0.00204	0.00969	299.1500	-0.00561	0.00928	321.0400
Tp36 (S,D)	-0.00224	0.00964	300.8400	-0.00557	0.00936	318.3400

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

Table 2: Biases, MSEs and PREs of Existing Estimators and Proposed Estimators of the proposed schemes 1, 2 and 3 using N=2000, n_1=1000, n=500, r=300 (Case 1 & 2)

Estimators	Case 1			Case 2		
	Biases	MSEs	PREs	Biases	MSEs	PREs
Sample mean T0	-0.00227	0.01853	100	0.00277	0.01875	100
Lee et al. (1994) T1	-0.00414	0.00423	438.36	-0.00157	0.00645	290.86
Singh & Horn (2000) T2	-0.00441	0.00407	454.81	-0.00221	0.00685	273.55
Kadilar & Cingi (2008) T3	0.01484	0.01824	101.6	0.01666	0.02677	70.04
Singh (2009) T4	-0.00115	0.00411	450.69	0.00135	0.00699	268.38
Audu et al. (2021b) T5	-0.02433	0.00531	348.68	-0.0247	0.00846	221.71
Musa et al. (2023)						
T6 *(a=Cx,b=Sx)*	-0.99812	1.02416	1.81	-0.97778	0.99538	1.88
T6 *(a=Cx,b=Skw)*	-1.12083	1.28358	1.44	-1.1005	1.25366	1.5
T6 *(a=Cx,b=Kurt)*	-1.14291	1.33356	1.39	-1.12257	1.30354	1.44
T6 *(a=Sx,b=Cx)*	-1.14699	1.34293	1.38	-1.12666	1.31289	1.43
T6 *(a=Sx,b=Skw)*	-1.14703	1.34303	1.38	-1.1267	1.31299	1.43
T6 *(a=Sx,b=Kurt)*	-1.14971	1.34918	1.37	-1.12938	1.31913	1.42
T6 *(a=Skw,b=Sx)*	-0.99725	1.02243	1.81	-0.97691	0.99367	1.89
T6 *(a=Skw,b=Cx)*	-1.12021	1.28219	1.45	-1.09988	1.25227	1.5
T6 *(a=Skw,b=Kurt)*	-1.14282	1.33337	1.39	-1.12249	1.30334	1.44
T6 *(a=Kurt,b=Cx)*	-1.05418	1.13884	1.63	-1.03384	1.10943	1.69
T6 *(a=Kurt,b=Sx)*	-0.90522	0.84804	2.18	-0.88493	0.82079	2.28
T6 *(a=Kurt,b=Skw)*	-1.05493	1.14042	1.62	-1.03459	1.111	1.69
Estimators of Proposed Scheme 1						
Tp11 (G,D)	-0.00472	0.00392	472.18	-0.00221	0.00697	268.78
Tp12 (G,S)	-0.00462	0.00402	460.55	-0.0021	0.00726	258.38
Tp13 (D,G)	-0.00452	0.00414	447.8	-0.00198	0.00752	249.15
Tp14 (D,S)	-0.00446	0.0042	441.54	-0.00193	0.00765	245.03
Tp15 (S,G)	-0.00467	0.00397	466.73	-0.00216	0.00711	263.53
Tp16 (S,D)	-0.00473	0.00391	473.57	-0.00223	0.00694	270.31
Estimators of Proposed Scheme 2						
Tp21 (G,D)	-0.00096	0.0203	91.26	0.00419	0.02055	91.22
Tp22 (G,S)	-0.00002	0.0216	85.78	0.00524	0.02188	85.67
Tp23 (D,G)	0.00079	0.02271	81.58	0.00614	0.02303	81.4
Tp24 (D,S)	0.00114	0.0232	79.85	0.00654	0.02354	79.63
Tp25 (S,G)	-0.00048	0.02097	88.37	0.00473	0.02123	88.3
Tp26 (S,D)	-0.00111	0.02011	92.15	0.00404	0.02035	92.12
Estimators of Proposed Scheme 3						
Tp31 (G,D)	-0.00481	0.00389	476.51	-0.00237	0.00645	290.58
Tp32 (G,S)	-0.00477	0.00396	467.51	-0.00233	0.00637	294.16
Tp33 (D,G)	-0.00471	0.00406	456.56	-0.00226	0.00635	295.35
Tp34 (D,S)	-0.00468	0.00411	450.95	-0.00222	0.00635	295.36
Tp35 (S,G)	-0.00479	0.00392	472.48	-0.00235	0.0064	292.72
Tp36 (S,D)	-0.00481	0.00388	477.45	-0.00237	0.00647	289.82

4. 2 Interpretation of the Results

The simulation results were interpreted in line with the primary objectives of the proposed modified regression-cum-dual mean imputation schemes: to reduce estimation bias, minimize Mean Squared Error (MSE), and improve the efficiency of population mean estimation under two-phase simple random sampling. To evaluate these objectives, extensive Monte Carlo simulation experiments were conducted under different sampling configurations and population characteristics. The performance of the proposed estimators was assessed using three standard measures—Bias, MSE, and Percentage Relative Efficiency (PRE)—which respectively measure estimation accuracy, precision, and relative efficiency.

The simulation results, presented in Tables 1 and 2 for Cases 1 and 2, compare the proposed estimators with existing estimators developed by Lee *et al.* (1994) T_1^*, Singh and Horn (2000) T_2^*, Kadilar and Cingi (2008) T_3^*, Singh (2009) T_4^*, Audu *et al.* (2021b) $T_{5(j)}^*$, $j=1,2,3,4$ and Musa *et al.* (2023) $T_{6(j)}^*$, $j=1,2,3,4$ ..

The comparative analysis shows that estimators with smaller absolute Bias values are more accurate, lower MSE values indicate greater precision, and higher PRE values reflect superior efficiency. Accordingly, the interpretation focuses on these performance measures to determine the extent to which the proposed modifications outperform the existing estimators and achieve their intended objective of improving population mean estimation under two-phase simple random sampling.

V. CONCLUSION

This study developed a class of modified regression-cum-dual mean imputation estimators for estimating population means under two-phase simple random sampling by effectively incorporating auxiliary information into the imputation process. The primary objective was to improve the accuracy and efficiency of population mean estimation in the presence of missing observations. The statistical properties of the proposed estimators, particularly their Biases and Mean Squared Errors (MSEs), were derived up to the third-order approximation using Taylor series expansion, providing a rigorous theoretical framework for evaluating their performance and establishing efficiency conditions relative to existing estimators.

The theoretical findings were validated through extensive Monte Carlo simulation studies conducted under different population structures and sampling configurations. The results consistently showed that the proposed estimators produced lower Biases, smaller MSEs, and higher Percentage Relative Efficiencies (PREs) than the competing estimators across all simulation scenarios. These findings confirm the robustness, stability, and superior performance of the proposed estimators under varying survey conditions.

Overall, the study demonstrates that the modified regression-cum-dual mean imputation schemes provide more accurate, precise, and efficient estimates of the population mean than existing methods. The proposed estimators represent a significant methodological contribution to survey sampling and offer a practical approach for handling missing data in socio-economic, agricultural, health, and official statistical surveys where auxiliary information is available.

REFERENCES

[1] Adejumobi, A., Audu, A., Yunusa, M. A. and Singh, R. V. K. (2022). Efficiency of Modified Generalized Imputation Scheme for Estimating Population Mean with Known Auxiliary Information. *Bayero Journal of Pure and Applied Sciences*, 15(1): 105-112

[2] Ahmed, M.S., Al-Titi, O., Al-Rawi and Abu-Dayyeh, W. (2006). Estimation of a population mean using different imputation methods. *Statistics in Transition,* 7(6): 1247 -1264.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

[3] Al-Omari, A. I., Bouza, C. N. and Herrera, C. (2013). Imputation methods of missing data for estimating the population mean using simple random sampling with known correlation coefficient. *Quality and Quantity*, 47, 353-365.

[4] Audu, A., Ishaq, O. O., Singh, R. V. K., Danbaba, A. and Manu, F. (2023a). On the study of efficiency of exponential-type estimator of population mean using robust regression methods. *Quality and Reliability Engineering International*, 39,190-205.

[5] Audu, A., Singh, R. and Khare, S. (2023b): New Regression-Type Compromised Imputation Class of Estimators with known Parameters of Auxiliary Variable. *Communication in Statistics-Simulation and Computation*, 52:10, 4789-4801.

[6] Audu, A., Ishaq, O.O., Isah, U., Muhammed, S., Akintola, K. A., A. Rashida, A., and Abubakar. A. (2020a) On the Class of Exponential-Type Imputation Estimators of Population Mean with Known Population Mean of Auxiliary Variable**.** *NIPES Journal of Science and Technology Research* 2(4) 2020 pp.1 - 11

[7] Audu, A., Ishaq O. O., Zakari, Y., Wisdom, D. D., Muili, J. and Ndatsu. A. M. (2020b): Regression-cum-exponential ratio imputation class of estimators of population mean in the presence of non-response. *Science Forum Journal of Pure and Applied Science*, 20, 58-63

[8] Barnard, J. and Meng, X. L. (1999). "Applications of multiple imputation in medical studies: from AIDS to NHANES". *Statistical Methods in Medical Research*. **8**(1): 17-36. doi:10.1177/096228029900800103.

[9] Bijoy K. H. (2012). A class of regression-cum-ratio estimators in two-phase sampling for utilizing information from high resolution satellite data**.** *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1, 71-76

[10] Chand, L. (1975). Some ratio type estimator based on two or more auxiliary variables. Ph.D. dissertation. Lowa State University. Ames, Lowa (unpublished).

[11] Chanu, W. W. and Singh, B. K. (2014). Improved Class of Ratio - Cum- Product Estimators of Finite Population Mean in two Phase Sampling. *Global Journal of Science Frontier Research: Mathematics and Decision Sciences,* 14(2), 68-81

[12] Choudhury S. and Singh B. K. (2012): A class of chain ratio-product type estimators with two auxiliary variables under double sampling scheme. *Jour. Of the Korean Stat. Soc*. 41, 247-256.

[13] Choudhury, S. and Singh, B. (2011). An improved class of ratio-cum-product estimator of finite population mean in simple random sampling. *International Journal of Statistics and Analysis*, 1(4):393–403.

[14] Cochran, W. G. (1942): Sampling Theory when the Sampling Units are of Unequal Sizes. *Journ. American Statistical Association*. 37: 199-212.

[15] Cochran, W. G. (1940): The estimation of the yield of cereal experiments by sampling for the ratio gain to total produce. *Journal of Agriculture Society,* 30, 262-275.

[16] Deming, W. E. (1956): *Sample Design in Business Research*. New York: John Wiley and Sons.

[17] Diana, G. and Perri, P. F. (2010). Improved estimators of the population mean for missing data. *Communications in Statistics-Theory and Methods*, Vol. 39, pp.3245–3251.

[18] Downton, F. (1966). Linear estimates with polynomial coefficients. *Biometrika,* 53:129–141.

[19] Enang, E. I., Akpan, V. M. and Ekpenyong, E. J. (2014): Alternative Ratio Estimator of Population Mean in Simple Random Sampling. *Journal of Mathematics Research*; 6(3), 54-61

[20] Gelman, A. and Jennifer H, (2006). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Ch.25.

[21] Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979). Probability-weighted moments: definition and relation to parameters of several distributions expressible in inverse form, *Water Resources Research*, 15, 1049–1054.

[22] Hansen , M. N. and Hurwitz, W. N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41:517–529.

[23] Hurvitz, D. G. (1952): Sampling and Field procedures of the Pittsburg Morbidity Survey. *Public Health Reports*. 67: 1003-1012.

[24] Kadilar, C and Cingi, H. (2006). Improvement in estimating the population mean in simple random sampling, *Applied Mathematics Letters*, 19, 75-79.

[25] Kadilar, C. and Cingi, H. (2004). New Ratio Estimators using Correlation Coefficient. *Interstat 4*:1-11

[26] Kalton, G. (1986) The treatment of Missing Survey Data, *Survey Methodology,* 12: 1-6

[27] Khan, H., Shouket, S., and Sanaullah, A. (2012). A Class of Improved Estimator for Estimating Population Mean Regarding Partial Information in Double Sampling. *Global Journal of Science Frontier Research, Mathematics and Decision Sciences*, 12 (14): 32-45.

[28] Khan, M. and Hussain, S. (2015). An efficient class of ratio estimators of finite population mean under simple random sampling, *Sci.Int.(Lahore),* 22(6), 4933-4937.

[29] Lee, H., Rancourt, E., and Sarndal, C.E. (1994). Experiments with variance estimation from survey data with imputed value. *Journal of Official Statistics, 10(3), 231-243*

[30] Malik, K. and Tailor, R. (2013). Ratio Type Estimator of Population Mean in Double Sampling. *International Journal of Advanced Mathematics and Statistics*. 1 (1): 34- 39