An Alternative Class of Ratio-Regression-Type Estimator under Two-Phase Sampling Scheme

Isah Muhammad[1], Yahaya Zakari[2] and Ahmed Audu[3]

In this study, a new exponential ratio-regression estimator is developed using an auxiliary variable for estimating the finite population mean under a two-phase sampling system. The Bias and Mean Square Error (MSE) of the proposed estimator are derived and compared with some of the estimators in extant literature. Thus, the conditions under which the proposed estimator is better than some existing estimators are provided. Empirically, using four real datasets and simulation study, the proposed estimator performs better than the classical ratio, classical regression, exponential ratio, and exponential regression cum ratio estimator when compared using the criteria of bias, mean square error and percentage relative efficiency. The proposed estimator can be used to estimate the averages of economic variables such as inflation, exchange rate, and standard of living for policy formulation

Keywords: Auxiliary variable, bias, double-sampling, efficiency, mean square error.
JEL Classification: C13, C15, C81, C82
DOI: 10.33429/Cjas.12221.1/5

1. Introduction

In an experimental survey, the use of auxiliary information always shows a substantial increase in the accuracy of population mean estimation. Regression and ratio approaches have been widely used where auxiliary information is known. Many mixtures of ratio estimators have been developed in literature by a number of authors applying linear transformation of the auxiliary variable (Ozgul & Cingi, 2014; Zakari, Muhammad & Sani, 2020). Thus, most scholars have commonly studied exponential estimators, for example, Bahl and Tuteja (1991), Singh *et al.,* (2008) and Grover and Kaur (2011). Many exponential estimators have been proposed using the population information of the auxiliary variable under different sampling schemes. However, when the information on the population mean of the auxiliary variable is

[1]Department of Statistics, Binyaminu Usman Polytechnic, Hadejia, Nigeria
[2]Corresponding Author: Department of Statistics, Ahmadu Bello University, Zaria, Nigeria, yahzaksta@gmail.com, Tel: +2348169766242
[3]Department of Mathematics, Usman Danfodiyo University, Sokoto, Nigeria

not available, one can use the two-phase sampling scheme in obtaining the improved estimator rather than the previous methods. Neyman (1938) was the first to develop the concept of two-phase sampling in estimating the population parameters. Two-phase sampling is a cost effective and a practical method. This sampling scheme is used to obtain the information about the auxiliary variable cheaply from a bigger sample at the first phase and relatively small sample at the second stage.

Ratio-regression-type estimators possess the properties of handling the assumptions of both classical ratio and classical regression estimators. Further developments include the work of Riaz *et al.,* (2014) that developed an estimator by combining the concept of Bahl and Tuteja (1991) exponential type estimator and classical regression estimator. Shabbir and Gupta (2010) proposed a regression-ratio-type exponential estimator by combining Rao (1991) and Bedi (1996) estimators. Ozgul and Cingi (2014) proposed a class of exponential regression cum ratio estimator. However, most of the ratio-based estimators can only be applied when the correlation between the study and auxiliary variables is positively strong. Similarly, the regression type estimator can be applied, when the regression slope does not pass through the origin, and for the product-based estimators, when the estimators are negatively correlated.

It is based on this background, that an alternative ratio-regression-type estimator that provides more efficient estimates than some existing estimators is being proposed. The specific objectives of the study are to: derive the properties of the developed alternative ratio-regression-type estimator such as bias and mean square error (MSE); derive the theoretical efficiency conditions of the proposed estimator over some existing estimators; compare the mean square error (MSE) and percentage relative efficiency (PRE) of the proposed estimator with some existing estimators using real datasets.

Following the introduction is Section 2, which contains the literature review while Section 3 presents the data and methodology of the study. Section 4 discusses the results while conclusion and policy recommendations are presented in Section 5.

2. Literature Review

2.1 Empirical Overview

Singh and Ruiz-Espejo (2003) proposed a new class of ratio-product estimators in two-phase sampling. They derived and obtained optimum values of the parameters along with the minimum mean square error of the proposed estimator. The study used the mean square error criterion in comparing the efficiency of the proposed and existing estimators. In the work of Samiuddin and Hanif (2007), regression and ratio estimation approaches were suggested to estimate the population mean using partial and no information cases in two-phase sampling. The properties of suggested estimators such as bias and mean square error were obtained and tested using real datasets. Based on the comparisons, they found that the proposed estimators performed better. Also, some estimators for two-phase and multiphase sampling were proposed by Ahmad (2008) using information on several auxiliary variables. The regression estimator was developed by Hanif *et al.,* (2010) using different auxiliary variables. The properties of suggested estimator such as bias and mean square error were obtained and tested using real datasets.

Singh and Vishwakarma (2007) modified the work of Bahl and Tuteja (1991) in two-phase sampling. The properties of suggested estimator such as bias and mean square error were obtained. They compared the proposed estimator with some existing estimator based on the criteria of mean square error and relative efficiency using real datasets. Ozgul and Cingi (2014) proposed a class of exponential regression cum ratio estimator for the estimation of population mean under two-phase sampling. Their estimator performed better based on minimum mean square error and percentage relative efficiency. Sukhatme (1962) used two-phase sampling scheme to propose a generalized ratio-type estimator. The classes of the proposed estimators were derived and applied to real datasets. Rao (1973) used two-phase sampling under stratification and non-response problems. The properties of proposed estimator such as bias and mean square error were obtained. The proposed estimator performed better based on minimum mean square error and percentage relative efficiency. Srivenkataramana (1980) proposed to transform an auxiliary variable to increase the performance of the population mean estimator. Sahoo *et al.,* (1993) provided regression approach in estimation by using two auxiliary variables for the two-phase sampling. The properties of proposed estimator such as bias and mean square error were obtained. The proposed estimator performed better based on minimum mean square error and percentage relative efficiency. In the sequence of improving the efficiency of the estimators, Singh & Upadhyaya (1995) suggested a generalized estimator for the population mean using two auxiliary variables in the two-phase sampling. The estimators for the population mean in double sampling considering an additional auxiliary variable have been discussed by Kiregyera (1980, 1984), Sahoo and Sahoo (1993), Sahoo *et al.,* (1993, 1994), Samiuddin and Hanif (2007), Singh and Vishwakarma (2007), Singh *et al.,* (2011), Singh and Choudhury (2012), Sanaullah *et al.,* (2014), Hamad *et al.,* (2013), Malik and Singh (2015), Yadav *et al.,* (2016), Shabbir and Gupta (2017) and Misra (2018).

However, using known values of certain population parameter(s) including coefficient of variation, coefficient of kurtosis, and correlation coefficient, several authors have suggested modified estimators for estimating population mean of the study variable. Recently, Yahaya and Kabir (2017) proposed a modified ratio product estimator of population mean of the variable of interest using median and coefficient of variation of the auxiliary variable in stratified random sampling scheme. But, the studies of these alternative estimators are still not efficient enough and can be improved upon by modification strategy. The ratio-regression-type estimator in this paper incorporates the properties of classical ratio and classical regression estimators.

3. Data and Methodology

3.1 Data

The efficiency of the proposed estimator compared to other double-sampling estimators are evaluated with five real datasets including those used by Kadilar and Cingi (2006) and Ozgul and Cingi (2014) to justify the performance of their estimators. Using the same datasets can be regarded as a fair comparison since the proposed estimator here is a modification of Ozgul and Cingi (2014).

Population I: Ozgul and Cingi (2014);

y: the number of teachers;

x: the number of students in both primary and secondary schools for 923 districts.
Population II: Sukhatme and Sukatme (1970)
y: the number of villages in the circle
x: the circle consisting more than five villages.
Population III: Kadilar and Cingi (2006);
y: Level of apple production;
x: Number of apple trees.
Population IV: Murthy (1967)
y: Output
x: Fixed capital
The parameters of the populations are given in the Table 1:

Table 1: Parameters of the populations used

Parameters	**Population I**	**Population II**	**Population III**	**Population IV**
N	923	89	104	80
n'	400	30	40	40
n	200	20	20	20
ρ_{yx}	0.955	3.360	0.865	51.862
$\overline{Y}$	436.3	0.124	625.37	11.265
$\overline{X}$	11440.50	0.604	13.930	0.354
C_y	1.72	2.190	1.866	0.751
C_x	1.86	0.766	1.653	0.9413

3.2 Notations and Existing Estimators

Consider a finite population, $U = U_1, ..., U_N$, of size N units. Let y denote the study variable taking the values y_i on the unit $U_i, (i = 1, ..., N)$ and Y represents unknown population mean. Also, let x be the auxiliary variable that takes the values x_i on the unit $U_i, (i = 1, ..., N)$ positively related to $\overline{Y}$ and $\overline{X}$ is unknown population mean. It is well known that two-phase sampling is used when the population average of the auxiliary variable is not known (Neyman, 1938).

Note that as described in Muili *et al.,* (2019), Zakari, Muili, *et al.,* (2020), Audu, Isaq Zakari *et al.,* (2020), Audu, Isaq, Muili, *et al.,* (2020), and Audu, Danbaba, *et al.,* (2020):

$$\overline{y} = \sum_{i\varepsilon S} y_i/n, \quad \overline{x} = \sum_{i\varepsilon S} x_i/n \ \ and \ \ \overline{x}' = \sum_{i\varepsilon S'} x_i/n'$$

where S' denotes the first phase sample of a fixed size n'; S denotes the second phase sample of a fixed size n; y denotes the study variable taking the values y_i on the unit $U_i, (i = 1,...,N)$ and $\overline{Y}$ represents unknown population mean; x denotes the auxiliary variable that takes the values x_i on the unit $U_i, (i = 1,...,N)$ positively related to $\overline{Y}$ and $\overline{X}$ is unknown population mean; $\overline{x}'$ denotes the primary sample mean of the auxiliary variable in the first phase sample of size n'; $\overline{x}$ denotes the sub-sample mean of the auxiliary variables in the second phase sample of size n and $\overline{y}$ denotes the mean of the study variable y in the second phase sample. For more detailed information see Ozgul and Cingi (2014).

We also define the following notations:

$\lambda = \left(\frac{1}{n} - \frac{1}{N}\right)$, $\lambda' = \left(\frac{1}{n'} - \frac{1}{N}\right)$, $C_y = S_y/\overline{Y}, C_x = S_x/\overline{X}, \rho_{yx} = S_{yx}/(S_y S_x)$,

$S_y^2 = \frac{\sum_{i=1}^{N}\left(y_i - \overline{Y}\right)^2}{N-1}$, $S_x^2 = \frac{\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2}{N-1}$, and $S_{yx} = \frac{\sum_{i=1}^{N}\left(y_i - \overline{Y}\right)\left(x_i - \overline{X}\right)}{N-1}$

where n' is the primary sample size; n is the sub-sample size; N is the number of units in the population; $\overline{Y}$ is the population mean of the study variable; $\overline{X}$ is the population mean of the auxiliary variable; λ is a known constant involving the samples and population units; ρ_{yx} is the population correlation coefficient between the auxiliary and the study variables; S_{yx} is the covariance between the auxiliary and the study variables; S_x^2 and S_y^2 are the variances of the auxiliary and the study variables, respectively; C_x and C_y are the population coefficients of variation of the auxiliary and study variables, respectively (Ozgul and Cingi, 2014).

Singh and Vishwakarma (2007) modified the two-phase sampling exponential ratio estimator as:

$$\overline{y}_{svr} = \overline{y} exp\left(\frac{\overline{x}' - \overline{x}}{\overline{x}' + \overline{x}}\right) \tag{1}$$

where $\overline{x}'$ is the primary sample mean of the auxiliary variable, $\overline{y}$ and $\overline{x}$ are the sub-sample means of the study and auxiliary variables, respectively.

The expression for the mean square error (MSE) equation of the estimator in (1), up to the first order of approximation is given by:

$$MSE\left(\overline{y}_{svr}\right) = \overline{Y}^2\left[\lambda C_y^2 + \lambda'\left(\frac{C_x^2}{4} - \rho_{yx} C_y C_p\right)\right] \tag{2}$$

Substituting regression estimator in (1), Ozgul and Cingi (2014) proposed a class of exponential regression cum ratio estimator as follows:

$$\hat{y}_{NH} = \left[k_1\bar{y} + k_2\left(\bar{x}' - \bar{x}\right)\right] exp\left(\frac{\bar{z}' - \bar{z}}{\bar{z}' + \bar{z}}\right) \tag{3}$$

where k_1 and k_2 are known constants, $\bar{z}'$ represents first phase transformation of the auxiliary variable as $\bar{z}' = a\bar{x}' + b$ and $\bar{z}$ is a second phase transformation of the auxiliary variable as $\bar{z} = a\bar{x} + b$. The optimum values of k_1 and k_2 along with the minimum mean square error are given, respectively as:

$$k_{1(opt)} = 1 - \frac{2-\lambda'\theta^2C_x^2}{\left[1+(\lambda-\lambda'\rho_{yx}^2)\right]}$$

$$k_{2(opt)} = \frac{\hat{Y}}{\hat{X}}\left[(\theta-1) + \frac{2-\lambda'\theta^2C_x^2}{\left[1+(\lambda-\lambda'\rho_{yx}^2)\right]}(2\theta - K_{yx})\right]$$

$$MSE\left(\bar{\hat{y}}_{NH}\right)^2 \frac{Var(\bar{y}_{lr})(1-\lambda'\theta^2C_x^2) - \frac{\lambda'^2\bar{Y}^2\theta^4C_x^4}{4}}{[\bar{Y}^2 + Var(\bar{y}_{lr})]}_{min} \tag{4}$$

$\overline{Y}$ is the population mean of the study variable; $\overline{X}$ is the population mean of the auxiliary variable; λ and λ' are a known constant involving the samples and population units; ρ_{yx} is the population correlation coefficient between the auxiliary and the study variables; C_x is the population coefficients of variation of the auxiliary variable; θ is known constant; $Var(\bar{y}_{lr})$ is the variance of linear regression estimator; K_{yx} is optimum value of k_1.

3.3 The Proposed Estimator

In application, the main purpose was to construct stable and efficient estimator using ratio of sample means as coefficient for $\bar{y}$ to minimize the influence of extreme values or outliers in Ozgul and Cingi (2014) estimator. We proposed an alternative ratio-regression-type estimator where the first component is a mixture of ratio with coefficient, α_1, and regression with coefficient, α_2, as $\left[\alpha_1\bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) + \alpha_2\left(\bar{x}' - \bar{x}\right)\right]$ and the second component is exponential estimator $exp\left(\frac{\bar{z}'-\bar{z}}{\bar{z}'+\bar{z}}\right)$. A linear combination approach was used to combine the ratio, regression and exponential estimators which yielded a new ratio-regression exponential-type estimator for finite population mean using the auxiliary information in two-phase sampling as:

$$\hat{y}_{IM} = \left[\alpha_1 \overline{y}\left(\frac{\overline{x}'}{\overline{x}}\right) + \alpha_2\left(\overline{x}' - \overline{x}\right)\right] exp\left(\frac{\overline{z}' - \overline{z}}{\overline{z}' + \overline{z}}\right) \tag{5}$$

where, $\hat{y}_{IM}$is the proposed estimator such that, when $\alpha_1 = 1$ and $\alpha_2 = 0$, the mixture estimator reduces to ratio-type exponential estimator while when $\alpha_1 = 0$ and $\alpha_2 = 1$, the mixture estimator reduces to regression-type exponential estimator, where α_1 and α_2 are real parameters to be determined such that the mean square error of the proposed estimator ($\hat{y}_{IM}$) is minimum.

To obtain the bias and MSE equations for the proposed estimator, we define the following notations:

$$\overline{y} = \overline{Y}\left(1 + e_0\right), \quad \overline{x} = \overline{X}\left(1 + e_1\right), \quad and\; \overline{x}' = \overline{X}\left(1 + e_1'\right)$$

such that

$$E(e_0) = E(e_1) = E\left(e_1'\right) = 0$$

$$E\left({e_0}^2\right) = \lambda C_y^2, \qquad E\left({e_1}^2\right) = \lambda C_x^2, \qquad E\left({e_1'}^2\right) = \lambda' C_x^2$$

$$E\left(e_0 e_1\right) = \lambda \rho C_y C_x, \quad E\left(e_0 e_1'\right) = \lambda' \rho C_y C_x, \quad E\left(e_1 e_1'\right) = \lambda' C_x^2$$

Expressing the estimator, $\hat{y}_{IM}$, in terms of e_i $(i = 0, \quad 1)$, we can write (5) as

$$\hat{y}_{IM} = a_1 \bar{Y}(1+e_0)(1+e_1)^{-1}(1+e_1') + a_2\left[\bar{X}(1+e_1') - \hat{X}(1+e_1)\right] exp\left(\frac{a_x\bar{X}(1+e_1') - a_x\hat{X}(1+e_1)}{a_x\hat{X}(1+e_1') + a_x\bar{X}(1+e_1) + 2b_x}\right) \tag{6}$$

Using Taylor series expansion of $(1+e_1)^{-1}$ to the first order of approximation, neglecting the terms of e's greater than two gives:

$$\hat{y}_{IM} = \left[a_1\hat{Y}(1+e_0+e_1'+e_0e_1')(1-e_1+e_1^2)+a_2\hat{X}(e_1'-e_1)\right] exp\left(\frac{a_x\hat{X}(e_1'-e_1)}{2\left(a_x\hat{X}+b_x\right)+a_x\hat{X}(e_1'+e_1)}\right) \quad (7)$$

By factorizing the exponential part and expanding the first term to the first order of approximation, multiplying out and neglecting the terms of e's greater than two, we get:

$$\hat{y}_{IM} = \left[\alpha_1\hat{Y}(1+e_0-e_1+e_1'+e_1^2-e_0e_1+e_0e_1'-e_1e_1')+\alpha_2\hat{X}(e_1'-e_1)\right] exp\left(\frac{\theta(e_1'-e_1)}{2\left(1+\frac{\theta(e_1'+e_1)}{2}\right)}\right) \quad (8)$$

where $\theta = \frac{a_x\hat{X}}{a_x\hat{X}+b_x}$
Taking out the common terms in the exponential part of (8), we get:

$$\hat{y}_{IM} = \left[a_1\hat{Y}(1+e_0-e_1+e_1'+e_1^2-e_0e_1+e_0e_1'-e_1e_1')+a_2\hat{X}(e_1'-e_1)\right] exp\left(\frac{\theta(e_1'-e_1)}{2}\left[1+\frac{\theta(e_1'+e_1)}{2}\right]^{-1}\right) \quad (9)$$

By exponential series and neglecting the terms of e's greater than two, the exponential part of (9), becomes

$$\approx \left(\alpha_1\bar{Y}-\bar{Y}-(\alpha_1\bar{Y}+\alpha_2\bar{X})e_1+\alpha_1\bar{Y}e_1^2+(\alpha_1\bar{Y}+\alpha_2\bar{X})e_1'-\alpha_1\bar{Y}e_1e_1'+...\right) \left(...\alpha_1\bar{Y}e_0-\alpha_1\bar{Y}e_0e_1+\alpha_1\bar{Y}e_0e_1'\right)\times\left(1+\left(\frac{\theta}{2}\right)e_1'-\frac{\theta e_1}{2}-\frac{\theta^2 e_1'^2}{8}+...\right) \left(...\frac{3\theta^2 {e_1}^2}{8}-\frac{\theta^2 e_1e_1'}{4}\right) \quad (10)$$

Expanding the RHS of (10) to the first order of approximation, neglecting the terms of e's greater than two and taking out the common terms gives:

$$(\hat{y}_{IM}) \approx \bar{Y} + \bar{Y}(\alpha_1 - 1) + e_1\bar{Y}e_0 + \alpha_1' \left(\alpha_1\bar{Y}\frac{\theta}{2} + (\alpha_1\bar{Y} + \alpha_2\bar{X})\right) - e_1 \left(\bar{Y}\alpha_1\frac{\theta}{2} + (\alpha_1\bar{Y} + \alpha_2\bar{X})\right) + e_1'^2\left(\frac{\theta}{2}(\alpha_1\bar{Y} + \alpha_2\hat{X}) - \bar{Y}\alpha_1\frac{\theta^2}{8}\right) + e_1^2 \left(3\bar{Y}\frac{\theta^2}{8} + \frac{\theta}{2}(\alpha_1\bar{Y} + \alpha_2\bar{X}) + \alpha_1\bar{Y}\right) - e_1e_1'\left(\bar{Y}\alpha_1\frac{\theta^2}{4} + (\alpha_1\bar{Y} + \alpha_2\bar{X})\theta + \alpha_1\bar{Y}\right) + e_0e_1'\left(\alpha_1\bar{Y}\frac{\theta}{2} + \alpha_1\bar{Y}\right) - e_0e_1\left(\alpha_1\bar{Y}\frac{\theta}{2} + \alpha_1\bar{Y}\right) \quad (11)$$

$$(\hat{y}_{IM} - \bar{Y}) \approx \bar{Y}(\alpha_1 - 1) + e_1\bar{Y}e_0 + \alpha_1' \left(\alpha_1\bar{Y}\frac{\theta}{2} + (\alpha_1\bar{Y} + \alpha_2\bar{X})\right) - e_1 \left(\bar{Y}\alpha_1\frac{\theta}{2} + (\alpha_1\bar{Y} + \alpha_2\bar{X})\right) + e_1'^2\left(\frac{\theta}{2}(\alpha_1\bar{Y} + \alpha_2\bar{X}) - \bar{Y}\alpha_1\frac{\theta^2}{8}\right) + e_1^2 \left(3\bar{Y}\frac{\theta^2}{8} + \frac{\theta}{2}(\alpha_1\bar{Y} + \alpha_2\bar{X}) + \alpha_1\bar{Y}\right) - e_1e_1'\left(\bar{Y}\alpha_1\frac{\theta^2}{4} + (\alpha_1\bar{Y} + \alpha_2\bar{X})\theta + \alpha_1\bar{Y}\right) + e_0e_1'\left(\alpha_1\bar{Y}\frac{\theta}{2} + \alpha_1\bar{Y}\right) - e_0e_1\left(\alpha_1\bar{Y}\frac{\theta}{2} + \alpha_1\bar{Y}\right) \quad (12)$$

From equation (12), the bias of the proposed estimator is obtained as:

$$Bias\,(\hat{y}_{IM}) = \bar{Y}(a_1 - 1) + \left(\lambda - \lambda'\right)Cx^2\left(\frac{3\theta^2}{8}\bar{Y} + \frac{\theta}{2}\left(\alpha_1\bar{Y} + \alpha_2\bar{X}\right) + \alpha_1\bar{Y}\right) + \left(\lambda - \lambda'\right)\left(\bar{Y}\alpha_1\frac{\theta}{2} + \alpha_1\bar{Y}\right)\rho C_yC_x \quad (13)$$

Subsequently, the MSE equation of the proposed estimator is also obtained from equation (12) as:

$$MSE\,(\hat{y}_{IM}) = \bar{Y}^2 + {a_1}^2W_1 - a_1W_2 + a_2^2W_3 - a_2W_4 + 2a_1a_2W_5 \quad (14)$$

where:

$$
\begin{aligned}
W_1 &= \bar{Y}^2\left(1+(\lambda-\lambda')C_x^2\left(\theta^2+2\theta+2\right)+\lambda C_y^2-2(\lambda-\lambda')\rho C_yC_x(\theta+1)\right)\\
W_2 &= \bar{Y}^2\left(2+(\lambda-\lambda')C_x^2\left(3+\frac{3\theta^2}{4}\right)-(\lambda-\lambda')\rho C_yC_x(\theta+2)\right)\\
W_3 &= (\lambda-\lambda')\bar{X}C_x^2\\
W_4 &= \bar{Y}\bar{X}\theta C_x^2(\lambda-\lambda')\\
W_5 &= (\lambda-\lambda')\left(C_x^2\bar{Y}\bar{X}(\theta+1)+\bar{Y}\bar{X}\rho C_yC_x\right)
\end{aligned}
$$

Differentiating (14) partially with respect to α_1 and α_2, and equating to zero, we obtain the optimum values of α_1 and α_2, respectively as:

$$\alpha_1 = \frac{W_2W_3-W_4W_5}{2\left(W_1W_3-W_5^2\right)} \tag{15}$$

and

$$\alpha_2 = \frac{W_1W_4-W_2W_5}{2\left(W_1W_3-W_5\right)} \tag{16}$$

Substituting the α_1 and α_2 optimum values into (14), we obtain the minimum mean square error of the proposed estimator as:

$$MSE\left(\bar{y}_{IM}\right)^2\left(\frac{W_2^2W_3+W_1W_4^2-2W_2W_4W_5}{4\left(W_1W_3-W_5^2\right)}\right)_{min} \tag{17}$$

Table 2: Some members of the suggested estimator, $\widehat{y}_{IM}$

Estimators	Values a_x	of b_x
$\hat{t}_{CE1} = \left[\alpha_1 \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) + \alpha_2\left(\bar{x}' - \bar{x}\right)\right] exp\left(\frac{\bar{x}'-\bar{x}}{\bar{x}'+\bar{x}}\right)$	1	0
$\hat{t}_{CE1} = \left[\alpha_1 \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) + \alpha_2\left(\bar{x}' - \bar{x}\right)\right] exp\left(\frac{\bar{x}'-\bar{x}}{\bar{x}'+\bar{x}+2\beta_2(\phi)}\right)$	1	$\beta_2(\phi)$
$\hat{t}_{CE2} = \left[\alpha_1 \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) + \alpha_2\left(\bar{x}' - \bar{x}\right)\right] exp\left(\frac{\bar{x}'-\bar{x}}{\bar{x}'+\bar{x}+2C_x}\right)$	1	C_x
$\hat{t}_{CE3} = \left[\alpha_1 \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) + \alpha_2\left(\bar{x}' - \bar{x}\right)\right] exp\left(\frac{\bar{x}'-\bar{x}}{\bar{x}'+\bar{x}+2\rho_{xy}}\right)$	1	ρ_{xy}
$\hat{t}_{CE4} = \left[\alpha_1 \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) + \alpha_2\left(\bar{x}' - \bar{x}\right)\right] exp\left(\frac{\beta_2(\phi)(\bar{x}'-\bar{x})}{\beta_2(\phi)(\bar{x}'+\bar{x})+2C_x}\right)$	$\beta_2(\phi)$	C_x
$\hat{t}_{CE5} = \left[\alpha_1 \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) + \alpha_2\left(\bar{x}' - \bar{x}\right)\right] exp\left(\frac{\beta_2(\phi)(\bar{x}'-\bar{x})}{\beta_2(\phi)(\bar{x}'+\bar{x})+2\rho_{xy}}\right)$	$\beta_2(\phi)$	ρ_{xy}
$\hat{t}_{CE6} = \left[\alpha_1 \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) + \alpha_2\left(\bar{x}' - \bar{x}\right)\right] exp\left(\frac{C_x(\bar{x}'-\bar{x})}{C_x(\bar{x}'+\bar{x})+2\beta_2(\phi)}\right)$	C_x	$\beta_2(\phi)$
$\hat{t}_{CE7} = \left[\alpha_1 \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) + \alpha_2\left(\bar{x}' - \bar{x}\right)\right] exp\left(\frac{C_x(\bar{x}'-\bar{x})}{C_x(\bar{x}'+\bar{x})+2\rho_{xy}}\right)$	C_x	ρ_{xy}
$\hat{t}_{CE8} = \left[\alpha_1 \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) + \alpha_2\left(\bar{x}' - \bar{x}\right)\right] exp\left(\frac{\rho_{xy}(\bar{x}'-\bar{x})}{\rho_{xy}(\bar{x}'+\bar{x})+2\beta_2(\phi)}\right)$	ρ_{xy}	$\beta_2(\phi)$
$\hat{t}_{CE9} = \left[\alpha_1 \bar{y}\left(\frac{\bar{x}'}{\bar{x}}\right) + \alpha_2\left(\bar{x}' - \bar{x}\right)\right] exp\left(\frac{\rho_{xy}(\bar{x}'-\bar{x})}{\rho_{xy}(\bar{x}'+\bar{x})+2C_x}\right)$	ρ_{xy}	C_x

where ρ_{yx} is the correlation coefficient between the auxiliary and the study variables; C_x is the coefficients of variation of the auxiliary variable; $\beta_2(\phi)$ is the coefficient of kurtosis of the auxiliary variable.

3.3 Efficiency Conditions of the Proposed Estimator

In this section, we illustrate how the proposed ratio-regression estimator will have least mean square error compared to the existing classical ratio estimator, classical regression estimator and some estimators in the literature for assessing the finite population mean. Also, the study explains the circumstances under which the estimator developed is more successful than estimators considered:

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

$$MSE(\hat{y}_{IM})\bar{y}_{Rmin};$$

$$\hat{Y}^2 - \frac{A}{B} < \hat{Y}^2[\lambda C_y^2 + \lambda'(C_x^2 - 2\rho_{yx}C_yC_x)]$$

$$\hat{Y}^2[\lambda C_y^2 + \lambda'(C_x^2 - 2\rho_{yx}C_yC_x)] - \hat{Y}^2 + \frac{A}{B} > 0 \tag{18}$$

where :

$$A = W_2^2 W_3 + W_1 W_4^2 - 2W_2 W_4 W_5$$

$$B = 4\left(W_1 W_3 - W_5^2\right)$$

Since condition (18) is satisfied, the proposed estimator is more efficient than the usual ratio estimator.

ii To compare the MSE of the proposed estimator with that of the usual regression estimator, the following is relevant:

$$MSE(\hat{y}_{IM})\overline{y}_{lr_{min}};$$

$$\hat{Y}^2 - \frac{A}{B} < \hat{Y}^2 C_y^2[\lambda - \lambda' \rho_{yx}^2]$$

$$\hat{Y}^2 C_y^2[\lambda - \lambda' \rho_{yx}^2] - \hat{Y}^2 + \frac{A}{B} > 0 \tag{19}$$

From (19), the RHS is greater than zero, meaning that the condition is satisfied. Thus, the proposed estimator is more efficient than the regression estimator.

iii In order to compare the MSE of the estimator proposed with that of Singh and Vishwakarma (2007) in (2), we have the following:

$$MSE(\hat{y}_{IM})\overline{y}_{svr_{min}};$$

$$\hat{Y}^2 - \frac{A}{B} < \hat{Y}^2\left(\lambda C_y^2 + \lambda'\left(\frac{C_x^2}{4} - \rho_{yx} C_y C_x\right)\right)$$

$$\hat{Y}^2\left(\lambda C_y^2 + \lambda'\left(\frac{C_x^2}{4} - \rho_{yx} C_y C_x\right)\right) - \hat{Y}^2 + \frac{A}{B} > 0 \tag{20}$$

From (20), the RHS is greater than zero, meaning that the condition is satisfied. Thus, the suggested estimator showed more efficiency than the work of Singh and Vishwakarma (2007) in (1).

iv To compare the proposed estimator's MSE with the exponential regression cum ratio estimator of Ozgul and Cingi (2014) MSE in (4), gives:

$$MSE(\bar{y}_{IM})\bar{y}_{NH_{min_{min}}};$$

$$\hat{Y}^2 - \frac{A}{B} < barY^2 \frac{Var(\bar{y}_{lr})(1-\lambda'\theta^2 C_x^2) - \frac{\lambda'^2 \bar{Y}^2 \theta^4 C_x^4}{4}}{\left[\hat{Y}^2 + Var(\bar{y}_{lr})\right]}$$

$$\hat{Y}^2 \frac{Var(\bar{y}_{lr})(1-\lambda'\theta^2 C_x^2) - \frac{\lambda'^2 \bar{Y}^2 \theta^4 C_x^4}{4}}{[\bar{Y}^2 + Var(\bar{y}_{lr})]} - \bar{Y}^2 + \frac{A}{B} > 0 \tag{21}$$

In (21), RHS is greater than zero. Then, (21) has satisfied the necessary condition, meaning that the proposed estimator is more efficient than Ozgul and Cingi (2014) exponential regression cum ratio estimator in (3).

4. Results and Discussion

4.1 Simulation

Here, simulation studies were performed to evaluate the efficiencies of the proposed estimator over other estimators considered in this paper. The simulated data were generated using the models defined in Table 3 and biases, MSEs, PREs were computed using the functions below:

$bias(T) = \frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{\bar{Y}} - \overline{Y}\right)$, $MSE(T) = \frac{1}{1000}\sum_{i=1}^{1000}\left(\hat{\bar{Y}} - \overline{Y}\right)^2$ and $PRE(T) = \frac{MSE(\bar{y})}{MSE(T)}X100$, where T is either of $\bar{y}$, $\bar{y}_R$, $\bar{y}_{svr}$, $\bar{y}_{NH}$ and $\hat{y}_{IM}$.

Table 3: Simulation Parameters and Models

Population	Auxiliary Variables (X)	Study Variables (Y)
1	$X \sim beta(1,4)$	Model 1: $Y = X + E$
2	$X \sim gamma(1,4)$	Model 2: $Y = X^2 + X + E$

Table 3 shows the simulation parameters and models adopted; the population 1 follows a mixed beta-normal distribution, while population 2 follows a mixed gamma-normal distribution. In each population, the population parameters were simulated based on linear and quadratic models.

Table 4: Various estimators' bias, MSE and PRE using Population 1

Estimators	Bias	MSE	PRE
Models 1: $Y = X + E$			
Sample mean $(\overline{y})$	0.00849	0.014011	100
Classical Ratio $(\overline{y}_R)$	0.00775	0.14079	99.52
Classical Regression $(\overline{y}_{lr})$	0.00926	0.14059	99.66
Singh and Vishwakarma (2007) $(\overline{y}_{svr})$	0.00808	0.14019	99.94
Ozgul and Cingi (2014) $(\overline{y}_{NH})$	-0.20717	0.17811	78.66
Proposed Estimator $(\hat{y}_{IM})$	-0.09962	0.00994	1409.56
Model 2: $Y = X^2 + X + E$			
Sample mean $(\overline{y})$	0.01532	0.13523	100
Classical Ratio $(\overline{y}_R)$	0.01543	0.13793	98.04
Classical Regression $(\overline{y}_{lr})$	0.01453	0.137	98.71
Singh and Vishwakarma (2007) $(\overline{y}_{svr})$	0.01533	0.13632	99.20
Ozgul and Cingi (2014) $(\overline{y}_{NH})$	-0.16534	0.15851	85.31
Proposed Estimator $(\hat{y}_{IM})$	-0.07571	0.00573	2360.03

Table 4 shows the various estimators' bias, mean square error and percentage relative efficiency values based on the population parameters obtained from linear and quadratic models when the underlining distribution is mixed beta-normal, namely population 1. For the linear model, it is observed that the proposed ratio-regression-type estimator (-0.09962, 0.00994 and 1409.56) has the minimum bias and mean square error and the supreme percentage relative efficiency values, respectively, compared to the sample mean (0.00849, 0.014011 and 100); classical ratio (0.00775, 0.14079 and 99.52); classical regression (0.00926, 0.14059 and 99.66); Singh and Vishwakarma (2007) exponential ratio-type (0.00808, 0.14019 and 99.94) and Ozgul and Cingi (2014) ratio-regression-type estimator (-0.20717, 0.17811 and 78.66). Similarly, for the quadratic model, it is also observed that the proposed ratio-regression-type estimator (-0.07571, 0.00573 and 2360.03) has the minimum bias and mean square error and the supreme percentage relative efficiency values, respectively, compared to the sample mean (0.01532, 0.13523 and 100); classical ratio (0.01543, 0.13793 and 98.04); classical regression (0.01453, 0.137 and 98.71); Singh and Vishwakarma (2007) exponential ratio-type (0.01533, 0.13632 and 99.20) and Ozgul and Cingi (2014) ratio-regression-type estimator (-0.16534, 0.15851 and 85.31).

Table 5: Various Estimators' MSE and PRE using Population 2

Estimators	Bias	MSE	PRE
Model 1: $Y = X + E$			
Sample mean ($\bar{y}$)	-0.00129	0.14824	100
Classical Ratio ($\bar{y}_R$)	-0.00342	0.15007	98.78
Classical Regression (y_{lr})	1e-04	0.14771	100.36
Singh and Vishwakarma (2007) ($\bar{y}_{svr}$)	-0.00244	0.14873	99.67
Ozgul and Cingi (2014) ($\bar{y}_{NH}$)	-0.30886	0.23737	62.45
Proposed Estimator ($\hat{y}_{IM}$)	-0.15403	0.02375	624.17
Model 2: $Y = X^2 + X + E$			
Sample mean ($\bar{y}$)	0.01123	0.15975	100
Classical Ratio ($\bar{y}_R$)	0.00931	0.16072	99.40
Classical Regression ($\bar{y}_{lr}$)	0.01255	0.15802	101.09
Singh and Vishwakarma (2007) ($\bar{y}_{svr}$)	0.01006	0.15971	100.03
Ozgul and Cingi (2014) ($\bar{y}_{NH}$)	-0.71299	0.6604	24.19
Proposed Estimator ($\hat{y}_{IM}$)	-0.34176	0.117	136.54

Table 5 shows the various estimators' biases, MSEs and PREs values based on the population parameters obtained from linear and quadratic models when the underlining distribution is mixed gamma-normal, namely population 2. For the linear model, it is observed that the proposed ratio-regression-type estimator (-0.15403, 0.02375 and 624.17) has the minimum bias and MSE and the supreme PRE values, respectively, compared to the sample mean (-0.00129, 0.14824 and 100); classical ratio (-0.00342, 0.15007 and 98.78); classical regression (1e-04, 0.14771 and 100.36); Singh and Vishwakarma (2007) exponential ratio-type (-0.00244, 0.14873 and 99.67) and Ozgul and Cingi (2014) ratio-regression-type estimator -0.30886, 0.23737 and 62.45). Similarly, for the quadratic model, it is also observed that the proposed ratio-regression-type estimator (-0.34176, 0.117 and 136.54) has the minimum bias and MSE and the supreme PRE values, respectively, compared to the sample mean (0.01123, 0.15975 and 100); classical ratio (0.00931, 0.16072 and 94.40); classical regression (0.01255, 0.58021 and 101.09); Singh and Vishwakarma (2007) exponential ratio-type (0.01006, 0.15971 and 100.03) and Ozgul and Cingi (2014) ratio-regression-type estimator (-0.71299, 0.6604 and 24.19). Therefore, based on the criteria of MSE and PRE, the proposed ratio-regression-type estimator performed better and is more efficient than the existing estimators considered. Thus, the proposed ratio-regression-type estimator in this study possesses the properties of handling the assumptions of both classical ratio and classical regression estimators.

4.2 Empirical Results

Table 6: Various estimators' MSE and PRE values with regard to $\bar{y}$

Estimators	Population I		Population II	
	MSE	PRE	MSE	PRE
Sample mean ($\bar{y}$)	2205.636	100	159654.6	100
Classical Ratio ($\bar{y}_R$)	944.1017	233.6227	680788.5	23.45142
Classical Regression ($\bar{y}_{lr}$)	921.6096	239.3243	119377.5	133.7393
Singh and Vishwakarma (2007) ($\bar{y}_{svr}$)	1163.268	189.6069	194613	82.03697
Ozgul and Cingi (2014) ($\bar{y}_{NH}$)	1142.667	193.0253	191126.3	83.53356
Proposed Estimator ($\hat{y}_{IM}$)	914.667	241.1409	1651.32	9668.302

An alternative ratio-regression-type estimator is developed using an auxiliary variable for estimating the finite population mean under a two-phase sampling scheme. The bias and MSE of the proposed estimator were derived and they were compared with some of the estimators in the extant literature has shown in Table 6. Theoretical proofs and numerical examples with five real datasets were carried out. The mean square error and percentage relative efficiency values of the proposed estimator and some existing estimators considered are presented in Table 6. Based on the results obtained from first dataset namely population I, it is observed that the proposed ratio-regression-type estimator (914.667 and 241.1409) has the minimum mean square error and the supreme percentage relative efficiency values, respectively, compared to the sample mean (2205.636 and 100); classical ratio (944.1017 and 233.6227); classical regression (921.6096 and 239.3243); Singh and Vishwakarma (2007) exponential ratio-type (1163.268 and 189.6069) and Ozgul and Cingi (2014) ratio-regression-type estimator (1142.667 and 193.0253). Also based on the results obtained from second dataset namely population II, it is observed that the proposed ratio-regression-type estimator (1651.32 and 9668.302) has the minimum mean square error and the supreme percentage relative efficiency values, respectively, compared to the sample mean (159654.6 and 100); classical ratio (680788.5 and 23.45142); classical regression (119377.5 and 133.7393); Singh and Vishwakarma (2007) exponential ratio-type (194613 and 82.03697) and Ozgul and Cingi (2014) ratio-regression-type estimator (191126.3 and 83.53356). Therefore, based on the criteria of mean square error and percentage relative efficiency, the proposed ratio-regression-type estimator performs better and is more efficient than the existing estimators considered. Thus, the proposed ratio-regression-type estimator in this study possesses the properties of handling the assumptions of both classical ratio and classical regression estimators.

Table 7: Various estimators' MSE and PRE values with regard to $\bar{y}$

Estimators	Population III		Population IV	
	MSE	PRE	MSE	PRE
Sample mean ($\bar{y}$)	54993.75	100	12.62215	100
Classical Ratio ($\bar{y}_R$)	29536.17	186.1912	16.88635	74.74765
Classical Regression ($\bar{y}_{lr}$)	29521.36	186.2846	5.16628	244.318
Singh and Vishwakarma (2007) ($\bar{y}_{svr}$)	35586.14	154.537	5.286313	238.7704
Ozgul and Cingi (2014) ($\bar{y}_{NH}$)	30361.55	181.1296	5.11836	246.6054
Proposed Estimator ($\hat{y}_{IM}$)	24711.69	22.5414	5.091813	247.8911

In Table 7, the mean square error and percentage relative efficiency values of the proposed estimator and some existing estimators considered are presented. Based on the results obtained from third dataset namely population III, it is observed that the proposed ratio-regression-type estimator (24711.69and 22.5414) has the minimum mean square error and the supreme percentage relative efficiency values, respectively, compared to the sample mean (54993.75 and 100); classical ratio (29536.17 and 186.1912); classical regression (29521.36 and 186.2846); Singh and Vishwakarma (2007) exponential ratio-type (35586.14 and 154.537) and Ozgul and Cingi (2014) ratio-regression-type estimator (30361.55 and 181.1296). Also based on the results obtained from fourth dataset namely population IV, it is observed that the proposed ratio-regression-type estimator (5.091813 and 247.8911) has the minimum mean square error and the supreme percentage relative efficiency values, respectively, compared to the sample mean (12.62215 and 100); classical ratio (16.88635 and 74.74765); classical regression (5.16628 and 244.318); Singh and Vishwakarma (2007) exponential ratio-type (5.286313 and 238.7704) and Ozgul and Cingi (2014) ratio-regression-type estimator (5.11836 and 246.6054). Therefore, based on the criteria of mean square error and percentage relative efficiency, the proposed ratio-regression-type estimator performs better and is more efficient than the existing estimators considered.

5. Conclusion and Policy Recommendations

5.1 Conclusion

An alternative ratio-regression estimator for estimating finite population mean in two-phase sampling is proposed in this paper. Expressions for bias and MSE of the estimator are derived up to first degree of approximation. The theoretical efficiency conditions in which the proposed estimator is more efficient than some of the current estimators have been presented. The proposed estimator is thus empirically compared with some existing estimators using real datasets and simulation study. Evidence from the study revealed that the proposed estimator performs better than the existing estimators considered when compared using bias, MSE and PRE, respectively.

5.2 Policy Recommendations

Based on our findings, the proposed estimator, can be used to estimate averages of some economic variables such as inflation, exchange rate, and standard of living for policy formulation.

Furthermore, for future work, one may modify the proposed estimator by incorporating two or more auxiliary variables to gain more efficiency in estimation process, since adopting auxiliary variable always increases the efficiency of estimators.

References

Ahmad, Z. (2008). Generalized multivariate ratio and regression estimators for multi-phase sampling. Doctoral Thesis, National College of Business Administration and Economics, Lahore, Pakistan.

Audu, A., Ishaq, O. O., Zakari, Y., Daniel, D. W., Muili, J. O., & Ndatsu, A. M. (2020). Regression-cum-exponential ratio imputation class of estimators of population mean in the presence of non-response. *Journal of Pure and Applied Sciences, 20,* 58-63.

Audu, A., Ishaq, O. O., Muili, J. O., Zakari, Y., Ndatsu, A. M., &Muhammed, S. (2020). On the efficiency of imputation estimators usingauxiliary attribute. *Continental Journal of Applied Sciences, 15*(1), 1-13.

Audu, A., Danbaba, A., Abubakar, A., Ishaq, O. O., & Zakari, Y. (2020). On the efficiency of calibration ratio estimators of population mean in stratified random sampling. Conference Proceedings of Royal Statistical Society Nigeria Local Group. 247-261.

Bahl, S., & Tuteja, R. K. (1991). Ratio and product exponential estimators. *Journal of Information and Optimization Sciences*, *12*(1), 159-164.

Bandyopadhyay, S. (1980). Improved ratio and product estimators. *Austrian Journal of Statistics. 36*(3), 217–225.

Bedi, P. K. (1996). Efficient utilization of auxiliary information at estimation stage. *Biometrical Journal, 38*(8), 973-976.

Grover, L. K., & Kaur, P. (2011). An improved estimator of the finite population mean in simple random sampling. *Model Assisted Statistics and Applications, 6*(1), 47-55.

Gupta, P. C. (1978). On some quadratic and higher degree ratio and product estimators. *Journal of India on Social and Agricultural Statistics, 30*(2), 71-80.

Hanif, M., Shahbaz, M. Q., & Ahmad, Z. (2010). Some improved estimators in multiphase sampling. *Pakistan Journal of Statistics, 26*(1), 195-202.

Hamad, N., Hanif, M., & Haider, N. (2013). A regression type estimator with two auxiliary variables for two-phase sampling. *Open Journal of Statistics, 3*(2), 74–81.

Kadilar, C., & Cingi, H. (2006). An improvement in estimating the population mean by using the correlation coefficient. *Hacettepe Journal of Mathematics and Statistics, 35*(1), 103-109.

Kiregyera, B. (1980). A chain ratio-type estimator in finite population double sampling using two auxiliary variables. *Metrika, 27*(1), 217–223.

Kiregyera, B. (1984). Regression-type estimator using two auxiliary variables and model of double sampling from finite populations. *Metrika, 31*(1), 215–216.

Malik, S., & Singh, R. (2015). Estimation of population mean using information on auxiliary attribute in two-phase sampling. *Applied Mathematics and Computation, 261,* 114–118.

Misra, P. (2018). Regression type double sampling estimator of population mean using auxiliary information. *Journal of Reliability and Statistical Studies*, *11*(1), 21–28.

Muili, J. O., Zakari, Y., & Audu, A. (2019). Modified class of estimator of finite population variance. *FUDMA Journal of Sciences, 3*(4), 67-78.

Murthy, M. N. (1964). Product method of estimation. *Sankhyā: The Indian Journal of Statistics, Series A, 26*(1), 69-74.

Neyman, J. (1938). Contribution to the theory of sampling human populations. *Journal of the American Statistical Association, 33*(201), 101-116.

Ozgul, N., & Cingi, H. (2014). A new class of exponential regression cum ratio estimator in two-phase sampling. *Hacettepe Journal of Mathematics and Statistics, 43*(1), 131-140.

Rao, J. N. K. (1973). On double sampling for stratification and analytical surveys. *Biometrika, 60*(1), 125–133.

Rao, T. J. (1991). On certain methods of improving ratio and regression estimator. *Communication in statistics - Theory and Methods, 20*(10), 3325-3340.

Reddy, V. N. (1973). On ratio and product method of estimation. *Sankhyā: The Indian Journal of Statistics, Series B, 35*(3), 307-316.

Riaz, N., Noo-ul-Amin, M., & Hanif, M. (2014). Regression-cum-exponential ratio type estimators for the population mean. Middle-East. *Journal of Scientific Research*, *19*(12), 1716-1721.

Robson, D. S. (1957). Application of multivariate polykays to the theory of unbiased ratio-type estimation. *Journal of the American Statistical Association, 52*(280), 511-522.

Sahai, A. (1979). An efficient variant of the product and ratio estimators. *Statistica Neerlandica*, *33*(1), 27-35.

Sahoo, J., & Sahoo, L. N. (1993). A class of estimators in two-phase sampling using two auxiliary variables. *Journal of the Indian Statistical Association, 31,* 107–114.

Sahoo, J., Sahoo, L. N., & Mohanty, S. (1993). A regression approach to estimation in two-phase sampling using two auxiliary variables. *Current Science*, *65*(1), 73–75.

Sahoo, J., Sahoo, L. N., & Mohanty, S. (1994). An alternative approach to estimation in two-phase sampling using two auxiliary variables. *Biometrical Journal, 36*(3), 293–298.

Samiuddin, M., & Hanif, M. (2007). Estimation of population mean in single and two phase sampling with or without additional information. *Pakistan Journal of Statistics*, *23*(2), 99-118.

Sanaullah, A., Ali, H. A., Amin, M. N., & Hanif, M. (2014). Generalized exponential chain ratio estimators under stratified two-phase random sampling. *Applied Mathematics and Computation, 226,* 541–547.

Shabbir, J., & Gupta, S. (2010). A new difference-cum-exponential type estimator of finite population mean in simple random sampling. *Colombian Journal of Statistics, 37*(1), 199-211.

Shabbir, J., & Gupta, S. (2017). On generalized exponential chain ratio estimators under stratified two-phase random sampling. *Communications in Statistics - Theory and Methods*, *46*(6), 2910–2920.

Singh, M. P. (1965). On the estimation of ratio and product of population parameters. *Sankhyā: The Indian journal of Statistics, Series B*, 27(3/4), 321-328.

Singh, R., Chauhan, P., & Sawan, N. (2011). Improvement in estimating population mean using two auxiliary variables in two-phase sampling. *Italian Journal of Pure and Applied Mathematics, 28,* 135–142.

Singh, B. K., & Choudhury, S. (2012). Dual to product estimator for estimating population mean in double sampling. *International Journal of Statistics and Systems, 7*(1), 31–39.

Singh, H.P., Tailor, R., Singh, S., & Kim, J.M (2008). A modified estimator of population mean using power of transformation. *Statistical Papers, 49*(1), 37-58.

Singh, H. P., & Ruiz-Espejo, M. (2003). On linear regression and ratio-product estimation of a finite population mean. *The Statistician, 52*(1), 59-67.

Singh, G. N., & Upadhyaya, L. N. (1995). A class of modified chain type estimators using two auxiliary variables in two-phase sampling. *Metron, 53* (3–4), 117–25.

Singh, H. P., & Vishwakarma, G. K. (2007). Modified exponential ratio and product estimators for finite population mean in two phase sampling. *Austrian Journal of Statistics, 36*(3), 217-225.

Srivastava, S. K. (1980). A class of estimator using auxiliary information in sample surveys. *Canadian Journal of Statistics, 8*(2), 253-254.

Srivastava, S. K. (1981). A generalized two-phase sampling estimator. *Journal of Indian Society of Agricultural Statistics*, *33*(1), 37-46.

Srivenkataramana, T. (1980). A dual to ratio estimator in sample surveys. *Biometrika*, *67*(1), 199-204.

Sukhatme, B. V. (1962). Some ratio-type estimators in two-phase sampling. *Journal of the American Statistical Association*, *57*(299), 628–632.

Upadhyaya, L. N., & Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, *41*(5), 627-636.

Walsh, J. E. (1970). Generalization of ratio estimates for population total. *Sankhyā: The India Journal of Statistics, Series A*, *32*(1), 99-106.

Yadav, S. K., Gupta, S., Mishra, S. S., & Shukla, A. K. (2016). Modified ratio and product estimators for estimating population mean in two-phase sampling. *American Journal of Operational Research, 6*(3), 61–82.

Yahaya, A., & Kabir, U. A. (2017). A modified ratio-product estimator of population mean in the presence of median and coefficient of variation of the auxiliary variable in stratified random sampling. *Bayero Journal of Pure and Applied Sciences, 10*(1), 6 – 17.

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., Muili, J. O., Tela, M. N., Danchadi, N. S., & Audu, A. (2020). Use of unknown weight to enhance ratio-type estimator in simple random sampling. *Lapai Journal of Applied and Natural Sciences, 5*(1), 74-81.

APPENDIX

R-CODES FOR SIMULATION STUDY

```
$Models for response variables
$ y=x+e
$y=x∧2+x+e
$Models for predictors
$ x¬beta(1000,1,4)
$ x¬gamma(1000, 1,4)
library(MASS)
x=rbeta(1000,1,4)
e=rnorm(1000,0,4)
w=0.1
y1=x∧2+x+e
yx=as.matrix(cbind(y1,x))
n=100
n1=200
```

```
N=1000
l=1/n-1/N
l1=1/n1-1/N
Cx=sd(x)/mean(x)
Cy=sd(y1)/mean(y1)
rho=cor(x,y1)
Yb=mean(y1)
Xb=mean(x)
R=Yb/Xb
a=1
b=0
tt=a*Xb/(2*(a*Xb+b))
k1=1-(2-l1*tt∧2*Cx∧2)/(1+(l-l1*rho))
k2=R*(tt-1+(2-l1*tt∧2*Cx∧2)*(2*tt-rho*Cy*Cx)/(1+(l-l1*rho)))
m1=(1+23*(l-l1)*Cx∧2/8-(4*l-3*l1)*rho*Cy*Cx/2)
m2=1+l*Cy∧2+(l-l1)*Cx∧2/8-(l-9*l1)*rho*Cy*Cx-(l*rho*Cy)∧2/(l-l1)
p1=m1/m2
p2=R*(0.5-p1*(2-l*rho*Cy/((l-l1)*Cx)))
bt0=NA; bt1=NA; bt2=NA; bt3=NA; bt4=NA; bt5=NA;
mt0=NA; mt1=NA; mt2=NA; mt3=NA; mt4=NA; mt5=NA;
for(i in 1:1000){
smp1=c(sample(1:1000,n1,replace=F))
smp2=c(sample(1:200,n,replace=F))
mar1=yx[smp1,]
mar2=mar1[smp2,]
yy=mar2[,1];xx=mar2[,2];xx1=mar1[,2];
yb=mean(yy);xb=mean(xx);xb1=mean(xx1);
zb=a*xb+b;zb1=a*xb1+b;
rho1=cor(xx,yy);
sy=sd(yy)
sx=sd(xx)
brg=rho1*sy/sx;
```

```
bt0[i]=yb-Yb;
bt1[i]=yb*xb1/xb-Yb;
bt2[i]=yb+brg*(xb1-xb)-Yb;
bt3[i]=yb*exp((xb1-xb)/(xb1+xb))-Yb;
bt4[i]=(k1*yb+k2*(xb1-xb))*exp((zb1-zb)/(zb1+zb))-Yb;
bt5[i]=(p1*yb*xb1/xb+p2*(xb1-xb))*exp((xb1-xb)/(xb1+xb))-Yb;
mt0[i]=(yb-Yb)∧2;
mt1[i]=(yb*xb1/xb-Yb)∧2;
mt2[i]=(yb+brg*(xb1-xb)-Yb)∧2;
mt3[i]=(yb*exp((xb1-xb)/(xb1+xb))-Yb)∧2;
mt4[i]=((k1*yb+k2*(xb1-xb))*exp((zb1-zb)/(zb1+zb))-Yb)∧2;
mt5[i]=((p1*yb*xb1/xb+p2*(xb1-xb))*exp((xb1-xb)/(xb1+xb))-Yb)∧2;
}
bb0=round(mean(bt0),5)
bb1=round(mean(bt1),5)
bb2=round(mean(bt2),5)
bb3=round(mean(bt3),5)
bb4=round(mean(bt4),5)
bb5=round(mean(bt5),5)
mm0=round(mean(mt0),5)
mm1=round(mean(mt1),5)
mm2=round(mean(mt2),5)
mm3=round(mean(mt3),5)
mm4=round(mean(mt4),5)
mm5=round(mean(mt5),5)
pr0=round(mm0/mm0*100,2)
pr1=round(mm0/mm1*100,2)
pr2=round(mm0/mm2*100,2)
pr3=round(mm0/mm3*100,2)
pr4=round(mm0/mm4*100,2)
pr5=round(mm0/mm5*100,2)
bb0;bb1;bb2;bb3;bb4;bb5;
```

mm0;mm1;mm2;mm3;mm4;mm5;
pr0;pr1;pr2;pr3;pr4;pr5;