



# Science Forum (Journal of Pure and Applied Sciences)



Journal Homepage: <https://atbuscienceforum.com.ng/index.php/jpas>

## Modified Regression-Cum-Dual Mean Imputation Schemes for Estimating Population Mean Under Two-Phase Simple Random Sampling

**B.B.Bayedo, K.E.Lasisi, A.Ahmed, A.A.Issa**

*Department of Statistics, Abubakar Tafawa Balewa University, Bauchi, Nigeria.*

### Abstract

In sample survey methodology, the incorporation of auxiliary information at the estimation stage is widely recognized as an effective strategy for improving the precision of population parameter estimates. This study develops modified regression-cum-dual mean imputation estimators for the estimation of the population mean under a two-phase simple random sampling (SRS) framework. The proposed estimators integrate auxiliary variables within a regression-assisted imputation structure, combined with dual estimation principles, to enhance efficiency in the presence of incomplete data. The methodological formulation considers two distinct cases within the two-phase sampling design, allowing for flexibility in handling varying patterns of auxiliary information and missing observations. The proposed estimators are derived analytically, and their statistical properties are examined, with particular emphasis on bias and mean square error (MSE) as measures of estimator performance. Efficiency comparisons are conducted against selected existing estimators commonly applied in similar sampling contexts. The results demonstrate that the modified regression-cum-dual mean imputation estimators consistently achieve lower mean square errors relative to the competing estimators across both cases considered, indicating superior efficiency and improved precision in estimating the population mean. These findings underscore the robustness of the proposed approach and its suitability for practical applications in survey sampling where auxiliary information is available and missing data issues are present. Overall, the study contributes to the advancement of imputation-based estimation techniques by providing a more efficient class of estimators within the two-phase SRS framework.

### Keywords

Auxiliary variable,  
Population Mean,  
Mean Squared Errors (MSE),  
Bias.

ISSN 1119-4618



1119 4618

JPAS



— Faculty of Science, ATBU Bauchi —

## 1.0 Introduction

Sampling theory constitutes a fundamental domain of statistical inference concerned with the systematic design, collection, analysis, and interpretation of data obtained from a subset (sample) of a target population. Its application extends beyond mere selection of observational units to the efficient incorporation of auxiliary information for improving estimator precision. In this context, considerable attention has been devoted to the development of modified and improved classes of estimators—particularly ratio, regression, and exponential-type estimators—through the utilization of known population parameters of an auxiliary variable ( $x$ ).

Notwithstanding these advances, a critical limitation arises in practical survey situations where complete population information on the auxiliary variable, such as its mean, is unavailable. Under such circumstances, two-phase (double) sampling provides a robust methodological framework. Originally formalized by Jerzy Neyman (1938), this design facilitates the preliminary estimation of auxiliary information in the first phase, followed by the collection of data on the primary study variable in a second phase. This approach is particularly advantageous in scenarios where measurement of the study variable is expensive or logistically demanding, while auxiliary variables are relatively inexpensive to observe and exhibit strong correlation with the study variable.

Consequently, two-phase sampling emerges as a cost-effective and statistically efficient strategy for obtaining reliable estimates of

population parameters when auxiliary information is partially unknown.

A substantial body of literature has contributed to the advancement of estimator development within the two-phase sampling paradigm. Notable contributions include those of Kumar and Bahl (2006), Singh and Vishwakarma (2007), Singh (2011), Ozgul and Cingi (2014), Kalita et al. (2016), Noor-Al-Amin et al. (2016), Bazad and Bazad (2019), Bhushan and Gupta (2019), Adamu et al. (2019), and Bhushan et al. (2023), among others, who have proposed diverse classes of estimators leveraging auxiliary information under two-phase sampling schemes.

However, the practical applicability and efficiency of these estimators are contingent upon the availability of complete and reliable sample information on both the study and auxiliary variables. In real-world surveys, non-response is a recurrent challenge that reduces effective sample size and consequently degrades estimator efficiency. To address this issue, several studies have explored estimator performance under non-response conditions within two-phase sampling. For instance, Sukhatme (1962), Choudhury and Singh (2015), Audu and Adewara (2017a, 2017b), and Adamu et al. (2019) examined classical ratio estimators under such conditions.

Further methodological refinements include the development of dual and exponential-type estimators. Following the foundational work of Srivenkataraman (1980), Kumar and Bahl (2006) proposed a class of dual-to-exponential ratio estimators within a two-phase framework. Singh and Vishwakarma (2007) extended the exponential ratio and

product estimators originally introduced by Bahl and Tuteja (1991). Subsequently, Singh (2011) and Ozgul and Cingi (2014) developed exponential regression-cum-ratio estimators. Kalita et al. (2016) introduced exponential ratio-cum dual-to-ratio estimators, while Bazad and Bazad (2019) further refined dual-to-ratio exponential estimators. Additionally, Bhushan and Gupta (2019) proposed log-type estimators, and Zaman and Kadilar (2021a, 2021b) introduced new classes of exponential estimators for finite population means. More recently, Bhushan et al. (2021a, 2021b) and Bhushan and Kumar (2023) have contributed additional efficient classes of estimators within the two-phase sampling framework.

Overall, despite the extensive methodological developments, challenges related to incomplete auxiliary information and non-response persist, thereby necessitating the continued development of more robust, efficient, and practically applicable estimators under two-phase sampling designs.

## 2.0 LITERATURE REVIEW

### 2.1 Some Existing Estimators under SRSWOR

The contributions of Cochran (1942), Deming (1956), Hurvitz (1952) and others helped in

laying the foundation of modern sampling theory. The work done during these periods made important contribution to the modern sampling theory by suggesting methods of utilizing the auxiliary information for the purpose of estimation of 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  $N$  and  $Y, X$  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

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

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

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

$$\text{Bias}(\bar{y}_r) = \bar{Y} \frac{1-f}{n} [C_x^2 - \rho_{xy} C_x C_y] \quad (2.4)$$

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

$$\bar{y}_p = \frac{\bar{y}}{\bar{X}} \bar{x} \tag{2.6}$$

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

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

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

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

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

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 Dual to ratio estimator  $\bar{y}_{st}$  have higher efficiencies when the correlation between the study and auxiliary variables is strong and positive while 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

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

and obtained that estimator attained optimality when  $k = \frac{(1+C)}{2}$  with

$C = C_y / C_x$ , where  $C_y$  is the coefficient of variation of  $y$  and  $C_x$  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

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

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

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

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

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

$$\bar{y}_{SP2} = \bar{y} \left[ \frac{\bar{X}M + M_d}{\bar{X}m + M_d} \right] \quad (2.36) \text{ where } M, M_d \text{ and } m \text{ are Median of the study variable,}$$

Median of the auxiliary variable and Sample median of study variable respectively.

The MSEs of modified estimators are given below;

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

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

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;

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

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

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

In his study, the modified estimators attain the optimality when  $\alpha k = 2\rho_{xy} C_y / C_x$ . The empirical study of 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).

### 3.0 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  $\Delta_{xi} = X_{(i+1)} - X_{(i)}$ , where  $X_{(i)}$  is the order statistics so that  $\Delta_{xi}, i = 1, \dots, n-1$  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  $X_1, X_2, \dots, X_n$ , be a random sample from a normal distribution with mean  $\mu$  and variance  $\sigma^2$ ; that is,  $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$ . Let  $X_{(1)} \leq X_{(2)} \leq \dots X_{(n)}$  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  $X_1, X_2, \dots, X_n$ , be a random sample from a normal distribution with mean  $\mu$  and variance  $\sigma^2$ ; that is,  $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$ . Let  $X_{(1)} \leq X_{(2)} \leq \dots X_{(n)}$  denote the corresponding order statistics. The PWMs is defined by Greenwood et al. (1979) and is given as

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

## 4.0 Data Analysis and Discussion of result

For empirical validation, four simulated population datasets with varying sampling conditions were generated to assess the performance of the proposed estimators within Survey Sampling. Their efficiency was evaluated against existing estimators using Bias, Mean Square Error (MSE), and Percentage Relative Efficiency (PRE), which

measure accuracy, variability, and relative performance in estimating the population mean. Bias and MSE assessed closeness to true parameters and estimation precision, while PRE enabled comparative efficiency analysis, providing a systematic framework that demonstrates the effectiveness and potential superiority of the proposed estimators within Statistics.

#### 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 $T_0$	0.00974	0.029	100.0000	0.00016	0.0298	100
Lee et al. (1994) $T_1$	-0.00107	0.0102	284.2900	-0.00471	0.01019	292.3500
Singh & Horn (2000) $T_2$	-0.00265	0.00994	291.8400	-0.00543	0.01088	273.8100
Kadilar & Cingi (2008) $T_3$	0.01267	0.02634	110.1000	0.02825	0.04539	65.6500
Singh (2009) $T_4$	0.00133	0.01002	289.5600	0.00058	0.01134	262.8300
Audu et al. (2021b) $T_5$	-0.02768	0.01163	249.4300	-0.04152	0.01544	192.9900
<b>Musa et al. (2023)</b>						
$T_6$ ( $a=Cx, b=Sx$ )	-1.08465	1.25058	2.3200	-1.09581	1.26833	2.3500
$T_6$ ( $a=Cx, b=Skw$ )	-1.18521	1.47997	1.9600	-1.19667	1.50113	1.9800
$T_6$ ( $a=Cx, b=Kurt$ )	-1.20351	1.52412	1.9000	-1.2151	1.5463	1.9300
$T_6$ ( $a=Sx, b=Cx$ )	-1.2069	1.5324	1.8900	-1.21852	1.55478	1.9200
$T_6$ ( $a=Sx, b=Skw$ )	-1.20694	1.53248	1.8900	-1.21855	1.55486	1.9200
$T_6$ ( $a=Sx, b=Kurt$ )	-1.20916	1.53792	1.8900	-1.22079	1.56043	1.9100
$T_6$ ( $a=Skw, b=Sx$ )	-1.08394	1.24905	2.3200	-1.09511	1.26679	2.3500
$T_6$ ( $a=Skw, b=Cx$ )	-1.1847	1.47874	1.9600	-1.19615	1.49987	1.9900
$T_6$ ( $a=Skw, b=Kurt$ )	-1.20344	1.52395	1.9000	-1.21502	1.54612	1.9300
$T_6$ ( $a=Kurt, b=Cx$ )	-1.13035	1.35207	2.1400	-1.14157	1.37098	2.1700
$T_6$ ( $a=Kurt, b=Sx$ )	-1.00982	1.09422	2.6500	-1.0211	1.11102	2.6800
$T_6$ ( $a=Kurt, b=Skw$ )	-1.13097	1.35346	2.1400	-1.14219	1.3724	2.1700
<b>Estimators of Proposed Scheme 1</b>						
$T_{p11}$ ( $G, D$ )	-0.00293	0.00959	302.3200	-0.00521	0.00993	300.0000
$T_{p12}$ ( $G, S$ )	-0.00311	0.00965	300.6500	-0.005	0.01024	290.9600
$T_{p13}$ ( $D, G$ )	-0.00323	0.00972	298.4500	-0.00479	0.01053	282.8800
$T_{p14}$ ( $D, S$ )	-0.00328	0.00976	297.2700	-0.00468	0.01067	279.2400
$T_{p15}$ ( $S, G$ )	-0.00303	0.00962	301.6000	-0.0051	0.01009	295.4400
$T_{p16}$ ( $S, D$ )	-0.0029	0.00959	302.4800	-0.00524	0.00989	301.3200
<b>Estimators of Proposed Scheme 2</b>						
$T_{p21}$ ( $G, D$ )	0.01144	0.03097	93.6300	0.00188	0.03217	92.6300
$T_{p22}$ ( $G, S$ )	0.01264	0.0324	89.5200	0.00312	0.0339	87.9000
$T_{p23}$ ( $D, G$ )	0.01363	0.0336	86.3000	0.00416	0.03538	84.2300
$T_{p24}$ ( $D, S$ )	0.01407	0.03413	84.9600	0.00461	0.03603	82.6900
$T_{p25}$ ( $S, G$ )	0.01206	0.0317	91.4700	0.00252	0.03305	90.1500
$T_{p26}$ ( $S, D$ )	0.01126	0.03075	94.2900	0.0017	0.0319	93.3900
<b>Estimators of Proposed Scheme 3</b>						
$T_{p31}$ ( $G, D$ )	-0.00219	0.00965	300.5000	-0.00558	0.00934	319.0300
$T_{p32}$ ( $G, S$ )	-0.0019	0.00974	297.6400	-0.00564	0.00924	322.4900
$T_{p33}$ ( $D, G$ )	-0.00165	0.00985	294.5000	-0.00566	0.0092	323.9800
$T_{p34}$ ( $D, S$ )	-0.00154	0.0099	292.9400	-0.00566	0.00919	324.2300
$T_{p35}$ ( $S, G$ )	-0.00204	0.00969	299.1500	-0.00561	0.00928	321.0400
$T_{p36}$ ( $S, D$ )	-0.00224	0.00964	300.8400	-0.00557	0.00936	318.3400

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 $T_0$	-0.00227	0.01853	100	0.00277	0.01875	100
Lee et al. (1994) $T_1$	-0.00414	0.00423	438.36	-0.00157	0.00645	290.86
Singh & Horn (2000) $T_2$	-0.00441	0.00407	454.81	-0.00221	0.00685	273.55
Kadilar & Cingi (2008) $T_3$	0.01484	0.01824	101.6	0.01666	0.02677	70.04
Singh (2009) $T_4$	-0.00115	0.00411	450.69	0.00135	0.00699	268.38
Audu et al. (2021b) $T_5$	-0.02433	0.00531	348.68	-0.0247	0.00846	221.71
<b>Musa et al. (2023)</b>						
$T_6$ ( $a=Cx, b=Sx$ )	-0.99812	1.02416	1.81	-0.97778	0.99538	1.88
$T_6$ ( $a=Cx, b=Skw$ )	-1.12083	1.28358	1.44	-1.1005	1.25366	1.5
$T_6$ ( $a=Cx, b=Kurt$ )	-1.14291	1.33356	1.39	-1.12257	1.30354	1.44
$T_6$ ( $a=Sx, b=Cx$ )	-1.14699	1.34293	1.38	-1.12666	1.31289	1.43
$T_6$ ( $a=Sx, b=Skw$ )	-1.14703	1.34303	1.38	-1.1267	1.31299	1.43
$T_6$ ( $a=Sx, b=Kurt$ )	-1.14971	1.34918	1.37	-1.12938	1.31913	1.42
$T_6$ ( $a=Skw, b=Sx$ )	-0.99725	1.02243	1.81	-0.97691	0.99367	1.89
$T_6$ ( $a=Skw, b=Cx$ )	-1.12021	1.28219	1.45	-1.09988	1.25227	1.5
$T_6$ ( $a=Skw, b=Kurt$ )	-1.14282	1.33337	1.39	-1.12249	1.30334	1.44
$T_6$ ( $a=Kurt, b=Cx$ )	-1.05418	1.13884	1.63	-1.03384	1.10943	1.69
$T_6$ ( $a=Kurt, b=Sx$ )	-0.90522	0.84804	2.18	-0.88493	0.82079	2.28
$T_6$ ( $a=Kurt, b=Skw$ )	-1.05493	1.14042	1.62	-1.03459	1.111	1.69
<b>Estimators of Proposed Scheme 1</b>						
$T_{p11}$ ( $G, D$ )	-0.00472	0.00392	472.18	-0.00221	0.00697	268.78
$T_{p12}$ ( $G, S$ )	-0.00462	0.00402	460.55	-0.0021	0.00726	258.38
$T_{p13}$ ( $D, G$ )	-0.00452	0.00414	447.8	-0.00198	0.00752	249.15
$T_{p14}$ ( $D, S$ )	-0.00446	0.0042	441.54	-0.00193	0.00765	245.03
$T_{p15}$ ( $S, G$ )	-0.00467	0.00397	466.73	-0.00216	0.00711	263.53
$T_{p16}$ ( $S, D$ )	-0.00473	0.00391	473.57	-0.00223	0.00694	270.31
<b>Estimators of Proposed Scheme 2</b>						
$T_{p21}$ ( $G, D$ )	-0.00096	0.0203	91.26	0.00419	0.02055	91.22
$T_{p22}$ ( $G, S$ )	-0.00002	0.0216	85.78	0.00524	0.02188	85.67
$T_{p23}$ ( $D, G$ )	0.00079	0.02271	81.58	0.00614	0.02303	81.4
$T_{p24}$ ( $D, S$ )	0.00114	0.0232	79.85	0.00654	0.02354	79.63
$T_{p25}$ ( $S, G$ )	-0.00048	0.02097	88.37	0.00473	0.02123	88.3
$T_{p26}$ ( $S, D$ )	-0.00111	0.02011	92.15	0.00404	0.02035	92.12
<b>Estimators of Proposed Scheme 3</b>						
$T_{p31}$ ( $G, D$ )	-0.00481	0.00389	476.51	-0.00237	0.00645	290.58
$T_{p32}$ ( $G, S$ )	-0.00477	0.00396	467.51	-0.00233	0.00637	294.16
$T_{p33}$ ( $D, G$ )	-0.00471	0.00406	456.56	-0.00226	0.00635	295.35
$T_{p34}$ ( $D, S$ )	-0.00468	0.00411	450.95	-0.00222	0.00635	295.36
$T_{p35}$ ( $S, G$ )	-0.00479	0.00392	472.48	-0.00235	0.0064	292.72
$T_{p36}$ ( $S, D$ )	-0.00481	0.00388	477.45	-0.00237	0.00647	289.82

## 4.2 Interpretation of the Results

The interpretation of the simulation outcomes is grounded in the core objectives of the proposed methodological framework, namely: reduction of estimator bias, minimization of Mean Square Error (MSE), and enhancement of estimation efficiency for the population mean within survey sampling. To achieve a rigorous and comprehensive evaluation, Monte Carlo simulations were conducted under alternative sampling configurations that mimic realistic survey conditions, particularly scenarios characterized by incomplete response patterns and partial auxiliary information.

The simulation results, summarized in Tables 1-2 for Cases 1 and 2 respectively, present the computed values of Bias, MSE, and Percentage Relative Efficiency (PRE) for the proposed class of regression-cum-dual mean imputation estimators. These are systematically compared with corresponding values from established estimators developed by Lee et al. (1994), Singh and Horn (2000), Kadilar and Cingi (2008), Singh (2009), Audu et al. (2021b), and Musa et al. (2023).

A detailed examination of the results indicates that the proposed estimators consistently achieve lower absolute bias values across both cases, suggesting improved unbiasedness properties. This reduction in bias can be attributed to the incorporation of dual auxiliary information and the imputation mechanism, which effectively compensates for missing observations and stabilizes the estimator structure. Furthermore, the MSE values associated with the proposed estimators are uniformly smaller than those of the competing estimators, demonstrating enhanced precision and reduced sampling

variability. Since MSE captures both variance and squared bias components, this improvement reflects a simultaneous optimization of both aspects of estimator performance.

In terms of efficiency, the PRE values of the proposed estimators exceed 100% relative to most of the existing estimators, indicating clear dominance in terms of relative efficiency. This superiority is observed consistently across different simulation scenarios, highlighting the robustness of the proposed estimators under varying sample sizes, non-response rates, and correlation structures between study and auxiliary variables. The improved performance is particularly pronounced in situations where the auxiliary variable exhibits strong correlation with the study variable, thereby maximizing the gain from auxiliary information utilization.

Additionally, the comparative patterns observed in Tables 1-2 suggest that the proposed estimators maintain stability across both Case 1 and Case 2 configurations, implying that their performance is not sensitive to moderate changes in sampling conditions. This stability is a desirable property in practical survey applications where population characteristics and response behaviors may vary unpredictably.

Overall, the empirical evidence demonstrates that the modified regression-cum-dual mean imputation estimators successfully achieve the intended objectives of bias reduction, MSE minimization, and efficiency improvement. The results provide strong support for the adoption of the proposed estimators as reliable alternatives to existing methods in survey sampling, particularly in the presence of

non-response and incomplete auxiliary information.

### 5.0 Conclusion

This study rigorously examined the development and performance of a class of modified regression-cum-dual mean imputation estimators for estimating the population mean under non-response within a two-phase sampling framework. The theoretical properties of the estimators were derived through higher-order approximations, specifically up to the third order, using Taylor series expansion. This approach enabled a more precise characterization of the estimators' Bias and Mean Square Error (MSE), thereby improving the reliability of analytical comparisons with existing estimators.

To further strengthen the analytical framework, binomial expansion techniques up to the third order were employed to obtain explicit expressions for the bias and MSE of the proposed estimators. These derivations facilitated the establishment of efficiency conditions by comparing the minimum MSE expressions of the proposed estimators with those of competing estimators. Consequently, the study was able to identify the parametric conditions—such as correlation levels, variability ratios, and non-response rates—under which the proposed estimators exhibit superior performance.

In addition to the theoretical analysis, an extensive simulation study was conducted to empirically validate the analytical findings. The simulation results provided strong evidence that the proposed estimators consistently outperform existing estimators, including those of Lee et al. (1994), Singh and Horn (2000), Kadilar and Cingi (2008), Singh (2009), Audu et al. (2021b), and Musa et al. (2023).

Specifically, the proposed estimators demonstrated lower bias, reduced MSE, and higher Percentage Relative Efficiency (PRE) across all considered cases.

Importantly, the integration of imputation techniques with regression and dual estimation principles enhances the capability of the proposed estimators to effectively handle missing data, which is a persistent challenge in real-world survey implementations. The improved performance under non-response conditions underscores the practical relevance of the proposed methodology for large-scale surveys where data incompleteness is unavoidable.

In conclusion, the combined theoretical and empirical evidence confirms that the proposed modified estimators offer a statistically efficient, robust, and practically viable approach for population mean estimation in two-phase sampling designs. The study therefore contributes significantly to the advancement of estimation techniques in survey sampling and provides a solid foundation for future research aimed at extending these methods to more complex sampling designs and real-life data applications.

### References

- 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
- 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.

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.

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.

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.

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

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

Audu, A., Ishaq, O. O., Abubakar, A. Akintola, K. A., Isah, U., Rashida, A. and Muhammad, S. (2021a): Regression-type Imputation Class of Estimators using Auxiliary Attribute. *Asian*

*Research Journal of Mathematics*, 17(5): 1-13.

Audu, A. Danbaba, A., Ahmad, S. K., Musa, N., Shehu, A., Ndatsu, A. M. and Joseph, A. O. (2021b). On the Efficient of Almost Unbiased Mean Imputation When the Population Mean of Auxiliary Variable is Unknown. *Asian Journal of Probability and Statistics*, 15(4): 235-250.