

Hybrid Exponential Type Estimators of Finite Population Mean in Double Sampling

T. UBA

Department of Statistics, Joseph Sarwuan Tarka University Makurdi, Benue State, Nigeria

DOI: <https://doi.org/10.51584/IJRIAS.2025.10020041>

Received: 23 January 2025; Accepted: 03 February 2025; Published: 15 March 2025

ABSTRACT

In this study, two-auxiliary exponential type estimators of finite population mean in Two-phase are proposed. The proposed estimators are extension of (1) estimator finite population mean in SRS to Two-phase sampling. The study investigated efficiency of the proposed estimators by utilizing the ratio of bias to standard error (RBSE) as a proxy to examine confidence limits for estimates. The expressions for the bias and Mean Square Error (MSE) of the estimators were derived. A comprehensive simulation study was carried out to show the efficacy of the estimators as compared to conventional estimators using Coefficient of Variation as a performance measure. Furthermore, a small sample from real data set was utilized to validate the performance of proposed estimators under two varying correlation coefficients amongst variables in the parameter space. The results of both the simulation study and real life studies have shown that the proposed estimators were not only asymptotic, more efficient but produces estimates that are more precise than most of the existing estimators considered in this study.

Keywords: Bias, Correlation Coefficient, Asymptotic Relative Efficiency, Ratio of Bias to Standard Error, Regression Estimators.

INTRODUCTION

In survey sampling, estimators are widely used within the design-based mode of inference both at the design and estimation stages for large-scale surveys (2). There are situations whereby single-auxiliary estimators of finite population characteristics, may not be efficient (3). In such situations there may be need to utilize multi-auxiliary variables that are correlated with the study variable in order to improve the sampling design and estimation. It may also be of interest to combine sampling techniques in a single survey in order to obtain the desired information, (4).

Basically, is known that the use of auxiliary information in survey sampling improves the precision of estimates of population parameters such as population total, population mean, population proportion, and so on (5). The usual estimators that use auxiliary information are the traditional ratio, product and regression estimators respectively as well as their improved or hybrid versions. When there is a positive correlation between the study variable and the auxiliary variable, the use of the classical ratio estimator is appropriate. On the other hand, if the correlation is negative, classical product estimator is preferable. Conversely, the regression estimator is used to estimate the population characteristics of interest when the regression line of the study variable and the auxiliary variable does not pass through the origin, but makes an intercept along the y-axis.

According to (6), John Graunt is believed to be the first to use auxiliary information in the 1660s for estimating the population England. Detailed work on auxiliary information were established in the work by (7), not forgetting the work on the classical ratio estimator by (8). (9) noted that the estimation of population parameters with greater precision is a relentless matter in sampling theory and that the precision of estimates can be improved by increasing the sample size but at the expense of the benefits of sampling hence the use of

auxiliary information. In many practical situations, the regression line does not pass through the origin. As such the ratio and product estimators do not perform equally well as compared to the regression estimator. For this reason, many authors such as (10), (11), (12) among many others have suggested modified ratio estimators in order to provide better alternatives. Recently, the use of auxiliary variable has enjoyed a boost due to exponentiation of the auxiliary variable in the work of (13), (14), (15), and so on. (16) introduced modifications of the exponential ratio estimator by using some transformations of the auxiliary variables. Exponentiation is widely applicable to ratio, product and regression estimators or a mixture (hybrid) of these, and appears to improve the precision of parameter estimates even when the relation between the study variable and the auxiliary variable is weak.

Under many sampling schemes, several exponential estimators, using the population information of the auxiliary variable, have been developed. However, if the information on the population mean of auxiliary variable is not readily available, two-phase (double) sampling method is the most appropriate sampling scheme to be adopted. It is cost effective, and is typically used when it is expensive to collect data on the variable of interest but relatively cheaper to collect variables that are correlated. If the population mean, \bar{X} of the auxiliary variable x is not known prior to the commencement of the survey, then it is appropriate to employ the two-phase sampling scheme. The following authors have proposed improved versions of the ratio exponential, regression exponential, or a regression-cum-ratio exponential type estimators using one or more auxiliary variables under double sampling, and these include the works by (17), (18). Taking a cue from the work of (19), (20) developed a computational approach to generalized estimator of population mean in double sampling using correlation coefficient between the study variable and the auxiliary variable. This generalized ratio estimator is a product of the mean of the study variable, the ratio estimator and suitable chosen constants such that their sum equals unity. This generalized estimator performed better than the sample mean, the classical ratio and the (19) estimators respectively, in terms of minimum mean square error. Another interesting hybrid estimator of finite population mean under Two-phase sampling is that of (21). This Regression-cum-ratio estimator takes advantage of the fact that it is easy to construct regression equation of y against x or y against z and using the known correlation coefficients between the auxiliary variables as additive factor to the second variable. The study examines the possibilities of different forms of auxiliary information derived from the remote sensing data in two phase sampling design and suggest an appropriate estimator that will be of broad interest and application.

Recently, (15) constructed an estimator of finite population mean in Two-phase sampling for stratification using the idea of chain-ratio estimators. The new improved estimator performed better than some selected unbiased standard estimators both in theoretical, natural and simulated populations. Similarly, (22) developed an efficient exponential chain ratio estimator for estimating finite population mean of the study character using imputation of missing data under two-phase sampling. While (23) proposed mixed estimators of finite population mean in Two-phase sampling using two auxiliary variables, they cautioned that purpose of mixture and conditions for which the individual estimators are developed must be stated and efficiency must be the target. (24) in their study proposed ratio-in-regression estimator of the mean using mean of the study and the auxiliary variables together with the rank of the auxiliary variable. In Two-phase sampling.

Apart from using the usual MSE of an estimator, which is the sum of the variance and the square of the bias (measures both accuracy and precision), the study deploy the ratio of bias to standard error (RBSE) to ensure that the best estimators are not only those whose estimates minimum MSEs or CVs but also are within confidence limits. The study also investigated asymptotic properties of the proposed estimators.

MATERIALS AND METHODS

Description of two-phase sampling design

Consider a finite population of N units, y_i , x_i and z_i ($i = 1, 2, \dots, n$), denote the values of the i^{th} unit of the characters Y, X and Z respectively. Here, y is our variable of interest (study variable), x and z are the auxiliary

variables that can be correlated with our variable of interest. Let S' be the first phase sample of size n' from the population of size N according to a simple random sampling without replacement (*srswor*) scheme, and \bar{x}' , \bar{z}' , the sample means of the two auxiliary variables are observed. Let S be the second phase sample of size n from the first phase sample ($n < n'$), and y , x , and z are observed. We also define: ρ_{HHH} : The parameter subspace, ($0.7 < \rho_1, \rho_2, \rho_3 < 1$) and ρ_{HLL} : the parameter subspace, ($0 < \rho_2, \rho_3 < 0.5 \cup 0.7 < \rho_1 < 1$), where, ρ_1 : Correlation coefficient between Y and Z , ρ_2 : Correlation coefficient between X and Z , ρ_3 : Correlation coefficient between Y and X .

Expectation under Two-Phase Sampling

We define the relative error in the first phase as follows:

$$e_{x'} = \frac{\bar{x}' - \bar{X}}{\bar{X}} \text{ and } e_{z'} = \frac{\bar{z}' - \bar{Z}}{\bar{Z}} \quad (1)$$

Thus,

$$\bar{x}' = (1 + e_{x'})\bar{X} \text{ and } \bar{z}' = (1 + e_{z'})\bar{Z} \quad (2)$$

$$\begin{aligned} E(e_x) = E(e_y) = E(e_z) = E(e_{x'}) = E(e_{z'}) = 0; E(e_y^2) = \theta_2 C_y^2; E(e_{x'}^2) = \theta_1 C_x^2; E(e_z^2) = \theta_2 C_z^2; E(e_{z'}^2) \\ = \theta_1 C_z^2; E(e_z^2) = \theta_2 C_z^2; E(e_y e_{x'}) = \theta_1 \rho_{yx} C_x C_y; E(e_y e_x) = \theta_2 \rho_{yx} C_x C_y; E(e_y e_{z'}) \\ = \theta_1 \rho_{yz} C_y C_z; E(e_y e_z) = \theta_2 \rho_{yz} C_y C_z; E(e_x e_{x'}) = \theta_1 C_x^2; E(e_{x'} e_{z'}) = \theta_1 \rho_{xz} C_x C_z; E(e_{x'} e_z) \\ = \theta_1 \rho_{xz} C_x C_z; E(e_x e_z) = \theta_2 \rho_{xz} C_x C_z; E(e_x e_{z'}) = \theta_1 \rho_{xz} C_x C_z; \text{ and } E(e_{z'} e_z) \\ = \theta_1 C_z^2 \end{aligned} \quad (3)$$

where,

$$\theta_1 = \frac{1}{n'} - \frac{1}{N} \text{ and } \theta_2 = \frac{1}{n} - \frac{1}{n'}$$

are the sampling fractions for the first and second sampling fractions respectively, (20).

EXISTING ESTIMATORS IN LITERATURE

(i) Classical regression estimator in two-phase sampling (8)

$$\bar{y}_{lr.ds} = \bar{y} + b_{y,x}(\bar{x}' - \bar{x}) \quad (4)$$

$$MSE(\bar{y}_{lr.ds}) = \bar{Y}^2 C_y^2 (\theta_2 - \theta_1 \rho_{y,x}^2) \quad (5)$$

(ii) Regression type estimator in two-phase sampling

The (25) regression type estimator for two-phase sampling is an extension of (26) estimator in survey sampling. It is given as:

$$\bar{y}_{H(2).ds} = \{\bar{y} + k_1(\bar{x}' - \bar{x})\} \left\{ k_2 \frac{\bar{z}'}{\bar{Z}} + (1 - k_2) \frac{\bar{z}}{\bar{Z}'} \right\} \quad (6)$$

$$-\infty \leq k_1 \leq \infty; 0 \leq k_2 \leq 1$$

$$Bias(\bar{y}_{H(2).ds}) = (\theta_2 - \theta_1) C_z \{ k_2 \bar{Y} C_z + (\bar{Y} C_y \rho_{yz} - k_1 \bar{X} C_x \rho_{xz}) \} (1 - 2k_2) \quad (7)$$

$$MSE(\bar{y}_{H(2).ds}) = \bar{Y}^2 C_y^2 \{ \theta_2 - (\theta_2 - \theta_1) [1 - (1 - \rho_{yz}^2)(1 - \rho_{yx.z}^2)] \} \quad (8)$$

where:

$$k_1 = \frac{\bar{Y} C_y (\rho_{yx} - \rho_{xz} \rho_{yz})}{\bar{X} C_x (1 - \rho_{xz}^2)}; \quad k_2 = \frac{1}{2} \left\{ 1 + \frac{C_y}{C_z} \left[\frac{\rho_{yz} - \rho_{yx} \rho_{xz}}{(1 - \rho_{xz}^2)} \right] \right\}$$

(iii) Ratio-chain type exponential estimator

Motivated by (27), (28) suggested the modified exponential ratio estimator to estimate finite population mean of study variable y in presence of two auxiliary variables, x and z . This estimator is given by

$$\bar{y}_{YUS.ds} = \bar{y} \exp \left[\frac{\bar{x}' \frac{(a\bar{z} - b)}{(a\bar{z} - b)} - \bar{x}}{\bar{x}' \frac{(a\bar{z} - b)}{(a\bar{z} - b)} + \bar{x}} \right] \quad (9)$$

$$\begin{aligned} Bias(\bar{y}_{YUS.ds}) = \bar{Y} \left[\frac{1}{2} \left(\frac{1-f}{n} \right) C_x^2 \left(\frac{3}{4} - K_{yx} \right) - \frac{1}{2} \left(\frac{1-f'}{n'} \right) \right. \\ \left. * \left(\frac{3}{4} C_x^2 - \frac{5}{4} \theta_0^2 C_z^2 - \rho_{yx} C_y C_x \right. \right. \\ \left. \left. + \theta_0 \rho_{yz} C_y C_z - \theta_0 \rho_{xz} C_x C_z \right) \right] \end{aligned} \quad (10)$$

$$MSE(\bar{y}_{YUS.ds}) = \bar{Y}^2 \left[\left(\frac{1-f}{n} \right) C_y^2 + \frac{1}{4} \left(\frac{1-f''}{n} \right) C_x^2 (1 - 4K_{yx}) - \left(\frac{1-f'}{n'} \right) \rho_{yz}^2 C_y^2 \right] \quad (11)$$

where:

$$\theta_0 = \frac{2\rho_{yx}C_yC_z}{C_z^2} = 2K_{yz}, \quad f = \frac{n}{N}; \quad f' = \frac{n'}{N}, \quad \text{and} \quad f'' = \frac{n}{n'}.$$

(iv) Exponential Chain ratio-type estimator in double sampling

(29) proposed a class of improved exponential chain ratio and product-type estimators of finite population mean \bar{Y} under Double (Two-phase) sampling scheme. The exponential chain ratio type estimator is given by:

$$\bar{y}_{KSC.ds} = \bar{y} \exp \left[\frac{\left(\frac{\bar{x}'}{\bar{x}} \right)^\alpha \frac{(a\bar{z} + b)}{(a\bar{z}' + b)} - 1}{\left(\frac{\bar{x}'}{\bar{x}} \right)^\alpha \frac{(a\bar{z} + b)}{(a\bar{z}' + b)} + 1} \right] \quad (12)$$

where, a and b are suitably chosen constants. The bias of the exponential chain ratio estimator is given as:

$$\begin{aligned} Bias(\bar{y}_{KSC.ds}) = \bar{Y} \left[\left(\frac{\alpha^2}{8} + \frac{\alpha}{4} \right) \lambda_3 C_x^2 + \frac{3}{8} \theta_0^2 \lambda_2 C_z^2 \right. \\ \left. - \frac{\alpha}{2} \lambda_3 \rho_{yx} C_y C_x - \frac{\theta_0}{2} \lambda_2 \rho_{yz} C_y C_z \right] \end{aligned} \quad (13)$$

$$MSE(\bar{y}_{KSC}) = \bar{Y}^2 \left[\lambda_1 C_y^2 + \frac{\alpha^2 C_x^2}{4} \lambda_3 + \frac{\theta_0^2 \lambda_2 C_z^2}{4} - \lambda_3 \rho_{yx} C_y C_x - \theta_0 \lambda_2 \rho_{yz} C_y C_z \right] \quad (14)$$

where,

$$\alpha = 2K_{y,x} = 2\rho_{yx} \frac{C_y}{C_x}; \quad \theta_0 = 2K_{y,z} = 2\rho_{yz} \frac{C_y}{C_z}$$

4. Proposed Estimator under Two-phase Sampling Scheme

By extending the [29] estimator of finite population mean in SRS, we propose the following hybrid exponential type regression-cum-ratio estimators of finite population mean in Two-phase random sampling under two conditions.

Case I: Proposed estimator under Two-phase sampling when auxiliary variables are not transformed

When the auxiliary variables are not transformed, our proposed alternative hybrid regression-cum-ratio exponential type estimator under of population mean under Two-phase sampling without replacement is given by:

$$\bar{y}_{uv.ds} = [\bar{y} + \beta_2(\bar{z}' - \bar{z})] \exp \left[\alpha_2 \left(\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}} \right) \right] \quad (15)$$

where, β_2 and α_2 are suitably constants which minimize the MSE of $\bar{y}_{uv.ds}$.

We define:

We define the relative error in the first phase as follows:

$$e_{x'} = \frac{\bar{x}' - \bar{x}}{\bar{x}} \text{ and } e_{z'} = \frac{\bar{z}' - \bar{z}}{\bar{z}}; \quad \bar{x}' = (1 + e_{x'})\bar{x} \text{ and } \bar{z}' = (1 + e_{z'})\bar{z}; \quad (16)$$

Theorem 1:

The bias of the proposed alternative hybrid regression-cum-ratio exponential type estimator under Two-phase sampling without replacement when the variables are not transformed is given by:

$$\begin{aligned} Bias(\bar{y}_{uv.ds}) = & \bar{Y} \left[\frac{1}{8} \alpha_2^2 (\theta_2 - \theta_1) C_x^2 - \frac{1}{2} \alpha_2 (\theta_2 - \theta_1) \rho_{yx} C_x C_y \right] \\ & + \frac{1}{2} \alpha_2 \beta_2 \bar{Z} [(\theta_2 - \theta_1) \rho_{xz} C_x C_z] \end{aligned} \quad (17)$$

Proof of Theorem 1:

Consider the expression in equation (15) above. Inserting the appropriate definitions of section 2 into the following expression, we have:

$$\begin{aligned} \bar{y}_{uv.ds} = & \{(1 + e_y)\bar{Y} + \beta_2[(1 + e_{z'})\bar{Z} - (\bar{Z} + \bar{Z}e_z)]\} \exp \left\{ \alpha_2 \left[\frac{(1 + e_{x'})\bar{X} - (1 + e_x)\bar{X}}{(1 + e_{x'})\bar{X} + (1 + e_x)\bar{X}} \right] \right\} \\ = & \{(1 + e_y)\bar{Y} + \beta_2[(\bar{Z} + \bar{Z}e_{z'}) - (\bar{Z} + \bar{Z}e_z)]\} * \exp \left\{ \alpha_2 \left[\frac{(1 + e_{x'} - 1 - e_x)\bar{X}}{(1 + e_{x'} + 1 + e_x)\bar{X}} \right] \right\} \end{aligned}$$

$$\begin{aligned}
&= \{(\bar{Y} + \bar{Y}e_y) + \beta_2[\bar{Z} + \bar{Z}e_{z'} - \bar{Z} - \bar{Z}e_z]\} * \exp\left\{\frac{1}{2}\alpha_2(e_{x'} - e_x)\left(1 + \frac{e_{x'} + e_x}{2}\right)^{-1}\right\} \\
&= [(\bar{Y} + \bar{Y}e_y) + \beta_2\bar{Z}(e_{z'} - e_z)] \exp\left[\frac{1}{2}\alpha_2(e_{x'} - e_x)\right] \quad (17a)
\end{aligned}$$

Expanding the exponential part of (17a) gives:

$$\begin{aligned}
&= [(\bar{Y} + \bar{Y}e_y) + \beta_2\bar{Z}(e_{z'} - e_z)] \left[1 + \frac{1}{2}\alpha_2(e_{x'} - e_x) + \frac{1}{2!} \cdot \frac{1}{4}\alpha_2^2(e_{x'} - e_x)^2 + \dots\right] \\
\bar{y}_{uv.ds} - \bar{Y} &= \bar{Y} \left[\frac{1}{2}\alpha_2(e_{x'} - e_x) + \frac{1}{8}\alpha_2^2(e_{x'}^2 - 2e_{x'}e_x + e_x^2) + e_y + \frac{1}{2}\alpha_2(e_{x'}e_y - e_xe_y)\right] \\
&\quad + \beta_2\bar{Z} \left\{e_{z'} - e_z + \frac{1}{2}\alpha_2(e_{x'}e_{z'} - e_xe_{z'})\right. \\
&\quad \left. - \frac{1}{2}\alpha_2(e_{x'}e_z - e_xe_z)\right\} \quad (17b)
\end{aligned}$$

Taking expectation on both sides of (17b) and applying the definitions in (3) gives

$$\begin{aligned}
Bias(\bar{y}_{uv.ds}) &= E(\bar{y}_{tpu} - \bar{Y}) \\
&= \bar{Y} \left[\frac{1}{2}\alpha_2 E(e_{x'} - e_x) + \frac{1}{8}\alpha_2^2 E(e_{x'}^2 - 2e_{x'}e_x + e_x^2) + E(e_y) + \frac{1}{2}\alpha_2 E(e_{x'}e_y - e_xe_y)\right] \\
&\quad + \beta_2\bar{Z} \left[E(e_{z'}) - E(e_z) + \frac{1}{2}\alpha_2 E(e_{x'}e_{z'} - e_xe_{z'})\right. \\
&\quad \left. - \frac{1}{2}\alpha_2 E(e_{x'}e_z - e_xe_z)\right] \\
&= \bar{Y} \left[\frac{1}{8}\alpha_2^2(\theta_2 - \theta_1)C_x^2 - \frac{1}{2}\alpha_2(\theta_2 - \theta_1)\rho_{yx}C_xC_y\right] + \beta_2\bar{Z} \left[\frac{1}{2}\alpha_2(\theta_2 - \theta_1)\rho_{xz}C_xC_z\right] \\
Bias(\bar{y}_{uv.ds}) &= \bar{Y} \left[\frac{1}{8}\alpha_2^2(\theta_2 - \theta_1)C_x^2 - \frac{1}{2}\alpha_2(\theta_2 - \theta_1)\rho_{yx}C_xC_y\right] \\
&\quad + \frac{1}{2}\alpha_2\beta_2\bar{Z}[(\theta_2 - \theta_1)\rho_{xz}C_xC_z]
\end{aligned}$$

Theorem 2:

The Mean Square Error (MSE) of the alternative hybrid regression-ratio exponential type estimator finite population mean under Two-phase sampling without replacement is given by:

$$\begin{aligned}
MSE(\bar{y}_{uv.ds}) &= \bar{Y}^2 C_y^2 (\rho_{xz}^2 - 1)^{-1} [(\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}^2 \rho_{xz}^2 \rho_{yz}^2)(\theta_2 - \theta_1) + (\rho_{xz}^2 - 1)\theta_2]; \rho_{y,x}, \\
&\quad \rho_{x,z}, \rho_{y,z} \neq \pm 1 \quad (18)
\end{aligned}$$

Proof of Theorem 2:

To find the MSE we square both sides of (17a) and take expectations.

$$(\bar{y}_{uv.ds} - \bar{Y})^2 = \left\{ \bar{Y} \left[\frac{1}{2} \alpha_2 (e_{x'} - e_x) + \frac{1}{8} \alpha_2^2 (e_{x'}^2 - 2e_{x'}e_x + e_x^2) + e_y + \frac{1}{2} \alpha_2 (e_{x'}e_y - e_xe_y) \right] + \beta_2 \bar{Z} \left[e_{z'} - e_z + \frac{1}{2} \alpha_2 (e_{x'}e_{z'} - e_xe_{z'}) - \frac{1}{2} \alpha_2 (e_{x'}e_z - e_xe_z) \right] \right\}^2$$

$$= \left\{ \bar{Y} \left[\frac{1}{2} \alpha_2 (e_{x'} - e_x) + e_y \right] + \beta_2 \bar{Z} [e_{z'} - e_z] \right\}^2$$

$$= \bar{Y}^2 \left[\frac{1}{2} \alpha_2 (e_{x'} - e_x) + e_y \right]^2 + 2\beta_2 \bar{Y} \bar{Z} \left[\frac{1}{2} \alpha_2 (e_{x'} - e_x) + e_y \right] [e_{z'} - e_z]$$

$$+ \beta_2^2 \bar{Z}^2 [e_{z'} - e_z]^2$$

$$MSE(\bar{y}_{uv.ds}) = E(\bar{y}_{tpu} - \bar{Y})^2$$

$$= \bar{Y}^2 \left[\frac{1}{4} \alpha_2^2 E(e_{x'}^2 - 2e_{x'}e_x + e_x^2) + \alpha_2 E(e_{x'}e_y - e_xe_y) + E(e_y^2) \right] + 2\beta_2 \bar{Y} \bar{Z} \left[\frac{1}{2} \alpha_2 E(e_{x'}e_{z'} - e_{x'}e_z - e_xe_{z'} + e_xe_z) + E(e_ye_{z'} - e_ye_z) \right] + \beta_2^2 \bar{Z}^2 E(e_{z'}^2 - 2e_{z'}e_z + e_z^2)$$

$$= \bar{Y}^2 \left[\frac{1}{4} \alpha_2^2 C_x^2 (\theta_2 - \theta_1) - \alpha_2 \rho_{yx} C_x C_y (\theta_2 - \theta_1) + C_y^2 \theta_2 \right] + 2\beta_2 \bar{Y} \bar{Z} \left[\frac{1}{2} \alpha_2 \rho_{xz} C_x C_z (\theta_2 - \theta_1) - \rho_{yz} C_y C_z (\theta_2 - \theta_1) \right] + \beta_2^2 \bar{Z}^2 C_z^2 (\theta_2 - \theta_1)$$

Therefore,

$$MSE(\bar{y}_{uv.ds}) = \bar{Y}^2 \left[\frac{1}{4} \alpha_2^2 C_x^2 (\theta_2 - \theta_1) - \alpha_2 \rho_{yx} C_x C_y (\theta_2 - \theta_1) + C_y^2 \theta_2 \right] + 2\beta_2 \bar{Y} \bar{Z} \left[\frac{1}{2} \alpha_2 \rho_{xz} C_x C_z (\theta_2 - \theta_1) - \rho_{yz} C_y C_z (\theta_2 - \theta_1) \right] + \beta_2^2 \bar{Z}^2 C_z^2 (\theta_2 - \theta_1) \quad (19)$$

Theorem 3

The values of α_2 and β_2 that minimizes the MSE of the proposed estimator, \bar{y}_{tpu} , in two-phase sampling when the supplementary variables are not transformed are given by:

$$\alpha_2 = \frac{2(\bar{Y} \rho_{yx} C_y - \beta_2 \bar{Z} \rho_{xz} C_z)}{\bar{Y} C_x} \text{ and } \beta_2 = \frac{\bar{Y} C_y (\rho_{yx} \rho_{xz} - \rho_{yz})}{\bar{Z} C_z (\rho_{xz}^2 - 1)} \quad (20)$$

Proof of Theorem 3:

In order to find the optimum values of α_2 and β_2 that minimize the MSE, we differentiate Equation (19) partially with respect to α_2 and β_2 and equate to zero.

$$\frac{\partial MSE(\bar{y}_{uv.ds})}{\partial \alpha_2} = \bar{Y}^2 \left[\frac{1}{2} \alpha_2 C_x^2 (\theta_2 - \theta_1) - \rho_{yx} C_x C_y (\theta_2 - \theta_1) \right] + 2\beta_2 \bar{Y} \bar{Z} \left[\frac{1}{2} \rho_{xz} C_x C_z (\theta_2 - \theta_1) \right] = 0$$

$$\bar{Y}^2 \left[\frac{1}{2} \alpha_2 C_x^2 (\theta_2 - \theta_1) - \rho_{yx} C_x C_y (\theta_2 - \theta_1) \right] + 2\beta_2 \bar{Y} \bar{Z} \left[\frac{1}{2} \rho_{xz} C_x C_z (\theta_2 - \theta_1) \right] = 0$$

$$\bar{Y}^2 \left[\frac{1}{2} \alpha_2 C_x - \rho_{yx} C_y \right] + 2\beta_2 \bar{Y} \bar{Z} \left[\frac{1}{2} \rho_{xz} C_z \right] = 0$$

$$\alpha_2 = \frac{2(\bar{Y} \rho_{yx} C_y - \beta_2 \bar{Z} \rho_{xz} C_z)}{\bar{Y} C_x} \quad (21)$$

$$\frac{\partial MSE(\bar{y}_{uv.ds})}{\partial \beta_2} = 2\bar{Y} \bar{Z} \left[\frac{1}{2} \alpha_2 \rho_{xz} C_x C_z (\theta_2 - \theta_1) - \rho_{yz} C_y C_z (\theta_2 - \theta_1) \right] + 2\beta_2 \bar{Z}^2 C_z^2 (\theta_2 - \theta_1) = 0$$

$$2\beta_2 \bar{Z} C_z = -\bar{Y} (\alpha_2 \rho_{xz} C_x - 2\rho_{yz} C_y)$$

$$\beta_2 = -\frac{1}{2} \frac{\bar{Y} (\alpha_2 \rho_{xz} C_x - 2\rho_{yz} C_y)}{\bar{Z} C_z} \quad (22a)$$

$$\beta_2 = \frac{\bar{Y} \rho_{yx} \rho_{xz} C_y - \bar{Y} \rho_{yz} C_y}{\bar{Z} \rho_{xz}^2 C_z - \bar{Z} C_z}$$

Therefore,

$$\beta_2 = \frac{\bar{Y} C_y (\rho_{yx} \rho_{xz} - \rho_{yz})}{\bar{Z} C_z (\rho_{xz}^2 - 1)} \quad (22b)$$

Putting (21) and (22b) into (19), and simplifying using maple, we have

$$MSE(\bar{y}_{uv.ds}) = \bar{Y}^2 \left[\frac{1}{4} \alpha_2^2 C_x^2 (\theta_2 - \theta_1) - \alpha_2 \rho_{yx} C_x C_y (\theta_2 - \theta_1) + C_y^2 \theta_2 \right] + 2\beta_2 \bar{Y} \bar{Z} \left[\frac{1}{2} \alpha_2 \rho_{xz} C_x C_z (\theta_2 - \theta_1) - \rho_{yz} C_y C_z (\theta_2 - \theta_1) \right] + \beta_2^2 \bar{Z}^2 C_z^2 (\theta_2 - \theta_1)$$

$$= \frac{\bar{Y}^2 C_y^2}{\rho_{xz}^2 - 1} \left[-2\rho_{yx}^2 \rho_{xz}^2 \rho_{yz}^2 (\theta_2 - \theta_1) + \rho_{yx}^2 (\theta_2 - \theta_1) + \rho_{yz}^2 (\theta_2 - \theta_1) + (\rho_{xz}^2 - 1) \theta_2 \right]$$

Therefore,

$$MSE(\bar{y}_{uv.ds}) = \bar{Y}^2 C_y^2 (\rho_{xz}^2 - 1)^{-1} \left[(\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}^2 \rho_{xz}^2 \rho_{yz}^2) (\theta_2 - \theta_1) + (\rho_{xz}^2 - 1) \theta_2 \right]; \rho_{xz}, \rho_{yz}, \rho_{yx} \neq \pm 1$$

Recall again from (17) that:

$$Bias(\bar{y}_{uv.ds}) = \bar{Y} \left[\frac{1}{8} \alpha_2^2 (\theta_2 - \theta_1) C_x^2 - \frac{1}{2} \alpha_2 (\theta_2 - \theta_1) \rho_{yx} C_x C_y \right] + \frac{1}{2} \alpha_2 \beta_2 \bar{Z} [(\theta_2 - \theta_1) \rho_{xz} C_x C_z]$$

Therefore, inserting the values of (21) and (22b) into the above and simplifying using maple gives us:

$$Bias(\bar{y}_{uv.ds}) = -\frac{1}{2} \frac{\bar{Y} C_y^2 (\theta_2 - \theta_1)}{(\rho_{xz}^2 - 1)^2} (\rho_{xz} \rho_{yz} - \rho_{yx})^2; \rho_{xz} \neq 1 \quad (23)$$

Case II: Proposed estimator in Two-Phase random sampling when auxiliary variables are transformed.

The proposed estimator when the variables are transformed is given as:

$$\bar{y}_{tv.ds} = [\bar{y} + \beta_3 (\bar{z}^* - \bar{z}')] \exp \left[\alpha_3 \left(\frac{\bar{x}^* - \bar{x}'}{\bar{x}' + \bar{x}^*} \right) \right] \quad (24)$$

where, α_3 and β_3 are determined such that the MSE of \bar{y}_{tpt} is minimized.

Theorem 4

The bias of the hybrid regression-cum-ratio exponential type estimator using SRSWOR, under two-phase sampling when the auxiliary variables are transformed is given by:

$$Bias(\bar{y}_{tv.ds}) = \bar{Y} \left[\frac{1}{8} \alpha_3^2 C_x^2 \Phi - \frac{1}{2} \alpha_3 \rho_{yx} C_x C_y \Psi \right] + \beta_3 \bar{Z} \left[\frac{1}{2} \alpha_3 \rho_{xz} C_x C_z \Phi \right] \quad (25)$$

Proof of Theorem 4:

Consider the expression in (24)

$$\bar{y}_{tv.ds} = [\bar{y} + \beta_3 (\bar{z}^* - \bar{z}')] \exp \left[\alpha_3 \left(\frac{\bar{x}^* - \bar{x}'}{\bar{x}' + \bar{x}^*} \right) \right]$$

And inserting the definitions of (16) and (24) into the above expression, we have:

$$\begin{aligned} \bar{y}_{tv.ds} &= \{(1 + e_y) \bar{Y} + \beta_3 [(1 - g e_z) \bar{Z} - (1 + e_{z'}) \bar{Z}]\} * \exp \left\{ \alpha_3 \left[\frac{(1 - g e_x) \bar{X} - (1 + e_{x'}) \bar{X}}{(1 + e_{x'}) \bar{X} + (1 - g e_x) \bar{X}} \right] \right\} \\ &= \{(\bar{Y} + \bar{Y} e_y) - \beta_3 \bar{Z} (e_{z'} + g e_z)\} * \exp \left\{ -\frac{1}{2} \alpha_3 (e_{x'} + g e_x) \left[1 + \frac{1}{2} (e_{x'} - g e_x) \right] \right\} \\ &= \{(\bar{Y} + \bar{Y} e_y) - \beta_3 \bar{Z} (e_{z'} + g e_z)\} * \exp \left\{ -\frac{1}{2} \alpha_3 \left[e_{x'} + \frac{1}{2} (e_{x'}^2 - g e_x e_{x'}) \right] + \left[g e_x + \frac{1}{2} (g e_{x'} e_x - g^2 e_x^2) \right] \right\} \end{aligned}$$

By first order approximation we have:

$$= [(\bar{Y} + \bar{Y} e_y) - \beta_3 \bar{Z} (e_{z'} + g e_z)] \exp \left[-\frac{1}{2} \alpha_3 (e_{x'} + g e_x) \right] \quad (26)$$

Which on expansion gives:

$$= [(\bar{Y} + \bar{Y} e_y) - \beta_3 \bar{Z} (e_{z'} + g e_z)] * \left\{ 1 - \left[\frac{1}{2} \alpha_3 (e_{x'} + g e_x) \right] + \frac{1}{2!} \cdot \frac{1}{4} \cdot \alpha_3^2 (e_{x'} + g e_x)^2 + \dots \right\}$$

Considering terms up to the power of two only we get:

$$\begin{aligned}
 &= [(\bar{Y} + \bar{Y}e_y) - \beta_3\bar{Z}(e_{z'} + ge_z)] * \left[1 - \frac{1}{2}\alpha_3(e_{x'} + ge_x) + \frac{1}{8}\alpha_3^2(e_{x'}^2 + 2ge_{x'}e_x + g^2e_x^2)\right] \\
 &\quad \bar{y}_{tpt} - \bar{Y} = \bar{Y} \left[-\frac{1}{2}\alpha_3(e_{x'} + ge_x) + \frac{1}{8}\alpha_3^2(e_{x'}^2 + 2ge_{x'}e_x + g^2e_x^2) \right] + \bar{Y} \left[e_y - \frac{1}{2}\alpha_3(e_{x'}e_y + ge_xe_y) \right] \\
 &\quad - \beta_3\bar{Z} \left\{ \left[e_{z'} - \frac{1}{2}\alpha_3(e_{x'}e_{z'} + ge_xe_{z'}) \right] \right. \\
 &\quad \left. + \left[ge_z - \frac{1}{2}\alpha_3(ge_{x'}e_z + g^2e_xe_z) \right] \right\} \quad (27)
 \end{aligned}$$

Taking expectation on both sides of (27)) we have:

$$\begin{aligned}
 Bias(\bar{y}_{tv.ds}) &= E(\bar{y}_{tpt} - \bar{Y}) \\
 &= \bar{Y} \left[-\frac{1}{2}\alpha_3E(e_{x'} + ge_x) + \frac{1}{8}\alpha_3^2E(e_{x'}^2 + 2ge_{x'}e_x + g^2e_x^2) + E(e_y) - \frac{1}{2}\alpha_3E(e_{x'}e_y + ge_xe_y) \right] \\
 &\quad + \beta_3\bar{Z} \left\{ \left[E(e_{z'}) - \frac{1}{2}\alpha_3E(e_{x'}e_{z'} + ge_xe_{z'}) \right] + \left[gE(e_z) - \frac{1}{2}\alpha_3E(ge_{x'}e_z + g^2e_xe_z) \right] \right\}
 \end{aligned}$$

Inserting the definitions of (3) in the above expression, we have:

$$\begin{aligned}
 &= \bar{Y} \left[-\frac{1}{2}\alpha_3(0 + 0) + \frac{1}{8}\alpha_3^2(\theta_1C_x^2 + 2g\theta_1C_x^2 + g^2\theta_2C_x^2) + 0 \right] - \frac{1}{2}\alpha_3(\theta_1\rho_{yx}C_xC_y + g\theta_2\rho_{yx}C_xC_y) \\
 &\quad - \beta_3\bar{Z} \left\{ \left[0 - \frac{1}{2}\alpha_3(\theta_1\rho_{xz}C_xC_z + g\theta_1\rho_{xz}C_xC_z) \right] + \left[0 - \frac{1}{2}\alpha_3(g\theta_1\rho_{xz}C_xC_z + g^2\theta_2\rho_{xz}C_xC_z) \right] \right\} \\
 &= \bar{Y} \left[\frac{1}{8}\alpha_3^2C_x^2(\theta_1 + 2g\theta_1 + g^2\theta_2) - \frac{1}{2}\alpha_3\rho_{yx}C_xC_y(\theta_1 + g\theta_2) \right] + \beta_3\bar{Z} \left[\frac{1}{2}\alpha_3\rho_{xz}C_xC_z(\theta_1 + 2g\theta_1 + g^2\theta_2) \right]
 \end{aligned}$$

Therefore,

$$\begin{aligned}
 Bias(\bar{y}_{tv.ds}) &= \bar{Y} \left[\frac{1}{8}\alpha_3^2C_x^2(\theta_1 + 2g\theta_1 + g^2\theta_2) - \frac{1}{2}\alpha_3\rho_{yx}C_xC_y(\theta_1 + g\theta_2) \right] \\
 &\quad + \beta_3\bar{Z} \left[\frac{1}{2}\alpha_3\rho_{xz}C_xC_z(\theta_1 + 2g\theta_1 + g^2\theta_2) \right] \\
 &= \bar{Y} \left[\frac{1}{8}\alpha_3^2C_x^2\Phi - \frac{1}{2}\alpha_3\rho_{yx}C_xC_y\Psi \right] + \beta_3\bar{Z} \left[\frac{1}{2}\alpha_3\rho_{xz}C_xC_z\Phi \right] \quad (28)
 \end{aligned}$$

Theorem 5

The MSE of the hybrid regression–ratio exponential type estimator in two-phase sampling is given by:

$$MSE(\bar{y}_{tv.ds}) = \frac{\bar{Y}^2C_y^2}{(\rho_{xz}^2 - 1)\Phi} [(\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{xz}\rho_{yz})\Psi^2 + (\rho_{xz}^2 - 1)\Phi\theta_2]; \rho_{yx}, \rho_{xz}, \rho_{yz}, \neq \pm 1, \Phi \neq 0 \quad (29)$$

where, $\rho_{xzh}, \rho_{xzh}, \rho_{yzh} \pm 1$.

Proof of Theorem 5

In order to obtain the MSE of $\bar{y}_{tp.st}$, we square both sides of Equation (27) and take expectations.

$$MSE(\bar{y}_{tv.ds}) = E(\bar{y}_{tpt} - \bar{Y})^2$$

$$\begin{aligned} (\bar{y}_{tv.ds} - \bar{Y})^2 &= \left\{ \bar{Y} \left[-\frac{1}{2} \alpha_3 (e_{x'} + g e_x) + \frac{1}{8} \alpha_3^2 (e_{x'}^2 + 2g e_{x'} e_x + g^2 e_x^2) \right] + \bar{Y} \left[e_y - \frac{1}{2} \alpha_3 (e_{x'} e_y + g e_x e_y) \right] \right. \\ &\quad \left. - \beta_3 \bar{Z} \left[\left[e_{z'} - \frac{1}{2} \alpha_3 (e_{x'} e_{z'} + g e_x e_{z'}) \right] + \left[g e_z - \frac{1}{2} \alpha_3 (g e_{x'} e_z + g^2 e_x e_z) \right] \right] \right\}^2 \\ &= \left\{ \bar{Y} \left[-\frac{1}{2} \alpha_3 (e_{x'} + g e_x) + e_y \right] + \beta_3 \bar{Z} (e_{z'} + g e_z) \right\}^2 \\ &= \bar{Y}^2 \left[\frac{1}{4} \alpha_3^2 (e_{x'}^2 + 2g e_{x'} e_x + g^2 e_x^2) - \alpha_3 (e_{x'} e_y + g e_x e_y) + e_y^2 \right] \\ &\quad + 2\beta_3 \bar{Z} \bar{Y} \left[-\frac{1}{2} \alpha_3 (e_{x'} e_{z'} + g e_{x'} e_z + g e_x e_{z'} + g^2 e_x e_z) + e_y e_{z'} + g e_y e_z \right] \\ &\quad + \beta_3^2 \bar{Z}^2 [e_{z'}^2 + 2g e_{z'} e_z + g^2 e_z^2] \end{aligned}$$

By taking expectations on both sides of the above and applying the definition of (3), we have

$$\begin{aligned} MSE(\bar{y}_{tv.ds}) &= \bar{Y}^2 \left[\frac{1}{4} \alpha_3^2 C_x^2 (\theta_1 + 2g\theta_1 + g^2\theta_2) - \alpha_3 \rho_{yx} C_x C_y (\theta_1 + g\theta_2) + \theta_2 C_y^2 \right] \\ &\quad + 2\beta_3 \bar{Z} \bar{Y} \left[-\frac{1}{2} \alpha_3 \rho_{xz} C_x C_z (\theta_1 + 2g\theta_1 + g^2\theta_2) + \rho_{yz} C_y C_z (\theta_1 + g\theta_2) \right] \\ &\quad + \beta_3^2 \bar{Z}^2 C_z^2 (\theta_1 + 2g\theta_1 + g^2\theta_2) \end{aligned}$$

Therefore,

$$\begin{aligned} MSE(\bar{y}_{tpt}) &= \bar{Y}^2 \left[\frac{1}{4} \alpha_3^2 C_x^2 \Phi - \alpha_3 \rho_{yx} C_x C_y \psi + \theta_2 C_y^2 \right] + 2\beta_3 \bar{Z} \bar{Y} \left[-\frac{1}{2} \alpha_3 \rho_{xz} C_x C_z \Phi + \rho_{yz} C_y C_z \psi \right] \\ &\quad + \beta_3^2 \bar{Z}^2 C_z^2 \Phi \end{aligned} \quad (30)$$

where,

$$\Phi = \theta_1 + 2g\theta_1 + g^2\theta_2; \quad \psi = \theta_1 + g\theta_2$$

Theorem 6

The values of α_3 and β_3 that minimizes the MSE of the proposed estimator, \bar{y}_{tpt} , in two-phase sampling when the auxiliary variables are transformed are given by:

$$\alpha_3 = \frac{2(\beta_3 \bar{Z} \rho_{xz} C_z \Phi + \bar{Y} \rho_{yx} C_y \psi)}{\bar{Y} C_x \Phi} \quad \text{and} \quad \beta_3 = -\frac{\bar{Y} C_y \psi (\rho_{xz} \rho_{yx} - \rho_{yz})}{\bar{Z} C_z \Phi (\rho_{xz}^2 - 1)} \quad (31)$$

Proof of Theorem 6:

To obtain the optimal values of α_3 and β_3 , we differentiate (30) partially with respect to α_3 and β_3 , and equating to zero:

$$\frac{\partial MSE(\bar{y}_{tv.ds})}{\partial \alpha_3} = \bar{Y}^2 \left(\frac{1}{2} \alpha_3 C_x^2 \Phi - \rho_{yx} C_x C_y \psi \right) - \beta_3 \bar{Z} \bar{Y} \rho_{xz} C_x C_z \Phi = 0 \quad (32)$$

$$\frac{\partial MSE(\bar{y}_{tv.ds})}{\partial \beta_3} = 2 \bar{Y} \bar{Z} \left(\frac{1}{2} \alpha_3 \rho_{xz} C_x C_z \Phi - \rho_{yz} C_y C_z \psi \right) - 2 \beta_3 \bar{Z}^2 C_z^2 \Phi = 0$$

From Equation (31) ,

$$\alpha_3 = \frac{2(\beta_3 \bar{Z} \rho_{xz} C_z \Phi + \bar{Y} \rho_{yx} C_y \psi)}{\bar{Y} C_x \Phi} \quad (33)$$

$$\beta_3 = -\frac{1}{2} \frac{\bar{Y} (\alpha_3 \rho_{xz} C_x C_z \Phi + 2 \rho_{yz} C_y \psi)}{\bar{Z} C_z \Phi} \quad (34a)$$

Putting (33) into (34a)

$$\beta_3 = \frac{\beta_3 \bar{Z} \rho_{xz}^2 C_z \Phi + \bar{Y} \rho_{xz} \rho_{yx} C_y \psi - \bar{Y} \rho_{yz} C_y \psi}{\bar{Z} C_z \Phi}$$

$$\beta_3 (\bar{Z} C_z \Phi - \bar{Z} \rho_{xz}^2 C_z \Phi) = \bar{Y} \rho_{xz} \rho_{yx} C_y \psi - \bar{Y} \rho_{yz} C_y \psi$$

$$\beta_3 = -\frac{\bar{Y} C_y \psi (\rho_{xz} \rho_{yx} - \rho_{yz})}{\bar{Z} C_z \Phi (\rho_{xz}^2 - 1)} \quad (34b)$$

Again, Putting (33) and (34b) into (30), we obtain the $MSE(\bar{y}_{tv.ds})$ have on simplification using maple

$$\begin{aligned} MSE(\bar{y}_{tv.ds}) &= \bar{Y}^2 \left[\frac{1}{4} \alpha_3^2 C_x^2 \Phi - \alpha_3 \rho_{yx} C_x C_y \psi + \theta_2 C_y^2 \right] + 2 \beta_3 \bar{Z} \bar{Y} \left[-\frac{1}{2} \alpha_3 \rho_{xz} C_x C_z \Phi + \rho_{yz} C_y C_z \psi \right] \\ &+ \beta_3^2 \bar{Z}^2 C_z^2 \Phi \end{aligned}$$

$$= \frac{\bar{Y}^2 C_y^2 [(\rho_{yx}^2 + \rho_{yz}^2 - 2 \rho_{yx} \rho_{xz} \rho_{yz}) \Psi^2 + (\rho_{xz}^2 - 1) \Phi \theta_2]}{(\rho_{xz}^2 - 1) \Phi}$$

$$MSE(\bar{y}_{tv.ds}) = \frac{\bar{Y}^2 C_y^2}{(\rho_{xz}^2 - 1) \Phi} [(\rho_{yx}^2 + \rho_{xz}^2 - 2 \rho_{yx} \rho_{xz} \rho_{yz}) \Psi^2 + (\rho_{xz}^2 - 1) \Phi \theta_2]$$

where,

$$\rho_{yx}, \rho_{xz}, \rho_{yz} \neq \pm 1.$$

From (28), we know that:

$$Bias(\bar{y}_{tv.ds}) = \bar{Y} \left[\frac{1}{8} \alpha_3^2 C_x^2 \Phi - \frac{1}{2} \alpha_3 \rho_{yx} C_x C_y \psi \right] + \beta_3 \bar{Z} \left[\frac{1}{2} \alpha_3 \rho_{xz} C_x C_z \Phi \right]$$

Therefore, putting (33) and (34b) into the above expression gives on simplification

$$Bias(\bar{y}_{tpt}) = \frac{1}{2} \frac{\bar{Y} C_y^2 \Psi^2}{(\rho_{xz}^2 - 1) \Phi} (\rho_{xz} \rho_{yz} - \rho_{yx}) (4 \rho_{xz}^2 \rho_{yx} - 3 \rho_{xz} \rho_{yz} - \rho_{yx}) \quad (35)$$

where, $\rho_{xz} \neq 1, \Phi \neq 0$

EFFICIENCY CRITERIA

For two unbiased estimators T and T^* of θ , we compare the variances of the estimators to determine their relative efficiency, i.e. Relative Efficiency of T and T^* is:

$$RE(T, T^*; \theta) = \frac{Var(T^*)}{Var(T)} \quad (36)$$

We extend this to large sample variances of asymptotically unbiased estimators.

Let $\{T_i^*\}, i = 1, 2$ and $\{T_j\}, j = 1, 2, 3$ be asymptotically unbiased for $\tau(\theta)$, then it is said that T_i^* is Asymptotically Efficient Relative to $\{T_j\}$ if:

$$ARE(T_j, T_i^*; \theta) = \lim_{n \rightarrow \infty} \frac{Var(T_i^*)}{Var(T_j)} \leq 1, \quad \theta \in \Theta, i = 1, 2; j = 1, 2, 3, 4 \quad (37)$$

where $\{T_i^*\}, i = 1, 2$ are the proposed and $\{T_j\}, j = 1, 2, 3$ and existing and competing estimators (8), (21), (28), and (29) under consideration in this study.

Asymptotic unbiasedness

Consider estimators based on n -sample, $T_n = T_n(X_1, X_2, \dots, X_n)$, where X_1, X_2, \dots, X_n are identical and independently distributed. Even estimators that are biased, may be close to unbiased for large n . An estimator T_n is an asymptotically unbiased estimator of deterministic θ if:

$$Bias(T_n) = E(T_n) - \theta \rightarrow 0 \text{ as } n \rightarrow \infty \quad (38)$$

Theorem 7

Let T_n denote an estimator of θ . If $Bias_{\theta}(T_n)$ and $Var_{\theta}(T_n)$ both approach zero as $n \rightarrow \infty$, then the mean squared error between T_n and θ approaches zero and hence T_n is a consistent estimator of θ .

Proof:

$$E[(T_n - \theta)^2] = Bias_{T_n}^2(\theta) + Var_{T_n}(\theta)$$

So, if:

$$Bias_{T_n}^2(\theta) \rightarrow 0$$

and

$$Var_{T_n}(\theta) \rightarrow 0$$

as $n \rightarrow \infty$, then,

$$\lim_{n \rightarrow \infty} E[(T_n - \theta)^2] = 0$$

which means convergence in mean square and hence T_n also converges to θ in probability.

We now show that \bar{y}_{uv} is asymptotically consistent.

Ratio of bias to standard error (RBSE)

(30) stated a mathematical expression to measure the effect of bias on the accuracy of an estimate as:

$$RBSE = \frac{|B|}{\sqrt{Var(T_n)}} \quad (39a)$$

where B is the bias of T_n (and T_n is unbiased).

as a working rule, if

$$RBSE = \frac{|B|}{\sqrt{MSE(T_n)}} < 0.1 \quad (39b)$$

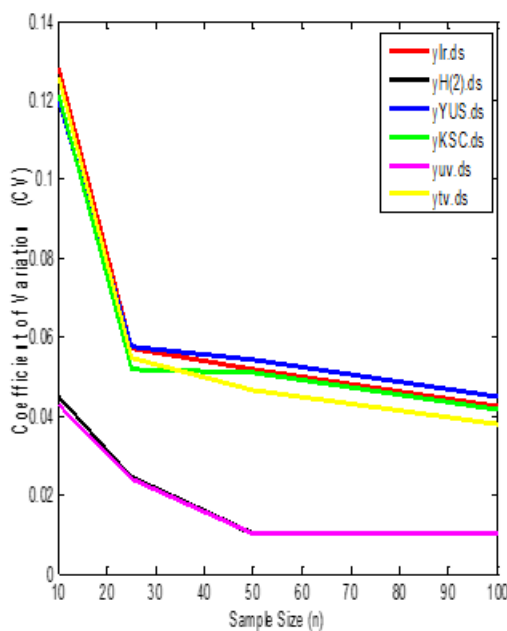
Then, effect of bias on the accuracy of an estimate is negligible otherwise, there might exist the problem of over-/under-estimation of the parameters in question

RESULTS

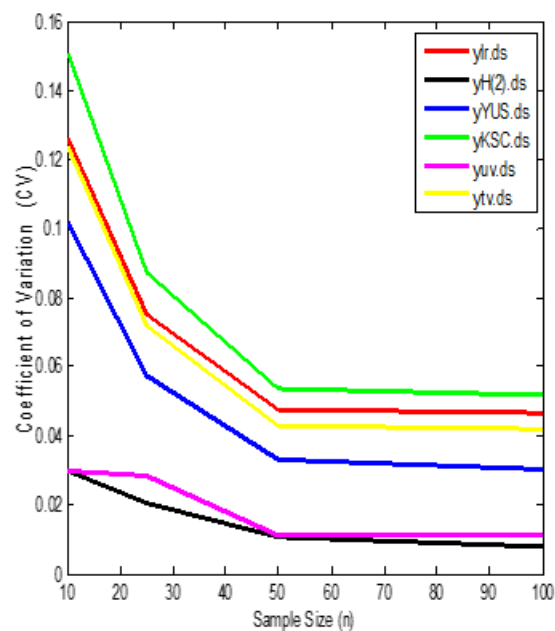
The results of analyses of both the synthetic and real data sets are as presented in Tables 1-6 and Figures 1-4 respectively.

Table 1: Estimates of Mean by Estimators (Populations I-II)

Corr. Coeff.	Estimator	Population I				Population II			
		Sample size (n)				Sample size (n)			
		10	25	50	100	10	25	50	100
ρ_{HLL}	$\bar{y}_{lr.ds}$	2997.2749	2999.8259	3004.6568	2999.5752	117.097	131.0738	118.5579	126.2957
	$\bar{y}_{H(2).ds}$	3003.8305	2999.8169	3008.7078	3001.5709	96.2106	179.6571	111.8701	142.1266
	$\bar{y}_{YUS.ds}$	3001.1202	3000.6463	3007.5985	3000.3857	104.2954	138.9594	109.5112	130.6289
	$\bar{y}_{KSC.ds}$	2996.5725	2999.9403	3003.165	2998.7430	131.3514	117.7131	123.7745	118.8470
	$\bar{y}_{uv.ds}$	3004.3829	3000.0405	3007.3698	3000.5766	81.4496	166.7773	114.2188	139.8101
	$\bar{y}_{tv.ds}$	2995.9190	2998.8601	2998.9709	2998.6614	109.9004	138.1936	128.3924	115.6996
ρ_{HHH}	$\bar{y}_{lr.ds}$	3001.5258	3003.1186	2999.5476	2999.3916	116.6134	129.2331	104.0943	121.2374
	$\bar{y}_{H(2).ds}$	2995.2180	3001.5362	2998.8251	2999.0559	105.9444	163.3707	105.6481	129.1032
	$\bar{y}_{YUS.ds}$	2998.3634	3002.8839	3000.2471	2999.0451	104.2954	138.9594	109.5112	130.6289
	$\bar{y}_{KSC.ds}$	3007.8232	3004.1914	3000.6680	2999.1729	130.6697	107.9089	111.5728	110.6655
	$\bar{y}_{uv.ds}$	3015.3354	3003.3525	3000.7607	2998.3208	95.39821	136.2343	110.3303	112.8225
	$\bar{y}_{tv.ds}$	2996.2477	3002.6030	2995.2379	2999.8598	108.5568	137.6100	116.7191	120.2998



(a)



(b)

Figure 1: Plot of Coefficient of Variation of Estimators for Population I
(a) ρ_{HLL} (b) ρ_{HHH}

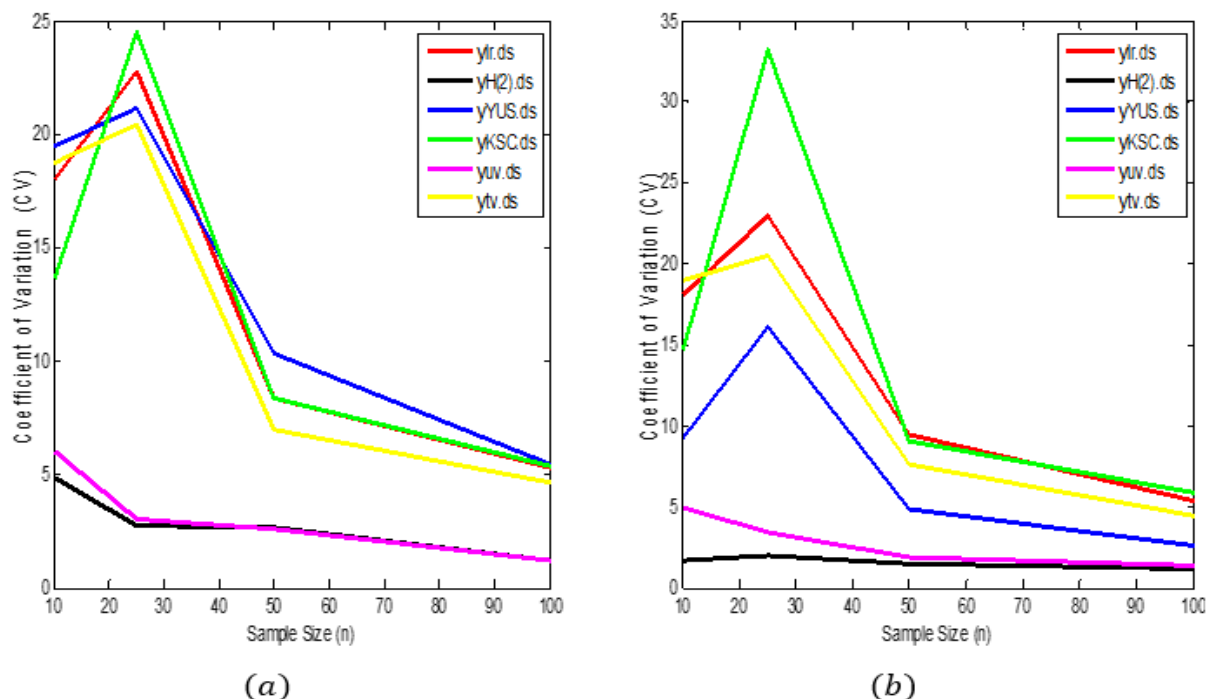


Figure 2: Plot of Coefficient of Variation of Estimators for Population II
(a) P_{HLL} (b) P_{HHH}

Table 3: Bias of Estimators under Two-Phase Sampling (Populations I-II)

		Population I				Population II			
		Sample size (n)				Sample size (n)			
Corr. Coeff.	Estimator	10	25	50	100	10	25	50	100
P_{HLL}	$\bar{y}_{lr.ds}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$\bar{y}_{H(2).ds}$	-0.0173	-0.005	-0.003	-0.003	-5.2033	-12.405	-1.6331	-0.6187
	$\bar{y}_{YUS.ds}$	$3E-05$	$1E-05$	$2E-05$	$1E-05$	-0.8852	-1.4866	-0.1517	-0.0660
	$\bar{y}_{KSC.ds}$	-0.0433	-0.02	-0.042	-0.027	-33.174	-177.03	-38.2519	-38.2101
	$\bar{y}_{uv.ds}$	$-5E-05$	$-9E-06$	$-2E-08$	$-1E-07$	-0.0015	-6.6E-06	-0.0008	-1.7E-05
	$\bar{y}_{tv.ds}$	$2E-12$	$6E-13$	$2E-14$	$3E-15$	$7.12E-10$	$1.43E-09$	$2.01E-10$	$5.78E-11$
P_{HHH}	$\bar{y}_{lr.ds}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$\bar{y}_{H(2).ds}$	-0.0226	-0.007	-0.002	-0.002	-3.4226	-9.9529	0.47814	-0.3437
	$\bar{y}_{YUS.ds}$	$4E-05$	$4E-05$	$3E-05$	$3E-05$	0.35795	6.5534	0.20616	0.11121
	$\bar{y}_{KSC.ds}$	-0.0469	-0.038	-0.035	-0.033	-33.345	-178.71	-38.195	-38.196
	$\bar{y}_{uv.ds}$	$-4E-05$	$-6E-05$	$-1E-05$	$-7E-05$	-0.2892	-0.2259	-0.5567	-0.0441
	$\bar{y}_{tv.ds}$	2996.247	3002.603	2995.237	2999.859	108.5568	137.6100	116.7191	120.2998

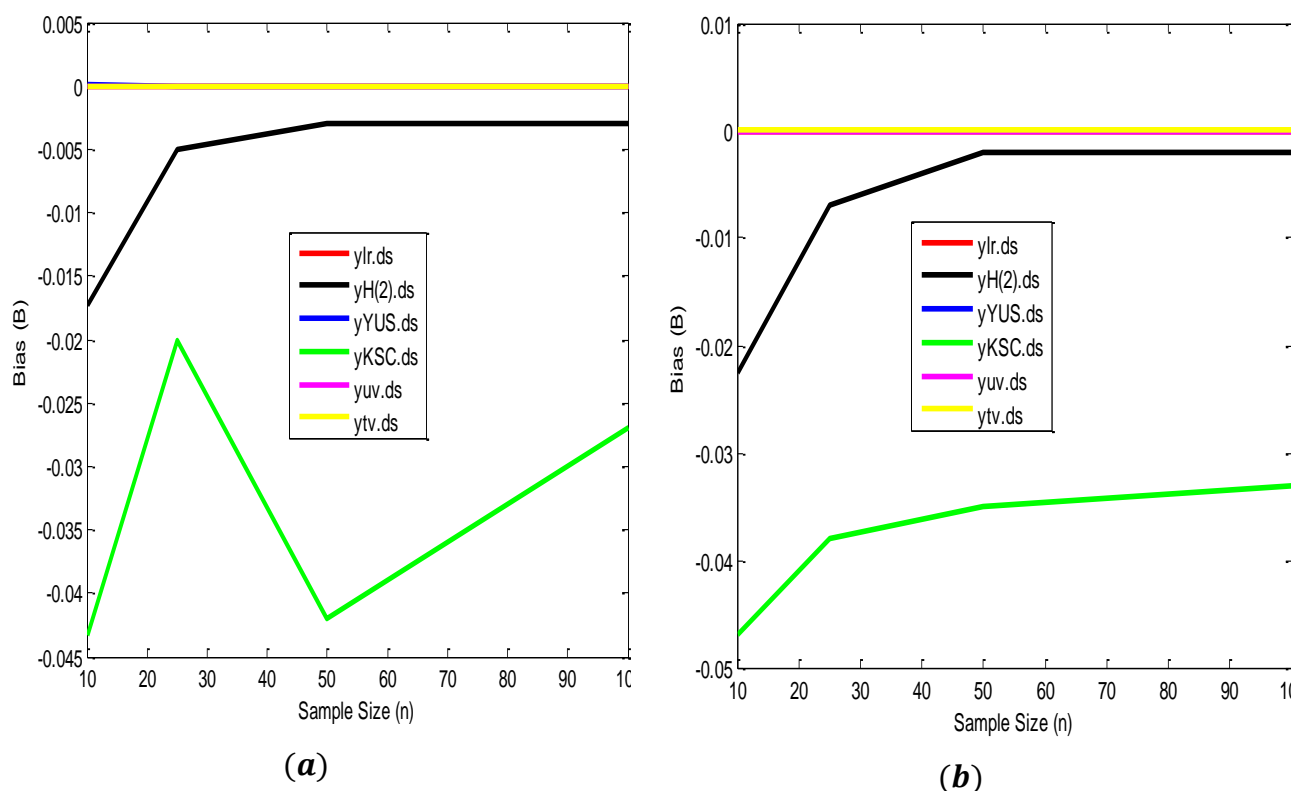


Figure 3: Plot of Bias of Estimators for Population I (a) ρ_{HLL} (b) ρ_{HHH}

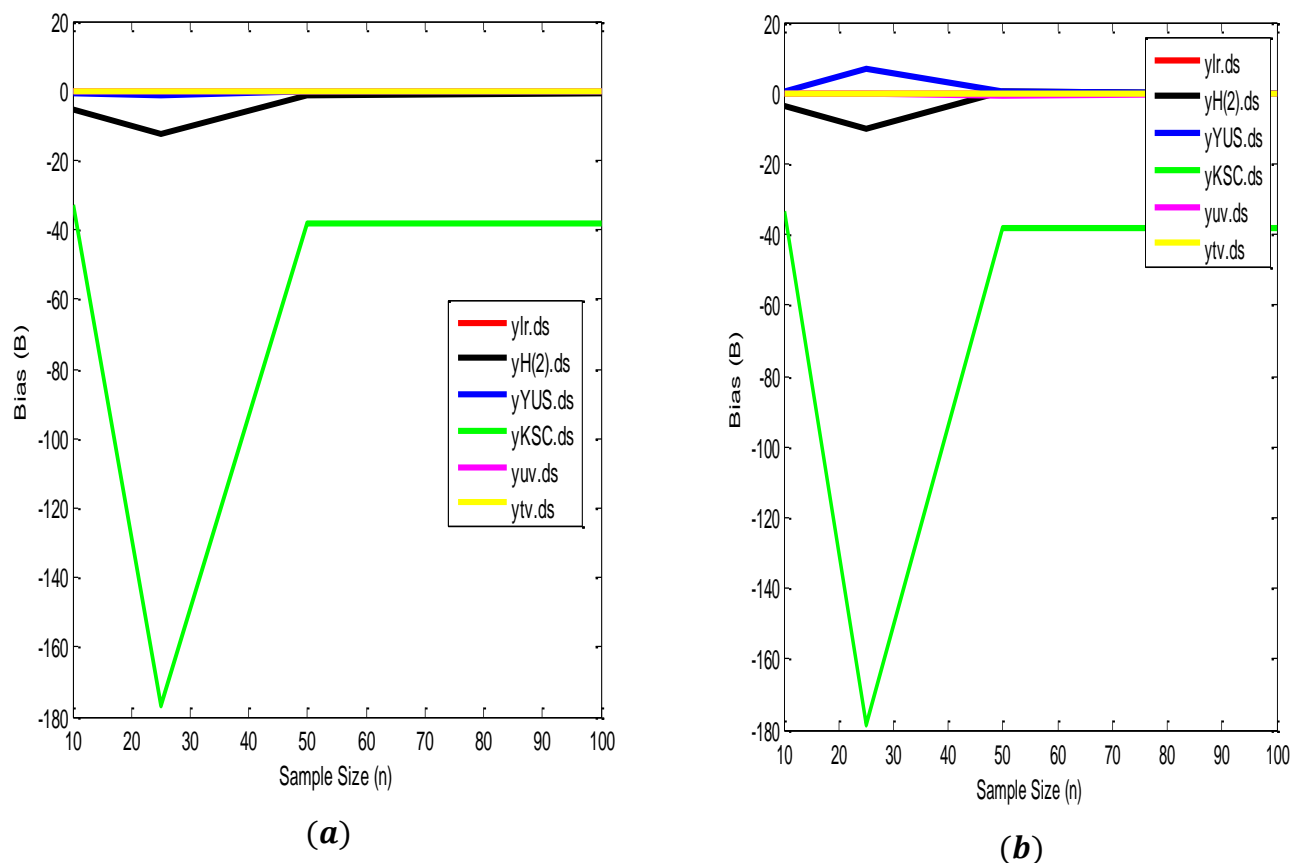


Figure 4: Plot of Bias of Estimators for Population II (a) ρ_{HLL} (b) ρ_{HHH}

Table 4: Ratio of Bias to Standard Errors of Estimators (Populations I-II)

		Population I				Population II			
		Sample size (n)				Sample size (n)			
Corr. Coeff.	Estimator	10	25	50	100	10	25	50	100
ρ_{HLL}	$\bar{y}_{lr.ds}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$\bar{y}_{H(2).ds}$	0.0128	0.0067	0.0093	0.0093	1.1133	2.492	0.5514	0.3675
	$\bar{y}_{YUS.ds}$	$8E-06$	$7E-06$	$1E-05$	$1E-05$	0.0436	0.0506	0.0134	0.0093
	$\bar{y}_{KSC.ds}$	0.0119	0.0132	0.0275	0.0218	1.8562	6.1408	3.6763	6.0178
	$\bar{y}_{uv.ds}$	$4E-05$	$1E-05$	$6E-08$	$5E-07$	0.0003	$1.31E-06$	0.0003	$1.04E-05$
ρ_{HHH}	$\bar{y}_{tv.ds}$	$6E-13$	$3E-13$	$1E-14$	$2E-15$	$3.45E-11$	$5.06E-11$	$2.25E-11$	$1.08E-11$
	$\bar{y}_{lr.ds}$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$\bar{y}_{H(2).ds}$	0.0255	0.0115	0.0065	0.0087	1.9216	3.0359	0.3096	0.2358
	$\bar{y}_{YUS.ds}$	$1E-05$	$2E-05$	$3E-05$	$4E-05$	0.0369	0.2925	0.0386	0.0332
	$\bar{y}_{KSC.ds}$	0.0103	0.0146	0.0219	0.0215	1.7432	4.9854	3.7871	5.8555
	$\bar{y}_{uv.ds}$	$4E-05$	$7E-05$	$3E-05$	0.0002	0.0605	0.0485	0.2667	0.0279
	$\bar{y}_{tv.ds}$	$2E-13$	$7E-14$	$5E-13$	$2E-13$	$5E-13$	$6E-11$	$1E-11$	$8E-11$

Table 5: Asymptotic Relative Efficiency (ARE) of Estimators (Populations I-II)

		Population I				Population II			
		Sample size (n)				Sample size (n)			
Corr. Coeff.	Estimator	10	25	50	100	10	25	50	100
ρ_{HLL}	$\bar{y}_{lr.ds}$	0.1118	0.1734	0.0369	0.0556	0.0546	0.0283	0.0877	0.0638
	$\bar{y}_{H(2).ds}$	0.9048	0.9431	0.9989	0.9922	1.1109	1.0189	0.9873	0.999
	$\bar{y}_{YUS.ds}$	0.1277	0.1730	0.0337	0.0492	0.0588	0.0293	0.0676	0.0561
	$\bar{y}_{KSC.ds}$	0.1251	0.2106	0.0385	0.0575	0.0760	0.0304	0.0800	0.0702
	$\bar{y}_{uv.ds}$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
ρ_{HHH}	$\bar{y}_{tv.ds}$	0.1166	0.1908	0.0462	0.0692	0.0570	0.0317	0.1083	0.0984
	$\bar{y}_{lr.ds}$	0.0549	0.1419	0.0521	0.0553	0.0515	0.0246	0.0451	0.0591
	$\bar{y}_{H(2).ds}$	1.0024	1.9008	1.0364	2.0676	7.1978	2.0216	1.8277	1.1760
	$\bar{y}_{YUS.ds}$	0.0851	0.2433	0.1072	0.1327	0.2427	0.0433	0.1525	0.2220
	$\bar{y}_{KSC.ds}$	0.0382	0.1050	0.0410	0.0448	0.0624	0.0169	0.0428	0.0587
	$\bar{y}_{uv.ds}$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\bar{y}_{tv.ds}$	0.0572	0.1554	0.0637	0.0674	0.0537	0.0273	0.0551	0.0872

Table 6: Parameter Estimates for Estimators using Real data (Population III) for Small Sample ($n = 10$)

Corr. Coeff.	Estimator	Mean	MSE	CV	Bias	RBSE
ρ_{HLL}	$\bar{y}_{lr.ds}$	5.8838	0.0065	1.3757	0.0000	0.0000
	$\bar{y}_{H(2).ds}$	5.9275	0.0206	2.4265	-0.000647	0.0045
	$\bar{y}_{YUS.ds}$	6.0066	0.0074	1.4401	0.000516	0.0059
	$\bar{y}_{KSC.ds}$	6.0138	0.0124	1.8528	0.000335	0.0030
	$\bar{y}_{uv.ds}$	5.8819	0.0086	1.5732	0.001169	0.0126
ρ_{HHH}	$\bar{y}_{tv.ds}$	6.0179	0.000046	0.1127	$1.11195E-06$	0.0002
	$\bar{y}_{lr.ds}$	6.1602	0.0108	1.6886	0.000	0.000
	$\bar{y}_{H(2).ds}$	6.6658	0.0342	2.7730	0.00031	0.0017
	$\bar{y}_{YUS.ds}$	6.6658	0.0212	1.9973	-0.00046	0.0032

	$\bar{y}_{KSC.ds}$	7.2845	0.0124	1.5286	0.00034	0.0031
	$\bar{y}_{uv.ds}$	6.1588	0.0155	2.0221	0.00212	0.0170
	$\bar{y}_{tv.ds}$	6.3013	8.3293E-05	0.1448	1.92319E-06	0.00021

DISCUSSION OF RESULTS

Population I (Simulated data)

Considering the first parameter subspace labeled ρ_{HLL} , one of the proposed estimators, $\bar{y}_{uv.ds}$ is perceived to be more efficient than other estimators throughout the parameter subspace while its variant, $\bar{y}_{tv.ds}$ is third best among the other competing estimators as $n \rightarrow \infty$ (Table 2 and Figure 1a). The proposed estimators are, asymptotically unbiased (Table 3 and Figure 3a), and consistent (Table 2 and Figure 1a). However, only estimator $\bar{y}_{uv.ds}$ is asymptotically efficient (Table 5).

In the second parameter space, ρ_{HHH} , The estimator, $\bar{y}_{H(2).ds}$ dominated the proposed and other estimators in the entire space (Table 2 and Figure 1b). Considering asymptotic properties, the proposed estimators are both asymptotically unbiased (Table 3 and Figure 1b), and consistent (Table 2 and Figure 2b), but again, only $\bar{y}_{uv.ds}$ is asymptotically efficient (Table 5). The ratio of bias to SE of all the competing estimators falls under the acceptable threshold ($RBSE \leq 0.1$).

Population II (Simulated data)

In the first parameter subspace, ρ_{HLL} , $\bar{y}_{H(2).ds}$ was apparently more efficient than the proposed and other estimators for small sample sizes. However, as $n \rightarrow \infty$, one of the proposed estimators, $\bar{y}_{uv.ds}$ dominated $\bar{y}_{H(2).ds}$ and the others throughout the subinterval (Table 2 and Figure 2a). It was however detected that ratio of bias to standard error of $\bar{y}_{H(2).ds}$ estimator is greater than the threshold of 0.1 (Table 4). This implies, estimated parameters by $\bar{y}_{H(2).ds}$ are likely to fall outside the 95% confidence intervals. Thus, the proposed estimators are preferable. The proposed estimators are again, asymptotically unbiased (Table 3 and Figure 3), and consistent (Table 2 and Figure 3a), once again, only $\bar{y}_{uv.ds}$ is asymptotically efficient (Table 5).

The most efficient estimator in the second subspace, ρ_{HHH} , is apparently, $\bar{y}_{H(2).ds}$ while one of the proposed estimators, $\bar{y}_{uv.ds}$ is the second most efficient (Table 1 and Figure 2b). However, RBSE of the estimator, $\bar{y}_{H(2).ds}$ exceeds 0.1 in the entire partition (Table 4). The implication is that, $\bar{y}_{H(2).ds}$ might have either underestimated or overestimated parameters in this interval (Sarmah *et al.*, 2013). Again the proposed estimators are preferable under this condition. Thus, the proposed estimators are respectively, asymptotically unbiased (Table 3 and Figure 4b), and consistent (Table and Figure 17b), but, only $\bar{y}_{uv.ds}$ is asymptotically efficient (Table 5).

Population III (Real data)

For Population III, only a sample size of $n = 10$ is considered. Here, one of the proposed estimators $\bar{y}_{tv.ds}$ outperformed all the other competing estimators under study in both the ρ_{HLL} and ρ_{HHH} parameter subspaces in terms of CV (Table 6). Its variant, $\bar{y}_{uv.ds}$ appears second in ρ_{HLL} subspace while $\bar{y}_{KSC.ds}$ is second in terms of CV in the ρ_{HHH} subspace. Thus the proposed estimators are preferable.

In view of the performance of the proposed and other competing estimators in Two-phase sampling, it is evident that the proposed estimators are more efficient than most of the existing estimators considered in this study namely, the classical regression estimator, (25) estimator, (31) estimator and (29) estimator.

CONCLUSION

This study proposed two new estimators for estimating the population mean in two-phase sampling. The

performance of the proposed estimators was evaluated using simulated and real data. The results show that the proposed estimators, particularly $y_{uv.ds}^-$, outperform most of the existing estimators considered in this study, including the classical regression estimator and several other estimators. The proposed estimators are asymptotically unbiased, consistent, and efficient, with $y_{uv.ds}^-$ being the most efficient in most cases. The ratio of bias to standard error of the proposed estimators falls within the acceptable threshold, indicating that they provide reliable estimates. Overall, the findings of this study suggest that the proposed estimators are preferable for estimating the population mean in two-phase sampling, particularly when the sample size is large.

REFERENCES

1. Uba, T., A. J. Ikughur and S. C. Nwaosu (2021). An Alternative Hybrid Estimator of Finite Population Mean in Simple Random Sampling. *Asian Journal of Probability and Statistics*, 15(4): 276-298.
2. Kalton, G. (2002). Models in the Practice of Survey Sampling (Revisited). *Journal of Official Statistics*, 18(2): 129-154
3. Diana, G., and Perri, P. F. (2007). Estimation of Finite Population Mean Using Multi-Auxiliary Information. *METRON- International Journal of Statistics*, 65(1): 99-112.
4. Ikughur, A. J., Ikyaagba, B. K., and Nwaosu, S. C. (2017). Exponential Ratio-Product Type Estimators in Stratified Random Sampling. *Journal of the Nigerian Statistical Association*, 29: 35-46.
5. Riaz, N., Noor-Ul-Amin, M. and Hanif, M. (2014). Regression-Cum-Exponential Ratio Type Estimators for the Population Mean. *Middle East Journal of Scientific Research*, 19(12): 1716-1721.
6. Neyman, J. (1938). Contribution to the Theory of Sampling Human Populations. *Journal of the American Statistical Association*, 33(201): 101-116.
7. Cochran, W. G. (1940). The Estimation of the Yields of the Cereal Experiments by Sampling for Ratio of Grain to Total Produce. *The Journal of Agricultural Science*, 30: 262-275.
8. Kanwai, M. E., Ashiribi, O. E., and Isah, A. (2016). Use of Auxiliary Variable and Asymptotically Optimum Estimators to Increase Efficiency of Estimation in Double Sampling. *International Journal of Statistics and Probability*, 5(3): 55-62.
9. Onyeka, A. C. (2013). A Class of Product-Type Exponential Estimators of the Population Mean in Simple Random Sampling Scheme, *Statistics in Transition*, 14(2): 189-200.
10. Saini, M., and Kumar, A. (2015). Exponential Type Product Estimator for Finite Population Mean with Information on Auxiliary Attribute. *Applications and Applied Mathematics*, 10(1): 106-113.
11. Mishra, S. S. (2018). Generalized Square Root Transformation and Exponential Ratio Type Estimator for Finite Population Mean in Sample Survey. *Bulletin of Mathematics and Statistics Research*, A peer reviewed International Journal, 6(2): 1-7.
12. Bahl, S., and Tuteja, R. K. (1991). Ratio and Product Type Exponential Estimator. *Information and optimization sciences*, 12: 159-163.
13. Singh, R., Chauhan, P., Sawan, N., and Smarandache, F. (2009). Improvement in Estimating the Population Mean using Exponential Estimator in Simple Random Sampling. *Bulletin of Statistics and Economics*, 3(9): 13-18.
14. Grover, L. K., and Kaur, P. (2011). An Improved Estimator of the Finite Population Mean in Simple Random Sampling. *Model Assisted Statistics and Applications*, 6(1): 47-55.
15. Rashid, R., Noor-Ul-Amin, M., and Hanif, M. (2015). Exponential Estimators for Population Mean using Transformed Variables. *Applied Mathematics and Information Sciences-an International Journal*, 9(4): 2107-2112.
16. Singh, H. P. and Espejo, R. (2003). On Linear Regression and Ratio Product Estimation of a Finite Population Mean. *The Statistician*, 52: 59-67.
17. Samiuddin, M., and Hanif, M. (2007). Estimation of Population Mean in Single and Two-Phase Sampling with or Additional Information. *Pakistan Journal of Statistics*, 23: 99-118.
18. Malik, K. A., and Tailor, R. (2013). Ratio type Estimator of Population Mean in Double Sampling. *International Journal of Advanced Mathematics and Statistics*, 1 (1):34-39.

19. Yadav, S. K., Mishra, S. S., Tiwari, V., and Shukla, A. K. (2014). Computational Approach to Generalized Ratio Type Estimator of Population Mean under Two Phase Sampling. *International Journal on Recent Trends in Computing and Communication*, 2(10): 3013-3017.
20. Handique, B. K. (2012). A class of Regression-cum-Ratio Estimators in Two-Phase Sampling for Utilizing Information from High Resolution Satellite Data. *ISPRS Annals of the Photogrametry, Remote Sensing and Spatial information Sciences*, 1(4): 71-76.
21. Deepan, G., Upasana, G. and B. K. Singh (2024) Exponential chain ratio estimator of finite population mean with imputation of missing data under two-phase sampling, *Advances and Applications in Statistics*, 91(9): 1113-1135. <https://doi.org/10.17654/0972361724059>
22. Dash, P. and Sunani, K. (2022) An Improved Class of Mixed Estimators of Population Mean under Double Sampling. *Indian Journal of Science and Technology* 15(7): 276-291. <https://doi.org/10.17485/IJST/v15i7.1396>
23. Tahir, M., Yude, B., Bashir, S., Hussain, S & Munir, T. (2023). A New Improved Estimator for the Population Mean using Twofold Auxiliary Information under Simple Random Sampling. *Management Science Letters*, 13(4): 265-276.
24. Hamad, N., Hanif, M., and Haider, N. (2013). A Regression Type Estimator with Two Auxiliary Variables for Two-Phase Sampling. *Open Journal of Statistics*, 3: 74-78.
25. Hanif, M., Hamad, N. and Shabbaz, M. Q. (2009). A Modified Regression Type Estimator in Survey Sampling. *World applied Science Journal*, 7(12): 1559-1561.
26. Singh, P., and Vishwakarma, G. K. (2007). Modified Exponential Ratio and Product Estimators for Finite Population Mean in Double Sampling. *Austrian Journal of Statistics*, 36: 217-225.
27. Estimator in Two-Phase Sampling using Auxiliary Information. *Statistical*, 73(2): 221-234.
28. Kalita, D., Singh, B. K., and Choudhary, S. (2013). Improved Exponential Chain Ratio and Product-Type Estimators for Finite Population mean in Double Sampling. *International Journal of Statistical Sciences*, 13: 21-37.
29. Cochran, W. G. (1977). *Sampling Techniques*. (3rd Ed.), USA, Wiley Eastern limited, Pp. 428.
30. Yadav, R., Upadhyaya, L. N., Singh, H. P., and Chatterjee, S. (2013). A Chain Ratio Exponential Type Estimator in Two-Phase Sampling using Auxiliary Information. *Statistical*, 73(2): 221-234.