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

Isah Muhammad¹, Yahaya Zakari² and Ahmed Audu³

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

¹Department of Statistics, Binyaminu Usman Polytechnic, Hadejia, Nigeria
²Corresponding Author: Department of Statistics, Ahmadu Bello University, Zaria, Nigeria, yahzaksta@gmail.com, Tel: +2348169766242
³Department 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 Samiuuddin 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>
<thead>
<tr>
<th>Parameters</th>
<th>Population I</th>
<th>Population II</th>
<th>Population III</th>
<th>Population IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>923</td>
<td>89</td>
<td>104</td>
<td>80</td>
</tr>
<tr>
<td>( n' )</td>
<td>400</td>
<td>30</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>( n )</td>
<td>200</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>( \rho_{yx} )</td>
<td>0.955</td>
<td>3.360</td>
<td>0.865</td>
<td>51.862</td>
</tr>
<tr>
<td>( \bar{y} )</td>
<td>436.3</td>
<td>0.124</td>
<td>625.37</td>
<td>11.265</td>
</tr>
<tr>
<td>( \bar{x} )</td>
<td>11440.50</td>
<td>0.604</td>
<td>13.930</td>
<td>0.354</td>
</tr>
<tr>
<td>( C_y )</td>
<td>1.72</td>
<td>2.190</td>
<td>1.866</td>
<td>0.751</td>
</tr>
<tr>
<td>( C_x )</td>
<td>1.86</td>
<td>0.766</td>
<td>1.653</td>
<td>0.9413</td>
</tr>
</tbody>
</table>

3.2 Notations and Existing Estimators

Consider a finite population, \( U = U_1, \ldots, U_N \), of size \( N \) units. Let \( y \) denote the study variable taking the values \( y_i \) on the unit \( U_i, (i = 1, \ldots, N) \) and \( Y \) represents unknown population mean. Also, let \( x \) be the auxiliary variable that takes the values \( x_i \) on the unit \( U_i, (i = 1, \ldots, N) \) positively related to \( Y \) and \( \bar{X} \) 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).


\[
\bar{y} = \sum_{i \in S} y_i / n, \quad \bar{x} = \sum_{i \in S} x_i / n \quad \text{and} \quad \bar{x}' = \sum_{i \in S'} x_i / n'
\]
An Alternative Class of Ratio-Regression-Type Estimator under Two-Phase Sampling Scheme Muhammad et al.

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 $\bar{Y}$ 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 $\bar{Y}$ and $\bar{X}$ is unknown population mean; $\bar{x}'$ denotes the primary sample mean of the auxiliary variable in the first phase sample of size $n'$; $\bar{x}$ denotes the sub-sample mean of the auxiliary variables in the second phase sample of size $n$ and $\bar{y}$ 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:

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

$$S_y^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{N-1}, \quad S_x^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}, \quad \text{and} \quad S_{yx} = \frac{\sum_{i=1}^{N} (y_i - \bar{y})(x_i - \bar{x})}{N-1},$$

where $n'$ is the primary sample size; $n$ is the sub-sample size; $N$ is the number of units in the population; $\bar{Y}$ is the population mean of the study variable; $\bar{X}$ is the population mean of the auxiliary variable; $\lambda$ is a known constant involving the samples and population units; $\rho_{yx}$ 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:

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

(1)

where $\bar{x}'$ is the primary sample mean of the auxiliary variable, $\bar{y}$ and $\bar{x}$ 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:

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

(2)
Substituting regression estimator in (1), Ozgul and Cingi (2014) proposed a class of exponential regression cum ratio estimator as follows:

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

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:

\[ k_{1(\text{opt})} = 1 - \frac{2 - \lambda' \theta^2 C_x^2}{[1 + (\lambda - \lambda')^2]} \]
\[ k_{2(\text{opt})} = \frac{\hat{y}}{\bar{x}} \left[ (\theta - 1) + \frac{2 - \lambda' \theta^2 C_x^2}{[1 + (\lambda - \lambda')^2]} (2\theta - K_{yx}) \right] \]

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

\( Y \) is the population mean of the study variable; \( X \) is the population mean of the auxiliary variable; \( \lambda \) and \( \lambda' \) are a known constant involving the samples and population units; \( \rho_{yx} \) is the population correlation coefficient between the auxiliary and the study variables; \( C_x \) is the population coefficients of variation of the auxiliary variable; \( \theta \) is known constant; \( \text{Var}(\bar{y}_{lr}) \) 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, \( \alpha_1 \), and regression with coefficient, \( \alpha_2 \), as \( \left[ \alpha_1 \bar{y} \left( \frac{\bar{x}}{\bar{x}} \right) + \alpha_2 (\bar{x}' - \bar{x}) \right] \) 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

\[ \text{MSE} \left( \hat{y}_{NH} \right)^2 \frac{\text{Var}(\bar{y}_{lr})(1 - \lambda' \theta^2 C_x^2) - \frac{\lambda'^2 \hat{y}^2 \theta^4 C_x^4}{4}}{[\hat{y}^2 + \text{Var}(\bar{y}_{lr})]} \min \]
using the auxiliary information in two-phase sampling as:

\[
\hat{y}_{IM} = \left[ \alpha_1 \bar{Y} \left( \frac{\bar{x}'}{\bar{x}} \right) + \alpha_2 (\bar{x}' - \bar{x}) \right] \exp \left( \frac{\bar{x}' - \bar{x}}{\bar{x} + \bar{x}} \right)
\]

(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 \( \alpha_1 \) and \( \alpha_2 \) 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:

\[
\bar{y} = \bar{Y} (1 + e_0), \quad \bar{x} = \bar{X} (1 + e_1), \quad \text{and} \quad \bar{x}' = \bar{X} (1 + e_1')
\]

such that

\[
E(e_0) = E(e_1) = E(e_1') = 0
\]

\[
E(e_0^2) = \lambda C_y^2, \quad E(e_1^2) = \lambda C_x^2, \quad E(e_1'^2) = \lambda' C_x^2
\]

\[
E(e_0 e_1) = \lambda \rho C_y C_x, \quad E(e_0 e_1') = \lambda' \rho C_y C_x, \quad E(e_1 e_1') = \lambda' C_x^2
\]

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

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

(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:
\[ \hat{y}_{IM} = \left[ a_1 \hat{Y} (1 + e_0 + e'_1 + e_0 e'_1)(1 - e_1 + e_1^2) + a_2 \hat{X} (e'_1 - e_1) \right] \]
\[ \exp \left( \frac{a_3 \hat{X} (e'_1 - e_1)}{2(a_3 \hat{X} + b_3) + a_3 \hat{X} (e'_1 + e_1)} \right) \]  
(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:

\[ \hat{y}_{IM} = \left[ \alpha_1 \hat{Y} (1 + e_0 - e_1 + e'_1 + e_1^2 - e_0 e_1 + e_0 e'_1 - e_1 e'_1) + \alpha_2 \hat{X} (e'_1 - e_1) \right] \]
\[ \exp \left( \frac{\theta (e'_1 - e_1)}{2 \left( 1 + \frac{\theta (e'_1 + e_1)}{2} \right)} \right) \]  
(8)

where \( \theta = \frac{a_3 \hat{X}}{a_3 \hat{X} + b_3} \)

Taking out the common terms in the exponential part of (8), we get:

\[ \hat{y}_{IM} = \left[ a_1 \hat{Y} (1 + e_0 - e_1 + e'_1 + e_1^2 - e_0 e_1 + e_0 e'_1 - e_1 e'_1) + a_2 \hat{X} (e'_1 - e_1) \right] \]
\[ \exp \left( \frac{\theta (e'_1 - e_1)}{2} \left[ 1 + \frac{\theta (e'_1 + e_1)}{2} \right]^{-1} \right) \]  
(9)

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

\[ \approx \left( \alpha_1 \hat{Y} - \hat{Y} - (\alpha_1 \hat{Y} + \alpha_2 \hat{X}) e_1 + \alpha_1 \hat{Y} e_1^2 + (\alpha_1 \hat{Y} + \alpha_2 \hat{X}) e'_1 - \alpha_1 \hat{Y} e_1 e'_1 + ... \right) \]
\[ (... \alpha_1 \hat{Y} e_0 - \alpha_1 \hat{Y} e_0 e_1 + \alpha_1 \hat{Y} e_0 e'_1) \times \left( 1 + \left( \frac{\theta}{2} \right) e'_1 - \frac{\theta e_1}{8} - \frac{\theta^2 e'_1^2}{8} + ... \right) \]
\[ \left( \frac{3 \theta^2 e_1^2}{8} - \frac{\theta^2 e_1 e'_1}{4} + ... \right) \]  
(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:

\[
(y_{IM}) \approx \bar{y} + \bar{y}(\alpha_1 - 1) + e_1 \bar{y}e_0 + \alpha'_i \left( \alpha_1 \bar{y} \frac{\theta}{2} + (\alpha_1 \bar{y} + \alpha_2 \bar{x}) \right) - e_1
\]

\[
\left( \bar{y} \alpha_1 \frac{\theta}{2} + (\alpha_1 \bar{y} + \alpha_2 \bar{x}) \right) + e_1' \left( \frac{\theta}{2} (\alpha_1 \bar{y} + \alpha_2 \bar{x}) - \bar{y} \alpha_1 \frac{\theta^2}{8} \right) + e_1^2
\]

\[
3 \bar{y} \frac{\theta^2}{8} + \frac{\theta}{2} (\alpha_1 \bar{y} + \alpha_2 \bar{x}) + \alpha_1 \bar{y} \right) - e_1 e_1' \left( \bar{y} \alpha_1 \frac{\theta^2}{4} + (\alpha_1 \bar{y} + \alpha_2 \bar{x}) \theta + \alpha_1 \bar{y} \right)
\]

\[
+ e_0 e_1' \left( \alpha_1 \bar{y} \frac{\theta}{2} + \alpha_1 \bar{y} \right) - e_0 e_1 \left( \alpha_1 \bar{y} \frac{\theta}{2} + \alpha_1 \bar{y} \right)
\]  

(11)

\[
(y_{IM} - \bar{y}) \approx \bar{y}(\alpha_1 - 1) + e_1 \bar{y}e_0 + \alpha'_i \left( \alpha_1 \bar{y} \frac{\theta}{2} + (\alpha_1 \bar{y} + \alpha_2 \bar{x}) \right) - e_1
\]

\[
\left( \bar{y} \alpha_1 \frac{\theta}{2} + (\alpha_1 \bar{y} + \alpha_2 \bar{x}) \right) + e_1' \left( \frac{\theta}{2} (\alpha_1 \bar{y} + \alpha_2 \bar{x}) - \bar{y} \alpha_1 \frac{\theta^2}{8} \right) + e_1^2
\]

\[
3 \bar{y} \frac{\theta^2}{8} + \frac{\theta}{2} (\alpha_1 \bar{y} + \alpha_2 \bar{x}) + \alpha_1 \bar{y} \right) - e_1 e_1' \left( \bar{y} \alpha_1 \frac{\theta^2}{4} + (\alpha_1 \bar{y} + \alpha_2 \bar{x}) \theta + \alpha_1 \bar{y} \right)
\]

\[
+ e_0 e_1' \left( \alpha_1 \bar{y} \frac{\theta}{2} + \alpha_1 \bar{y} \right) - e_0 e_1 \left( \alpha_1 \bar{y} \frac{\theta}{2} + \alpha_1 \bar{y} \right)
\]  

(12)

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

\[
Bias(y_{IM}) = \bar{y}(a_1 - 1) + (\lambda - \lambda') C \beta^2 \left( \frac{3 \theta^2}{8} \bar{y} + \frac{\theta}{2} (\alpha_1 \bar{y} + \alpha_2 \bar{x}) + \alpha_1 \bar{y} \right)
\]

\[
+ (\lambda - \lambda') \left( \bar{y} \alpha_1 \frac{\theta}{2} + \alpha_1 \bar{y} \right) \rho C_y C_x
\]  

(13)

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

\[
MSE(y_{IM}) = \bar{y}^2 + a_1 \bar{w}_1 - a_1 \bar{w}_2 + a_2 \bar{w}_3 - a_2 \bar{w}_4 + 2a_1 a_2 \bar{w}_5
\]  

(14)
where:

\[ W_1 = \bar{Y}^2 (1 + (\lambda - \lambda') C_x^2 (\theta^2 + 2\theta + 2) + \lambda C_y^2 - 2(\lambda - \lambda')\rho C_y C_x (\theta + 1)) \]

\[ W_2 = \bar{Y}^2 \left( 2 + (\lambda - \lambda') C_x^2 \left( 3 + \frac{3\theta^2}{4} \right) - (\lambda - \lambda')\rho C_y C_x (\theta + 2) \right) \]

\[ W_3 = (\lambda - \lambda') \tilde{X} C_x^2 \]

\[ W_4 = \bar{Y} \tilde{X} \theta C_x^2 (\lambda - \lambda') \]

\[ W_5 = (\lambda - \lambda') \left( C_x^2 \bar{Y} \tilde{X} (\theta + 1) + \bar{Y} \tilde{X} \rho C_y C_x \right) \]

Differentiating (14) partially with respect to \( \alpha_1 \) and \( \alpha_2 \), and equating to zero, we obtain the optimum values of \( \alpha_1 \) and \( \alpha_2 \), respectively as:

\[ \alpha_1 = \frac{W_2 W_3 - W_4 W_5}{2 (W_1 W_3 - W_2^2)} \]  \hspace{1cm} (15)

and

\[ \alpha_2 = \frac{W_1 W_4 - W_2 W_5}{2 (W_1 W_3 - W_5)} \]  \hspace{1cm} (16)

Substituting the \( \alpha_1 \) and \( \alpha_2 \) optimum values into (14), we obtain the minimum mean square error of the proposed estimator as:

\[ MSE(\tilde{y}_M)^2 \left( \frac{W_2^2 W_3 + W_1 W_4^2 - 2W_2 W_4 W_5}{4 (W_1 W_3 - W_5^2)} \right)_{min} \]  \hspace{1cm} (17)
An Alternative Class of Ratio-Regression-Type Estimator under Two-Phase Sampling Scheme

Muhammad et al.

Table 2: Some members of the suggested estimator, \( \hat{y}_{1M} \)

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Values of ( a_x )</th>
<th>Values of ( b_x )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{t}_{CE1} = \alpha_t \frac{\hat{y}}{x} + \alpha_e (\hat{x} - \bar{x}) ) exp ( \frac{\bar{y} - \bar{x}}{\bar{x} - \bar{x}} )</td>
<td>1 ( \beta_2(\phi) )</td>
<td></td>
</tr>
<tr>
<td>( \hat{t}_{CE1} = \alpha_t \frac{\hat{y}}{x} + \alpha_e (\hat{x} - \bar{x}) ) exp ( \frac{\bar{y} - \bar{x}}{\bar{x} + x + 2\beta_2(\phi)} )</td>
<td>1 ( C_x )</td>
<td></td>
</tr>
<tr>
<td>( \hat{t}_{CE2} = \alpha_t \frac{\hat{y}}{x} + \alpha_e (\hat{x} - \bar{x}) ) exp ( \frac{\bar{y} - \bar{x}}{\bar{x} + 2x} )</td>
<td>1 ( \beta_2(\phi) )</td>
<td></td>
</tr>
<tr>
<td>( \hat{t}<em>{CE3} = \alpha_t \frac{\hat{y}}{x} + \alpha_e (\hat{x} - \bar{x}) ) exp ( \frac{\bar{y} - \bar{x}}{\bar{x} + x + 2\rho</em>{xy}} )</td>
<td>1 ( \rho_{xy} )</td>
<td></td>
</tr>
<tr>
<td>( \hat{t}_{CE4} = \alpha_t \frac{\hat{y}}{x} + \alpha_e (\hat{x} - \bar{x}) ) exp ( \frac{\beta_2(\phi)(\bar{y} - \bar{x})}{\bar{x} + x + 2\beta_2(\phi)} )</td>
<td>( \beta_2(\phi) ) ( C_x )</td>
<td></td>
</tr>
<tr>
<td>( \hat{t}<em>{CE5} = \alpha_t \frac{\hat{y}}{x} + \alpha_e (\hat{x} - \bar{x}) ) exp ( \frac{\beta_2(\phi)(\bar{y} - \bar{x})}{\bar{x} + x + 2\rho</em>{xy}} )</td>
<td>( \beta_2(\phi) ) ( \rho_{xy} )</td>
<td></td>
</tr>
<tr>
<td>( \hat{t}<em>{CE6} = \alpha_t \frac{\hat{y}}{x} + \alpha_e (\hat{x} - \bar{x}) ) exp ( \frac{\beta_2(\phi)(\bar{y} - \bar{x})}{\bar{x} + x + 2\rho</em>{xy}} )</td>
<td>( \beta_2(\phi) ) ( \rho_{xy} )</td>
<td></td>
</tr>
<tr>
<td>( \hat{t}<em>{CE7} = \alpha_t \frac{\hat{y}}{x} + \alpha_e (\hat{x} - \bar{x}) ) exp ( \frac{\beta_2(\phi)(\bar{y} - \bar{x})}{\bar{x} + x + 2\rho</em>{xy}} )</td>
<td>( \beta_2(\phi) ) ( C_x )</td>
<td></td>
</tr>
<tr>
<td>( \hat{t}<em>{CE8} = \alpha_t \frac{\hat{y}}{x} + \alpha_e (\hat{x} - \bar{x}) ) exp ( \frac{\beta_2(\phi)(\bar{y} - \bar{x})}{\bar{x} + x + 2\rho</em>{xy}} )</td>
<td>( \beta_2(\phi) ) ( C_x )</td>
<td></td>
</tr>
</tbody>
</table>

where \( \rho_{xy} \) 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:

\[
MSE(\hat{y}_{1M}) \leq MSE(\hat{y}_{Rmin}) \leq \hat{y}^2 \left[ \lambda C_y^2 + \lambda' (C_x^2 - 2\rho_{xy} C_x C_y) \right] - \hat{y}^2 + \frac{A}{B} > 0
\]
where:

\[ A = W_2^2W_3 + W_1W_4^2 - 2W_2W_4W_5 \]

\[ B = 4 \left( W_1W_3 - W_5^2 \right) \]

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:

\[
\begin{align*}
MSE(\hat{y}_{IM})_{Y\text{min}}; \\
\hat{y}^2 - \frac{A}{B} < \hat{y}^2 C_y^2 [\lambda - \lambda' \rho_{yx}^2] \\
\hat{y}^2 C_y^2 [\lambda - \lambda' \rho_{yx}^2] - \hat{y}^2 + \frac{A}{B} > 0
\end{align*}
\]

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:

\[
\begin{align*}
MSE(\hat{y}_{IM})_{Y_{svr\text{min}}}; \\
\hat{y}^2 - \frac{A}{B} < \hat{y}^2 \left( \lambda C_y^2 + \lambda' \left( \frac{C_x^2}{4} - \rho_{yx} C_y C_x \right) \right) \\
\hat{y}^2 \left( \lambda C_y^2 + \lambda' \left( \frac{C_x^2}{4} - \rho_{yx} C_y C_x \right) \right) - \hat{y}^2 + \frac{A}{B} > 0
\end{align*}
\]

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:

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

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:

\[
\text{bias}(T) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{Y} - \bar{Y}), \quad \text{MSE}(T) = \frac{1}{1000} \sum_{i=1}^{1000} \left(\hat{Y} - \bar{Y}\right)^2 \quad \text{and}
\]

\[
\text{PRE}(T) = \frac{\text{MSE}(T)}{\text{MSE}(y)} \times 100, \quad \text{where} \quad T \text{ is either of } y, \bar{y}_{Rv}, \bar{y}_{svr}, \bar{y}_{NH} \text{ and } \hat{y}_{IM}.
\]

<table>
<thead>
<tr>
<th>Table 3: Simulation Parameters and Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Bias</th>
<th>MSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Models 1: $Y = X + E$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample mean ($\bar{y}$)</td>
<td>0.00849</td>
<td>0.014011</td>
<td>100</td>
</tr>
<tr>
<td>Classical Ratio ($\bar{y}_R$)</td>
<td>0.00775</td>
<td>0.14079</td>
<td>99.52</td>
</tr>
<tr>
<td>Classical Regression ($\bar{y}_{lr}$)</td>
<td>0.00926</td>
<td>0.14059</td>
<td>99.66</td>
</tr>
<tr>
<td>Singh and Vishwakarma (2007) ($\bar{y}_{svr}$)</td>
<td>0.00808</td>
<td>0.14019</td>
<td>99.94</td>
</tr>
<tr>
<td>Ozgul and Cingi (2014) ($\bar{y}_{NH}$)</td>
<td>-0.20717</td>
<td>0.17811</td>
<td>78.66</td>
</tr>
<tr>
<td>Proposed Estimator ($\hat{y}_{IM}$)</td>
<td>-0.09962</td>
<td>0.00994</td>
<td>1409.56</td>
</tr>
</tbody>
</table>

Model 2: $Y = X^2 + X + E$

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Bias</th>
<th>MSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample mean ($\bar{y}$)</td>
<td>0.01532</td>
<td>0.13523</td>
<td>100</td>
</tr>
<tr>
<td>Classical Ratio ($\bar{y}_R$)</td>
<td>0.01543</td>
<td>0.13793</td>
<td>98.04</td>
</tr>
<tr>
<td>Classical Regression ($\bar{y}_{lr}$)</td>
<td>0.01453</td>
<td>0.137</td>
<td>98.71</td>
</tr>
<tr>
<td>Singh and Vishwakarma (2007) ($\bar{y}_{svr}$)</td>
<td>0.01533</td>
<td>0.13632</td>
<td>99.20</td>
</tr>
<tr>
<td>Ozgul and Cingi (2014) ($\bar{y}_{NH}$)</td>
<td>-0.16534</td>
<td>0.15851</td>
<td>85.31</td>
</tr>
<tr>
<td>Proposed Estimator ($\hat{y}_{IM}$)</td>
<td>-0.07571</td>
<td>0.00573</td>
<td>2360.03</td>
</tr>
</tbody>
</table>

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 underlying 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

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Bias</th>
<th>MSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: $Y = X + E$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample mean ($\bar{y}$)</td>
<td>-0.00129</td>
<td>0.14824</td>
<td>100</td>
</tr>
<tr>
<td>Classical Ratio ($\bar{y}_R$)</td>
<td>-0.00342</td>
<td>0.15007</td>
<td>98.78</td>
</tr>
<tr>
<td>Classical Regression ($\bar{y}_{lr}$)</td>
<td>1e-04</td>
<td>0.14771</td>
<td>100.36</td>
</tr>
<tr>
<td>Singh and Vishwakarma (2007) ($\bar{y}_{svr}$)</td>
<td>-0.00244</td>
<td>0.14873</td>
<td>99.67</td>
</tr>
<tr>
<td>Ozgul and Cingi (2014) ($\bar{y}_{NH}$)</td>
<td>-0.30886</td>
<td>0.23737</td>
<td>62.45</td>
</tr>
<tr>
<td>Proposed Estimator ($\bar{y}_{IM}$)</td>
<td>-0.15403</td>
<td>0.02375</td>
<td>624.17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Bias</th>
<th>MSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 2: $Y = X^2 + X + E$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample mean ($\bar{y}$)</td>
<td>0.01123</td>
<td>0.15975</td>
<td>100</td>
</tr>
<tr>
<td>Classical Ratio ($\bar{y}_R$)</td>
<td>0.00931</td>
<td>0.16072</td>
<td>99.40</td>
</tr>
<tr>
<td>Classical Regression ($\bar{y}_{lr}$)</td>
<td>0.01255</td>
<td>0.15802</td>
<td>101.09</td>
</tr>
<tr>
<td>Singh and Vishwakarma (2007) ($\bar{y}_{svr}$)</td>
<td>0.01006</td>
<td>0.15971</td>
<td>100.36</td>
</tr>
<tr>
<td>Ozgul and Cingi (2014) ($\bar{y}_{NH}$)</td>
<td>-0.71299</td>
<td>0.6604</td>
<td>24.19</td>
</tr>
<tr>
<td>Proposed Estimator ($\bar{y}_{IM}$)</td>
<td>-0.34176</td>
<td>0.1117</td>
<td>136.54</td>
</tr>
</tbody>
</table>

Table 5 shows the various estimators’ biases, MSEs and PREs values based on the population parameters obtained from linear and quadratic models when the underlying 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, 0.1117 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>
<thead>
<tr>
<th>Estimators</th>
<th>Population I</th>
<th></th>
<th>Population II</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>PRE</td>
<td>MSE</td>
<td>PRE</td>
</tr>
<tr>
<td>Sample mean ((\bar{y}))</td>
<td>2205.636</td>
<td>100</td>
<td>159654.6</td>
<td>100</td>
</tr>
<tr>
<td>Classical Ratio ((y_R))</td>
<td>944.1017</td>
<td>233.6227</td>
<td>680788.5</td>
<td>23.45142</td>
</tr>
<tr>
<td>Classical Regression ((y_{lr}))</td>
<td>921.6096</td>
<td>239.3243</td>
<td>119377.5</td>
<td>133.7393</td>
</tr>
<tr>
<td>Singh and Vishwakarma (2007) ((y_{svr}))</td>
<td>1163.268</td>
<td>189.6069</td>
<td>194613</td>
<td>82.03697</td>
</tr>
<tr>
<td>Ozgul and Cingi (2014) ((y_{NH}))</td>
<td>1142.667</td>
<td>193.0253</td>
<td>191126.3</td>
<td>83.53356</td>
</tr>
<tr>
<td>Proposed Estimator ((\hat{y}_{IM}))</td>
<td>914.667</td>
<td>241.1409</td>
<td>1651.32</td>
<td>9668.302</td>
</tr>
</tbody>
</table>

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

Muhammad et al.

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 $\bar{y}$

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Population III</th>
<th>Population IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>PRE</td>
</tr>
<tr>
<td>Sample mean ($\bar{y}$)</td>
<td>54993.75</td>
<td>100</td>
</tr>
<tr>
<td>Classical Ratio ($\bar{y}_R$)</td>
<td>29536.17</td>
<td>186.1912</td>
</tr>
<tr>
<td>Classical Regression ($\bar{y}_r$)</td>
<td>29521.36</td>
<td>186.2846</td>
</tr>
<tr>
<td>Singh and Vishwakarma (2007) ($\bar{y}_{svr}$)</td>
<td>35586.14</td>
<td>154.537</td>
</tr>
<tr>
<td>Ozgul and Cingi (2014) ($\bar{y}_{NH}$)</td>
<td>30361.55</td>
<td>181.1296</td>
</tr>
<tr>
<td>Proposed Estimator ($\bar{y}_{IM}$)</td>
<td>24711.69</td>
<td>22.5414</td>
</tr>
</tbody>
</table>

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 (24711.69 and 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


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


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

Muhammad et al.


**APPENDIX**

**R-CODES FOR SIMULATION STUDY**

$ y=x+e$

$y=x^2+x+e$

$\$Models for predictors

$ x\sim beta(1000,1,4)$

$ x\sim 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)))
b0=NA; b1=NA; b2=NA; b3=NA; b4=NA; b5=NA;
m0=NA; m1=NA; m2=NA; m3=NA; m4=NA; m5=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,]
ny=mar2[,1];nx=mar2[,2];nx1=mar1[,2];
yb=mean(yy);xb=mean(xx);xb1=mean(xx1);
z=xb+a*xb+b;z=xb1+a*xb1+b;
rho1=cor(xx,yy);
sy=sd(yy)
sx=sd(xx)
brg=rho1*sy/sx;
bt0[i] = \(yb - Yb\);
bt1[i] = \(yb \times x_{b1}/x_b - Yb\);
bt2[i] = \(yb + brg \times (x_{b1} - x_b) - Yb\);
bt3[i] = \(yb \times \exp((x_{b1} - x_b)/(x_{b1} + x_b)) - Yb\);
bt4[i] = \((k1 \times yb + k2 \times (x_{b1} - x_b)) \times \exp((z_{b1} - z_b)/(z_{b1} + z_b)) - Yb\);
bt5[i] = \((p1 \times yb \times x_{b1}/x_b + p2 \times (x_{b1} - x_b)) \times \exp((x_{b1} - x_b)/(x_{b1} + x_b)) - Yb\);

mt0[i] = \((yb - Yb)^2\);
mt1[i] = \((yb \times x_{b1}/x_b - Yb)^2\);
mt2[i] = \((yb + brg \times (x_{b1} - x_b) - Yb)^2\);
mt3[i] = \((yb \times \exp((x_{b1} - x_b)/(x_{b1} + x_b)) - Yb)^2\);
mt4[i] = \(((k1 \times yb + k2 \times (x_{b1} - x_b)) \times \exp((z_{b1} - z_b)/(z_{b1} + z_b)) - Yb)^2\);
mt5[i] = \(((p1 \times yb \times x_{b1}/x_b + p2 \times (x_{b1} - x_b)) \times \exp((x_{b1} - x_b)/(x_{b1} + x_b)) - Yb)^2\);

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 \times 100,2)
pr1 = round(mm0/mm1 \times 100,2)
pr2 = round(mm0/mm2 \times 100,2)
pr3 = round(mm0/mm3 \times 100,2)
pr4 = round(mm0/mm4 \times 100,2)
pr5 = round(mm0/mm5 \times 100,2)

bb0;bb1;bb2;bb3;bb4;bb5;
An Alternative Class of Ratio-Regression-Type Estimator under Two-Phase Sampling Scheme Muhammad et al.

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