Abstract

In this paper, we propose an alternative hybrid estimator of finite population mean in simple random sampling without replacement (SRSWOR). This proposed estimator is a modification of Rashid et al. [1] estimator. The expressions for the bias and Mean Square Error (MSE) of the estimator are derived. A comprehensive simulation study to show the efficacy of the estimator as compared to conventional estimators using Coefficient of Variation as a performance measure. The results are also supported with empirical illustrations using real life data which have shown that the proposed estimator was more efficient than almost all the existing estimators considered in this study.

Keywords: Auxiliary variable; hybrid estimators; mean square error; ratio estimators; regression estimators.

1 Introduction

The use of auxiliary information in estimation of population mean, total, or ratio got a boost when Bahl and Tuteja [2] introduced their exponential ratio and product estimators of population mean respectively. These estimators use a single auxiliary variable and produce more efficient estimates than the usual existing
estimators. As noted by Rashid et al. [1], exponential estimators are preferable to classical ratio and product estimators, especially when the linear relationship between the variable of interest and the auxiliary variable is weak. Several authors have over the years proposed estimators based on exponentiation of the traditional ratio, product and regression estimators respectively or a mixture of these.

Many other authors, have in one way or the other tried to make significant improvements on the efficiency of their ratio estimators by making use of the parameters of the auxiliary variables and known constants to propose new ratio estimators. For example, Singh and Tailor [4,5] in separate works proposed improved ratio estimators of finite population by utilizing known correlation coefficients and coefficient of variations respectively of the auxiliary variable. Kadilar and Cingi [5,6] made use of coefficient of Kurtosis, coefficient of variation and correlation coefficient and their combinations to construct improved ratio estimators of population mean. The work by Jitthavech and Lorchirachoonkul [7] investigated four new estimators in simple random sampling, biased sample mean, ratio estimator and two linear estimators utilizing known coefficients of variation of the variable of interest and the supplementary variable. The results of the study shows that the two proposed linear regression estimators are more efficient than the new biased sample mean and at least as efficient as the three traditional estimators (sample mean, ratio and product) and Sisodia and Dwivedi [8] estimator. The results further proved that at least one of the proposed linear regression estimators is always more efficient than the new ratio estimator and Searls sample mean.

In most of the regression type estimators encountered in literature, only a few have dealt with the problem of handling large scale data [9] (Hanif et al., 2010; Kanwai et al., 2016), but all of them are univariate. Authors are more interested in developing estimators that are more efficient than existing ones in terms of minimum Mean Square Error (MSE), relative efficiency or coefficient of variation using available small sample data sets. However, due to demand from big tech industries, governmental and NGOs, producing large chunks of data, there is every need for researchers to rise to the task of evolving estimators that are asymptotic in nature.

This study therefore, modifies an existing estimator of finite population mean by Rashid et al. [1] to construct a hybrid regression-cum-ratio exponential type estimators of finite population in simple random sampling using two supplementary variables. Four partitions of the correlation coefficient parameter space are considered under varying sample sizes to investigate the effect of correlation coefficient and ascertain asymptotic properties of the proposed estimators as compared with some selected existing estimators.

2 Literature Review

According to Kanwai et al. (2016), the oldest estimator of population mean in the history of sampling survey under simple random sampling is the sample mean, defines as the sum of all the possible observations or units of a given sample divided by the total number of units in the sample. For simple random sampling, it is obvious that this sample mean estimator and are consistent estimates of the population mean and the mean total, respectively. Searls (1964) proposed a modified version of the sample mean estimator, where is a constant. This modified estimator like the sample mean estimator is unbiased, consistent and has been proved to be more efficient than the sample mean estimator.

The aim of sample survey is to get information about the population by taking random samples from the population. Population could be considered as a collection of units defined according to the objective of the study. Estimation of the population mean is a tenacious issue in sampling surveys and several efforts have been made by many researchers to improve the accuracy or precision of the estimates by using supplementary information. In sampling survey, information may or may not be readily available on every unit of the population under consideration. If there exists a variable whose attributes are known for every unit of the population but is not the study variable but rather can be used to improve the sampling plan or to enhance the estimation of the variable of interest, then this particular variable is called an auxiliary variable. The auxiliary variable about a study population may include a known variable to which the variable of interest (study variable) is approximately related (positively or negatively). This information could be used at the planning stage, the estimation stage or both [11]. The estimation of population parameters with greater precision is a relentless issue in sampling theory and the precision of the estimates can be improved by increasing the sample size, but by doing so tends to sabotage the essence of sampling (saving time, labor and cost). Thus, an alternative is to employ the use of auxiliary (supplementary) variables with a combination of an appropriate estimation
procedure in order to increase the precision of the estimates. The auxiliary variable must be closely related to the study variable and the employed estimator must be asymptotically optimum (Kanwai et al., 2016).

John Graunt, cited in Riaz, et al. [11] was the first believed to have used the auxiliary information to estimate the first realistic estimates of the number of men and women in London and the whole population of England and showed that both were increasing, with steady migration into London. However, Neyman (1934) study may be referred to as an initial work where auxiliary information has been discussed in detail whereas Watson [12] made use of the regression method of estimation to estimate the average area of the leaves on a plant. Cochran [13,14] used auxiliary information in single phase sampling at estimation stage to develop the classical ratio type estimator for the estimation of population mean as well as the regression estimator respectively. The ratio estimator is more efficient as compared to the sample mean estimator provided the auxiliary variable and the variables of interest are highly positively correlated and the regression line passes through the origin.

While Robson [15] and Murthy (1964) worked independently on the classical product estimator of the population mean. The product type estimator like the ratio estimator is more efficient than the sample mean estimator, in situations where the auxiliary variable has strong negative correlation with the variable of interest. If, however, the regression line has an intercept, the regression estimator is preferable to the both the ratio and product estimators as the case may be applicable. The historical development on the improvements of the ratio method of estimation was done by Sen [16]. These improved ratio estimators, though, biased, are more efficient than the classical ratio estimator.

Rashid et al. [1] suggested two exponential type, ratio-cum-ratio and product-cum-product class of estimators of finite population mean. These estimators are a product of the study variable and exponent of the linear combination of two auxiliary variables such that the sum of the constants is unity. They firstly developed the generalized forms of the estimators and discussed special cases and conditions under which they produce optimum estimates. Meanwhile, Shabbir et al. [17] and Haji and Lata [18] worked independently to improve the difference estimator through exponentiation. The new improved estimator was achieved by averaging exponential ratio and product estimators respectively, after drawing inspiration from Yadav and Kadilar [19] estimator. This new estimator was validated by using ten different real datasets. Similar exponential ratio type estimators were developed by Singh and Vishwakarma [9], Vishwakarma and Kumar [20] and Singh and Khahid [21] with application in two phase sampling.

On the other hand, Kumar et al. [22] proposed a class of exponential chain type ratio estimator for population mean with imputation of missing data in Two-Phase sampling. The work dealt with the challenge of non-response in situations where the information on another additional auxiliary is available alongside the main auxiliary variable.

Hamad et al. [23], in extending the work done by Hanif et al. (2009) developed a regression type estimator with two auxiliary variables for two-phase sampling when there is no available information about the auxiliary variables at the population level. This estimator is a product of the classical regression estimator, and the linear combination of two ratio estimators. To avoid the problem of multi-linearity, they assumed there is minimum correlation between the supplementary variables.

Saini and Kumar [24] estimator is a modified unbiased exponential type product estimator of the population mean. This particular estimator has a unique property of a bi-serial correlation between the variable of interest and auxiliary attributes. By using a linear combination of two auxiliary variables, Lu et al. [25] presented a new exponential type estimator. The chosen weights satisfy the condition that their sum equals unity and by employing Tailor series method, obtained the bias and the MSE by first order approximation.

Yadav et al. [26] showed a deviation from estimation of population mean to that of population variance using auxiliary variables which was achieved by utilizing the auxiliary information in the context of coefficient of kurtosis and the population mean of the auxiliary variable. Meanwhile, Jabbar et al. [27] developed an exponential estimator of population variance in two stage sampling under the conditions where sum of the weights was not equal to unity and secondly when the sum of weights equals unity.

Interestingly, Yadav and Misra [28] constructed an exponential estimator for population mean using median of the variable of interest. This estimator appeared useful in practical situations where it is difficult to get
information on the mean of the study variable from the population. Mishra [29] suggested a more generalized square root transformed ratio type estimator and exponential ratio type estimator similar to Gupta et al. [30] estimator except that it combined two ratio estimators in the linear combination and two ratio estimator in the exponential component. Meanwhile, Riaz et al. [11] had constructed regression-cum-ratio/product exponential type estimator by combining the concept of Bahl and Tuteja [2] and the regression estimator.

On the other hand, many authors have used median, coefficient of kurtosis, coefficient of skewness, deciles, quartiles deviation, etc. [3,8,28,31-38].

Abid et al. [39] proposed a new linear combinations of ratio type estimators in SRS using non-conventional measures such as Hodges Lehman estimator, population mid-range and population tri-mean as a supplementary information. It is however observed that upon all these efforts, none of these seemed to have greater efficiency than the regression estimator, but some had greater gain in efficiency as compared to the classical ratio estimator. Motivated by Jeelani et al. [34], Misra et al. [40] developed an improved ratio type estimator of population mean using predictive approach of estimation by using linear combination of the coefficient of skewness and quartile deviation of the auxiliary variable. Motivated by the work done by Kadilar and Cingi [5], Subzar et al. [36] developed a new class of more efficient ratio estimators utilizing the linear combination of coefficient of skewness and population deciles in place of the coefficient of Kurtosis and variation respectively. In another vein Onyeka et al. [41] constructed a class of estimators for population ratio (R) in simple random sampling scheme using transformation of the auxiliary variable principle. The study shows large gains in efficiency over traditional ratio and product estimators depending on whether there is strong positive or negative relationship between the study variable and the auxiliary variable. Diana and Perri [42] estimator of population mean utilizes both known mean and variance of p auxiliary variables to estimate population mean of study variable, that is, it uses multi-auxiliary variable with known mean and variances.

In another development, Hassan et al. [43] developed a regression type estimator for either positive or negative correlation between variable of interest and supplementary variables. Unlike the usual regression estimator, this estimator is more efficient than the classical ratio and product estimators respectively irrespective of the nature of correlation coefficient. Recently, Shabbir et al. [44] proposed a ratio-exponential-log type estimator of finite population mean in simple random sampling using two auxiliary variables when the population parameters are known. Relatedly, Ahmad et al. (2021), constructed an improved class of estimators of finite population mean in both simple random sampling and stratified Two-phase sampling using population proportion as attribute. Zaman and Kadilar [45] ratio and product estimator in Stratified Two-phase sampling considered two cases: when second sample of size n is drawn from first sample of size n', and when the second and first samples are drawn independently from the parent population of size N. Similarly, Vishwakarma and Zeeshan [46] proposed generalized ratio-cum-product estimator for finite population mean under Two-phase sampling using optimal samples sizes for the given cost function. While Hussain et al. [47] provided an improved version of the Bahl and Tuteja [2] ratio estimator, Etebong et al. [48] introduces a new method of producing more accurate and efficient estimates of ratio and product estimators that are considerably adjustable to both negatively and positively correlated populations.

This study modifies Rashid et al. [1] estimator by replacing the sample mean with the regression mean estimator to form a regression and ratio exponential type estimator of finite population mean in simple random sampling without replacement. It further applies a transformation due to Srivenkataramana [49] to ascertain whether or not the efficiency of the proposed estimator is improved.

3 Preliminaries and Notations

Consider a finite population, \( U = \{U_1, U_2, ..., U_N\} \). Suppose that a sample of size n is drawn from this population using Simple Random Sampling without replacement (SRSWOR) scheme. Let y be the study variable of interest, x and z, be the respective auxiliary variables and \( y_i, x_i \) and \( z_i \) be the observations in the \( i^{th} \) unit of the study variable and the two auxiliary variables under consideration.

Define \( e_y \) as error term of the study variable; \( e_x \); error term of the x variable; \( e_z \); Error term of the z variable; \( \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \); sample mean of study variable; \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \); sample mean of x variable; \( \bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i \); sample
mean of z variable; $\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i$; population mean of study variable; $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} x_i$; population mean of x variable; $\bar{Z} = \frac{1}{N} \sum_{i=1}^{N} z_i$; population mean of z variable; $S^{2}_{X} = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2$; population variance of x variable; $S^{2}_{Z} = \frac{1}{N-1} \sum_{i=1}^{N} (Z_i - \bar{Z})^2$; population variance of z variable; $S_{yx} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{X})(y_i - \bar{Y})$; population covariance between x and y; $S_{yz} = \frac{1}{N-1} \sum_{i=1}^{N} (z_i - \bar{Z})(y_i - \bar{Y})$; population covariance between y and z;

$S_{xz} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{X})(z_i - \bar{Z})$: population covariance between x and z; $\rho_{yz} = \frac{S_{yz}}{S_{y}S_{z}}$: correlation coefficient between y and z denoted by $\rho_{1}$; $\rho_{x} = \frac{S_{xy}}{S_{x}S_{y}}$: correlation coefficient between x and y denoted by $\rho_{2}$; $\rho_{xz} = \frac{S_{xz}}{S_{x}S_{z}}$: correlation coefficient between x and z denoted by $\rho_{3}$; $C_{y} = \frac{S_{y}}{\bar{Y}}$: coefficient of variation of the study variable; $C_{x} = \frac{S_{x}}{\bar{X}}$: coefficient of variation of x variable;

$C_{z} = \frac{S_{z}}{\bar{Z}}$: coefficient of variation of z variable. Furthermore, let $E(e_{i}) = 0$ for $(i = x, y, z)$; $E(e_{y}^{2}) = \theta C_{y}^{2}$; $E(e_{x}^{2}) = \theta C_{x}^{2}$; $E(e_{y}^{2}) = \theta C_{y}^{2}$; $E(e_{x}e_{y}) = \theta \rho_{yx}C_{x}C_{y}$; $E(e_{x}e_{z}) = \theta \rho_{xz}C_{x}C_{z}$; $E(e_{y}e_{z}) = \theta \rho_{yz}C_{y}C_{z}$ where, $\theta = \frac{1}{n} - \frac{1}{N}$

4 Existing Ratio-exponential Estimators in Simple Random Sampling

(i) Classical regression estimator

Cochran [14] estimator of finite population mean uses one auxiliary variable like the classical ratio estimator but produces more efficient estimates when the regression line has an intercept. It is an unbiased estimator given as:

$$\bar{y}_{yr} = \bar{y} + \beta_{y,x}(\bar{X} - \bar{x}) \tag{1}$$

The MSE of the classical regression estimator is given by:

$$MSE(\bar{y}_{yr}) = \bar{Y}^{2}C_{y}^{2}\theta(1 - \rho_{y,x}^{2}) \tag{2}$$

(ii) Exponential ratio-cum-ratio estimator

Rashid et al. [1] used two transformed auxiliary variables to develop this estimator under single phase sampling which is an improvement on Bahl and Tuteja [2] estimator. It is given as:

$$\bar{y}_{RAG} = \bar{y}exp\left[a\left(\frac{\bar{x}^{*} - \bar{X}}{\bar{X} + \bar{x}^{*}}\right) + b\left(\frac{\bar{z}^{*} - \bar{Z}}{\bar{Z} + \bar{z}^{*}}\right)\right] \tag{3}$$

where, $a = \frac{2\bar{Y}(\rho_{yx} - \rho_{yx}\rho_{xz})}{g\bar{C}_{x}(1 - \rho_{xz}^{2})}; b = \frac{2\bar{Y}(\rho_{yz} - \rho_{yz}\rho_{xz})}{g\bar{C}_{z}(1 - \rho_{xz}^{2})}; g = \frac{n}{N-n}$. $\bar{x}^{*}$ and $\bar{z}^{*}$ are transformed auxiliary variables such that $\bar{x}^{*} = (1 - g\bar{e}_{x})\bar{X}$ and $\bar{z}^{*} = (1 - g\bar{e}_{z})\bar{Z}$. The bias and MSE are respectively defined as:

$$Bias(\bar{y}_{RAG}) = \frac{\bar{Y}}{B}\left\{g^{2}[C_{y}^{2}(2abK_{xz} + b^{2}) + a^{2}C_{x}^{2}] - 4g(aK_{xz}C_{y}^{2} + bK_{yz}C_{z}^{2})\right\} \tag{4}$$

$$MSE(\bar{y}_{RAG}) = \bar{y}^{2}C_{y}^{2}\left[\frac{1 - (\rho_{yx}^{2} + \rho_{yz}^{2} - 2\rho_{yx}\rho_{yz})}{1 - \rho_{xz}^{2}}\right] \tag{5}$$
where, \( K_{y,x} = \rho_{y,x} C_y C_x \), \( \theta = \frac{1}{n} - \frac{1}{N} = \frac{N-n}{Nn} \).

(iii) The Exponential ratio type estimator

Ekpenyong and Enang (2015) exponential ratio type estimator is an improvement on the classical regression and ratio estimators. It is preferable to both the classical regression and ratio estimators respectively, in situations where there is low positive correlation between the study variable and the auxiliary variable. It is given as:

\[
\bar{y}_{EE} = \theta_1 \bar{y} + \theta_2 (\bar{X} - \bar{x}) \exp \left( \frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}} \right)
\]

where: \( \theta_1 \) and \( \theta_2 \) are suitably chosen scalars, such that \( \theta_1 > 0 \) and \( -\infty < \theta_2 < \infty \). Its MSE and bias are given as:

\[
B(\bar{y}_{EE}) = \bar{y} \left[ (\theta_1 - 1) + \theta_2 K_{y,x} \theta \frac{C_y^2}{2} \right]
\]

\[
MSE(\bar{y}_{EE}) = \bar{y}^2 [1 + \theta_1^2 \gamma_1 - 2\theta_1 - 2\theta_1 \theta_2 m \gamma_2 - 2\theta_2 m^2 \gamma_3 + \theta_2^2 m^2 \gamma_4]
\]

where: \( \gamma_1 = 1 + \theta C_y^2 \); \( \gamma_2 = \theta C_y^2 \left( K_{y,x} - \frac{1}{2} \right) \); \( \gamma_3 = \theta \frac{C_y^2}{2} \); \( \gamma_4 = \theta C_y^2 \); \( m = \frac{\bar{X} + \bar{x}}{\bar{y}} \); \( \theta_1 = \frac{y_1 + y_2 y_3}{y_1 y_4 - y_2} \); \( \theta_2 = \frac{R(y_2 + y_1 y_3)}{y_1 y_4 - y_2} \); \( R = \frac{\bar{y}}{\bar{X}} \).

(iv) Exponential regression-ratio/product estimator

Riaz et al. [11] developed the regression-ratio/product exponential estimator by combining the concept of Bahl and Tuteja [2] exponential type estimator and the classical regression estimator. It is given as:

\[
\bar{y}_{RNH} = [\bar{y} + a_1 (\bar{Z} - \bar{x})] \exp \left[ \gamma \frac{\bar{X} - \bar{x}}{\bar{X} + (b_1 - 1) \bar{x}} \right]
\]

where \( a_1, b_1 \) are real positive constants and \( \gamma \) may take the values -1 and 1.

\[
Bias(\bar{y}_{RNH}) = a_1 \gamma \theta \bar{Z} \rho_{x,z} \frac{C_y C_z}{b_1} + \frac{\theta \gamma \bar{y} C_y^2}{2 b_1^2} [1 + 2(b_1 - 1) - 2b_1 K_{y,x}]
\]

where:

\[
a_1 = \frac{\bar{y}}{\bar{Z}} \left( K_{x,z} - K_{y,z} K_{y,x} \right) \text{ and } b_1 = \frac{\gamma (1 - \rho_{x,z}^2)}{K_{y,x} - K_{x,z} K_{y,x} \frac{C_y^2}{C_z^2}}
\]

\[
MSE(\bar{y}_{RNH}) = \bar{y}^2 C_z^2 \theta \left( 1 - \gamma^2 \rho_{x,z}^2 - \rho_{y,z}^2 - \gamma^2 \rho_{y,x}^2 + 2\gamma^2 \rho_{y,x} \rho_{x,z} \rho_{y,z} \right)
\]

5 Proposed Estimator

The proposed alternative hybrid regression-cum-ratio exponential type estimators of finite population mean modifies Rashid et al. [1] estimator. Here, the sample mean is replaced with the classical regression estimator. Further, it is assumed that \( z \) and \( x \) have strong and weak positive relationship with study variable respectively.

The estimator is given as:

\[
\bar{y}_{uv} = [\bar{y} + \beta (\bar{Z} - \bar{x})] \exp \left[ a \left( \frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}} \right) \right]
\]
where, $\alpha$ and $\beta$ suitably chosen constants such that $\text{MSE}(\bar{y}_{uv})$ is minimized.

**Theorem 1:** The bias of the proposed alternative hybrid estimator when auxiliary variables are not transformed is given as:

$$\text{Bias}(\bar{y}_{uv}) = \bar{y} \left( \frac{1}{B} \alpha^2 \theta C_x^2 - \frac{1}{2} \alpha \rho_{yx} C_x C_y \right) + \frac{1}{2} \beta \tilde{Z} \alpha \theta \rho_{xz} C_x C_z$$  \hspace{1cm} (12)

**Theorem 2:** The MSE of the proposed alternative hybrid estimator when auxiliary variables are not transformed is given as:

$$\text{MSE}(\bar{y}_{uv}) = \bar{y}^2 \theta C_y^2 (\rho_{xz}^2 - 1)^{-1} (\rho_{yz}^2 + \rho_{xz}^2 + \rho_{yz}^2 - 2 \rho_{yz} \rho_{xz} \rho_{yz} - 1)$$  \hspace{1cm} (13)

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

When the auxiliary variables are transformed, our proposed estimator becomes:

$$\tilde{y}_{uv} = [\bar{y} + \beta_1 (\bar{x} - \hat{X})] \exp \left[ \alpha_1 \left( \frac{\bar{x} - \hat{X}}{\bar{x} + \bar{X}} \right) \right]$$  \hspace{1cm} (14)

where, $\alpha_1$ and $\beta_1$ are suitably chosen constants that minimize $\text{MSE}$ of $\tilde{y}_{uv}$. $\bar{x}^*$ and $\bar{z}^*$ are transformed auxiliary variables such that $\bar{x}^* = (1 - g_{x}) \bar{X}$ and $\bar{z}^* = (1 - g_{z}) \bar{Z}$ and $g = \frac{n}{n-n}$.

**Theorem 3:** The bias of the proposed alternative hybrid estimator when auxiliary variables are transformed is given as:

$$\text{Bias}(\tilde{y}_{uv}) = \bar{y} \left[ \frac{1}{B} \alpha_1^2 \theta C_z^2 - \frac{1}{2} \alpha_1 g \theta \rho_{yx} C_x C_y \right] + \beta_1 \tilde{Z} \left[ \frac{1}{2} \alpha_1 g^2 \theta \rho_{xz} C_x C_z \right]$$  \hspace{1cm} (15)

**Theorem 4:** The MSE of the proposed alternative hybrid estimator when auxiliary variables are transformed is given as:

$$\text{MSE}(\tilde{y}_{uv}) = \bar{y}^2 \theta C_y^2 (\rho_{xz}^2 - 1)^{-1} (\rho_{yz}^2 + \rho_{xz}^2 + \rho_{yz}^2 - 2 \rho_{yz} \rho_{xz} \rho_{yz} - 1)$$  \hspace{1cm} (16)

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

**NB:** Proofs of Theorems 1-4 are found in Appendices 1-4 below.

**Corollary 1:** The bias of the alternative hybrid estimator when the auxiliary variables are not transformed, $\text{Bias}(\bar{y}_{uv})$ is -1 times the bias of the alternative hybrid estimator when the auxiliary variables are transformed, $\text{Bias}(\tilde{y}_{uv})$. i.e. $\text{Bias}(\bar{y}_{uv}) = -\text{Bias}(\tilde{y}_{uv})$

**Corollary 2:** The MSE of $\bar{y}_{uv}$ the alternative hybrid estimator when the auxiliary variables are transformed, is independent of $g$ and equals the MSE of $\bar{y}_{uv}$ the alternative hybrid estimator when the auxiliary variables are not transformed. i.e. $\text{MSE}(\bar{y}_{uv}) = \text{MSE}(\tilde{y}_{uv})$

### 6 Efficiency Comparison of Estimators

This study employs the Coefficient of Variation (CV) to compare performance of estimators considered in this study. Bowerman [50] defined Coefficient of Variation as a statistical tool used to measure the size of the standard deviation relative to the size of the population or sample mean. This is given as:

$$CV = \frac{\text{Var}(X)}{X} \times 100\%$$
For estimators that are biased, the Coefficient of Variation is given as:

\[
CV = \sqrt{\frac{MSE(X)}{\bar{X}}} \times 100%
\]

The estimator with the least CV is considered the “best” in the class of estimators.

**7 Empirical Study**

To investigate the performance of various estimators of population mean \( \bar{Y} \) of study variable \( y \), we generated synthetic data generated according to the Uniform Distribution with the following statistics:

**Statistics of Study Populations:**

**7.1 Population I [Source: Generated according to Normal Distribution using RNG in Excel]**

\[
N = 1000; \bar{Y} = 100.0786; n_1 = 10, n_2 = 25, n_3 = 50, n_4 = 100; \bar{X} = 25.90251; \\
\bar{x}_1 = 25.42673; \bar{x}_2 = 26.29295; \bar{x}_3 = 24.82662; \bar{x}_4 = 27.15513; \bar{Z} = 75.27509; \\
\bar{z}_1 = 59.3777; \bar{z}_2 = 82.72477; \bar{z}_3 = 70.25747; \bar{z}_4 = 89.85487; C_x = 0.71745; \\
C_y = 0.55689; C_z = 0.56869; \rho_{yx} = -0.0198; \rho_{yz} = 0.009504; \rho_{xz} = -0.0036
\]

**7.2 Population II [Source: Gul, 1991]**

This data is a study of the effect of Managing Accounting System (X) and Perceived Environmental Uncertainty (Z) on small business manager’s Perception of their performance (Y).

For this data we have:

\[
N = 40; \bar{Y} = 6.1532; n_1 = 10, n_2 = 25, n_3 = 35, \bar{X} = 4.4659; \bar{Z} = 4.3583; \bar{x}_1 = 4.3847 \\
\bar{z}_2 = 4.6830; \bar{z}_3 = 4.6589; \bar{z}_4 = 4.5185; \bar{z}_5 = 4.296; \bar{z}_6 = 4.5191; C_y = 0.7977; C_x = 0.8531 ; C_z = 0.9601; \rho_{yx} = -0.02495; \rho_{yz} = -0.01002; \rho_{xz} = 0.160939
\]

We denote the four correlation coefficient partitions by the following:

(i) \( \rho_{HLL} \) is the region, (0.7 < \( \rho_1 < 1 \) and \( 0 < \rho_2, \rho_3 < 0.5 \))
(ii) \( \rho_{HHH} \) is the region, (0.7 < \( \rho_1, \rho_2, \rho_3 < 1 \))
(iii) \( \rho_{LLL} \) is the region, (0 < \( \rho_1, \rho_2, \rho_3 < 0.5 \))
(iv) \( \rho_{LLH} \) is the region (−0.5 < \( \rho_1, \rho_2, \rho_3 < 0 \))

where, \( \rho_1 = \rho_{yx}, \rho_2 = \rho_{yx} \) and \( \rho_3 = \rho_{xz} \)

**7.3 Statistical Software used for Data Analysis**

All calculations in this work were implemented in Maple 7 while data analysis was done in Excel, the graph were drawn using Matlab, 2007.
Table 1. Estimates of population I mean as $n \to \infty$

<table>
<thead>
<tr>
<th>S/No</th>
<th>Corr. Coeff.</th>
<th>Estimator</th>
<th>Sample Size (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>$\rho_{HLL}$</td>
<td>$\bar{y}_{uv}$</td>
<td>147.4198</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rah}$</td>
<td>153.4693</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rnh}$</td>
<td>194.1192</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{tv}$</td>
<td>145.7691</td>
</tr>
<tr>
<td></td>
<td>$\rho_{HHH}$</td>
<td>$\bar{y}_{uv}$</td>
<td>146.7985</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rah}$</td>
<td>146.7881</td>
</tr>
<tr>
<td>2</td>
<td>$\rho_{LLL}$</td>
<td>$\bar{y}_{uv}$</td>
<td>135.2068</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rah}$</td>
<td>103.7209</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rnh}$</td>
<td>103.0653</td>
</tr>
<tr>
<td></td>
<td>$\rho_{LLL}$</td>
<td>$\bar{y}_{tv}$</td>
<td>95.81516</td>
</tr>
<tr>
<td>3</td>
<td>$\rho_{LLL}$</td>
<td>$\bar{y}_{tv}$</td>
<td>104.8436</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{uv}$</td>
<td>144.9099</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rah}$</td>
<td>146.3571</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rnh}$</td>
<td>134.8345</td>
</tr>
<tr>
<td>4</td>
<td>$\rho_{LLL}$</td>
<td>$\bar{y}_{tv}$</td>
<td>149.7035</td>
</tr>
</tbody>
</table>

Table 2. Mean square error of population I as $n \to \infty$

<table>
<thead>
<tr>
<th>S/No</th>
<th>Corr. Coeff.</th>
<th>Estimator</th>
<th>Sample Size (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>$\rho_{HLL}$</td>
<td>$\bar{y}_{uv}$</td>
<td>121.1768</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rah}$</td>
<td>323.3008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rnh}$</td>
<td>121.1768</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{tv}$</td>
<td>121.1768</td>
</tr>
<tr>
<td>2</td>
<td>$\rho_{HHH}$</td>
<td>$\bar{y}_{uv}$</td>
<td>83.9331</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rah}$</td>
<td>10765.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rnh}$</td>
<td>83.9331</td>
</tr>
<tr>
<td></td>
<td>$\rho_{LLL}$</td>
<td>$\bar{y}_{tv}$</td>
<td>83.9331</td>
</tr>
<tr>
<td>3</td>
<td>$\rho_{LLL}$</td>
<td>$\bar{y}_{tv}$</td>
<td>223.4875</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rah}$</td>
<td>258.637</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rnh}$</td>
<td>223.4875</td>
</tr>
<tr>
<td></td>
<td>$\rho_{LLL}$</td>
<td>$\bar{y}_{tv}$</td>
<td>223.4875</td>
</tr>
<tr>
<td>4</td>
<td>$\rho_{LLL}$</td>
<td>$\bar{y}_{tv}$</td>
<td>329.9688</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rah}$</td>
<td>344.059</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rnh}$</td>
<td>329.9688</td>
</tr>
</tbody>
</table>

Table 3. Coefficient of variation of population I as $n \to \infty$

<table>
<thead>
<tr>
<th>S/No</th>
<th>Corr. Coeff.</th>
<th>Estimator</th>
<th>Sample Size (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>$\rho_{HLL}$</td>
<td>$\bar{y}_{uv}$</td>
<td>746.7133</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rah}$</td>
<td>1171.607</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{rnh}$</td>
<td>567.0758</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\bar{y}_{tv}$</td>
<td>755.1692</td>
</tr>
<tr>
<td></td>
<td>$\rho_{LLL}$</td>
<td>$\bar{y}_{uv}$</td>
<td>624.087</td>
</tr>
</tbody>
</table>
Fig. 1. Graph showing Coefficient of Variation of Population I (a) $\rho_{HLL}$ (b) $\rho_{HHH}$ (c) $\rho_{LLL}$ (d) $\rho_{LLL}$
Table 4. Estimated mean of population II

<table>
<thead>
<tr>
<th>S/No</th>
<th>Corr. Coeff.</th>
<th>Estimator</th>
<th>Sample Size (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>ρ_HLL</td>
<td>y_{uv}</td>
<td>6.3383</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rah}</td>
<td>6.6081</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rnh}</td>
<td>6.5481</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{ly}</td>
<td>6.5898</td>
</tr>
<tr>
<td>2</td>
<td>ρ_HHH</td>
<td>y_{uv}</td>
<td>6.6134</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rah}</td>
<td>6.6110</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rnh}</td>
<td>6.6056</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{ly}</td>
<td>6.6111</td>
</tr>
<tr>
<td>3</td>
<td>ρ_LLL</td>
<td>y_{uv}</td>
<td>6.0836</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rah}</td>
<td>6.0651</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rnh}</td>
<td>6.0330</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{ly}</td>
<td>6.0673</td>
</tr>
<tr>
<td>4</td>
<td>ρ_LLL</td>
<td>y_{uv}</td>
<td>6.5918</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rah}</td>
<td>6.6082</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rnh}</td>
<td>6.5364</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{ly}</td>
<td>6.6101</td>
</tr>
</tbody>
</table>

Table 5. Mean square error of population II

<table>
<thead>
<tr>
<th>S/No</th>
<th>Corr. Coeff.</th>
<th>Estimator</th>
<th>Sample Size (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>ρ_HLL</td>
<td>y_{uv}</td>
<td>1.0242</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rah}</td>
<td>1.0629</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rnh}</td>
<td>0.9914</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{ly}</td>
<td>0.9851</td>
</tr>
<tr>
<td>2</td>
<td>ρ_HHH</td>
<td>y_{uv}</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rah}</td>
<td>0.1146</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rnh}</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{ly}</td>
<td>0.0031</td>
</tr>
<tr>
<td>3</td>
<td>ρ_LLL</td>
<td>y_{uv}</td>
<td>0.0254</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rah}</td>
<td>0.0295</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rnh}</td>
<td>0.0254</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{ly}</td>
<td>0.0254</td>
</tr>
<tr>
<td>4</td>
<td>ρ_LLL</td>
<td>y_{uv}</td>
<td>0.0259</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rah}</td>
<td>0.0262</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rnh}</td>
<td>0.0259</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{ly}</td>
<td>0.0259</td>
</tr>
</tbody>
</table>

Table 6. Coefficient of variation of population II

<table>
<thead>
<tr>
<th>S/No</th>
<th>Corr. Coeff.</th>
<th>Estimator</th>
<th>Sample Size (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>1</td>
<td>ρ_HLL</td>
<td>y_{uv}</td>
<td>1.0242</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rah}</td>
<td>1.0629</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rnh}</td>
<td>0.9914</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{ly}</td>
<td>0.9851</td>
</tr>
<tr>
<td>2</td>
<td>ρ_HHH</td>
<td>y_{uv}</td>
<td>0.8289</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rah}</td>
<td>5.1207</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{rnh}</td>
<td>0.8299</td>
</tr>
<tr>
<td></td>
<td></td>
<td>y_{ly}</td>
<td>0.8292</td>
</tr>
<tr>
<td>S/No</td>
<td>Corr. Coeff.</td>
<td>Estimator</td>
<td>Sample Size (n)</td>
</tr>
<tr>
<td>------</td>
<td>-------------</td>
<td>-----------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>$\hat{y}_{uv}$</td>
<td>2.6196</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{y}_{rah}$</td>
<td>2.8336</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{y}_{rnh}$</td>
<td>2.6415</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{y}_{ly}$</td>
<td>2.6266</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>$\hat{y}_{uv}$</td>
<td>2.4431</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{y}_{rah}$</td>
<td>2.4510</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{y}_{rnh}$</td>
<td>2.4638</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{y}_{ly}$</td>
<td>2.4363</td>
</tr>
</tbody>
</table>

Fig. 2. Graph showing coefficient of variation of population II (a) $\rho_{HLL}$ (b) $\rho_{HHH}$ (c) $\rho_{LLL}$ (d) $\rho_{LLL}$
6 Discussion of Results

In this study we have considered the performance of our proposed estimators and some selected estimators in literature on four partitions of the correlation coefficient interval \((-1, 1)\). (i) \(\rho_{HLL}\), (ii) \(\rho_{HHH}\) (iii) \(\rho_{LLL}\) and (iv) \(\rho_{LLL}\)

From Table 1, observe that as \(n \to \infty\), all the estimators pinned good estimates of the population mean in the same neighborhood at all levels of correlation coefficients. However, for small samples, \(\bar{y}_{RNH}\) exclusively overestimated the mean at \(\rho_{HLL}\) and underestimated the mean at \(\rho_{LLL}\).

In Table 2, the MSE of \(\bar{y}_{RAH}\) is consistently higher for all sample sizes at \(\rho_{HLL}\), \(\rho_{HHH}\) and \(\rho_{LLL}\), but fairly stable when \(n = 50\) at \(\rho_{LLL}\). Observe also that the MSEs of \(\bar{y}_{uv}, \bar{y}_{tv}\), and \(\bar{y}_{RNH}\) in all the sampling regions decrease as \(n \to \infty\), and are the same for all sample sizes. This family of estimators are generally preferable for large scale surveys.

From Table 3, in the parameter subspace denoted by \(\rho_{HLL}, \bar{y}_{RNH}\) performed better on the CV scale than the others estimator for small sample sizes. This is followed by the proposed estimators \(\bar{y}_{uv}, \bar{y}_{tv},\) and \(\bar{y}_{RAH}\). As sample size increases, the proposed estimators \(\bar{y}_{uv}\) and \(\bar{y}_{tv}\) became more efficient than the other competing estimators. Under this particular interval, we can confidently say that, \(\bar{y}_{RNH}\) estimator performs better when sample size is very small. Whereas \(\bar{y}_{uv}\) is preferable when sample size is moderate, while, for large samples, \(\bar{y}_{tv}\) is preferred as shown in Fig. 1(a). In case of the real data presented in Table 6 and depicted by Fig. 2(a), the proposed estimators are more efficient and precise than the other estimators on the entire parameter space.

Considering the second parameter subspace, \(\rho_{HHH}, \bar{y}_{tv}\) appears to dominate throughout the parameter space except for moderate sample sizes for the simulation study. This is closely followed by \(\bar{y}_{uv}\) (see Fig. 1(b)). On the other hand, the proposed estimators are more efficient for small samples (Table 6 and Fig. 2b) but one of the competing estimators appeared more efficient as sample size increases for the real data set.

In the third region, \(\rho_{LLL}\), where all the correlation coefficients are low but positive, it is observed that, \(\bar{y}_{tv}\) and \(\bar{y}_{uv}\) in a closely dominate the for small sample size of between 10 and 25. Between \(n=25\) and \(n=50\), \(\bar{y}_{RNH}\) and \(\bar{y}_{RAH}\) had better performance while for \(n = 50\), the duo, \(\bar{y}_{uv}\) and \(\bar{y}_{tv}\) have dominance over \(\bar{y}_{RAH}\) and \(\bar{y}_{RNH}\) (see Fig. 1(c)). Considering the real data, the proposed estimators are more efficient than other estimators for small samples (see Table 6 and Fig. 2c).

The fourth experiment considered the region, \(\rho_{LLL}\). For the simulation study, one of the proposed estimators, \(\bar{y}_{tv}\) dominated all other estimators throughout the parameter domain. This is followed by \(\bar{y}_{uv}\) and then \(\bar{y}_{RAH}\) (see Table 3 and Fig. 1d). In case of real data, The proposed are more efficient for small samples as shown in Table 6 and depicted in Fig. 2d. Both \(\bar{y}_{tv}\) and \(\bar{y}_{RAH}\) utilize two auxiliary variables that are transformed according to Srivenkataramana (1980) yet \(\bar{y}_{RAH}\) performed better than \(\bar{y}_{RAH}\) for all the experiments conducted in this study. Reason is not far-fetched, as \(\bar{y}_{tv}\) belongs to the regression family whereas, \(\bar{y}_{RAH}\) is a member of the ratio family. Thus, \(\bar{y}_{RAH}\) would dominate \(\bar{y}_{tv}\) only if the regression line between \(y\) and \(x\) or \(z\) passes through the origin [10]. The transformation of the auxiliary variables has also shown improvement on efficiency of \(\bar{y}_{tv}\) as compared to that of \(\bar{y}_{uv}\).

7 Conclusion

This study proposed an alternative hybrid exponential type estimator of finite population mean in simple random sampling under two cases:

- When the auxiliary variables are not transformed
- When the auxiliary variables are transformed

The study also considered four partitions of the correlation coefficient parameter space for which:

- Only one correlation coefficient is high, and all three are positive, \(\rho_{HLL}\).
- All correlation coefficients are high and positive, $\rho_{HHH}$.
- All correlation coefficients are low and positive, $\rho_{LLL}$.
- All correlation coefficients are low and negative, $\rho_{LL}$.

Coefficient of Variation (CV) of different estimators for different sample sizes as given in Tables 3 and 6 respectively, for the synthetic and real data, have shown that the proposed estimators, $\tilde{y}_{HP}$ and $\tilde{y}_{LP}$, are:

- More efficient than the existing estimators for both small and large samples when at least one of the correlation coefficient is high.
- More efficient than the existing estimators for only small samples when the correlation coefficients are low and positive or negative.

The proposed estimators are therefore recommended for use in simple random sampling for both small- and large-scale surveys.

**Competing Interests**

Authors have declared that no competing interests exist.

**References**


Appendix

Appendix 1. Proof to Theorem 1

Consider the expression in (11) above,

$$\bar{y}_{uv} = [\bar{y} + \beta (\bar{Z} - \bar{x})] \exp \left[ \alpha \left( \frac{\bar{X} - \bar{x}}{\bar{Y} + \bar{x}} \right) \right]$$

By substituting the definitions for $\bar{x}$, $\bar{Y}$ and $\bar{Z}$ from section 1, we have:

$$= \left[ (1 + e_y)\bar{y} + \beta (\bar{Z} - (1 + e_x)\bar{Z}) \right] \exp \left[ \alpha \left( \frac{\bar{X} - (1 + e_x)\bar{X}}{\bar{Y} + (1 + e_x)\bar{Z}} \right) \right]$$

$$= \left[ \bar{Y} + \bar{Y} e_y + \beta (\bar{Z} - \bar{Z} - \bar{Z} e_x) \right] \exp \left[ \alpha \left( \frac{-\bar{X} e_x}{\bar{Y} + \bar{X} + \bar{Z} e_x} \right) \right]$$

$$= \left[ \bar{Y} + \bar{Y} e_y + \beta (-\bar{Z} e_x) \right] \exp \left[ \alpha \left( \frac{-\bar{X} e_x}{\bar{Y} + \bar{Z} e_x} \right) \right]$$

$$= \left[ \bar{Y} + \bar{Y} e_y + \beta (-\bar{Z} e_x) \right] \exp \left[ \alpha \left( \frac{-\bar{X} e_x}{\bar{Y} + \bar{Z} e_x} \right) \right]$$

$$= \left[ \bar{Y} + \bar{Y} e_y + \beta (-\bar{Z} e_x) \right] \exp \left[ \frac{-\bar{X} e_x}{\bar{Y} + \bar{Z} e_x} \right]$$

By First order approximation principle, we have:

$$= \left[ \bar{Y} + \bar{Y} e_y + \beta (-\bar{Z} e_x) \right] \exp \left[ -\frac{1}{2} \bar{e}_x \right] \tag{17}$$

Expanding the exponential part expression in (17), we have:

$$= \left[ \bar{Y} + \bar{Y} e_y + \beta (-\bar{Z} e_x) \right] \exp \left[ \frac{-1}{2} \bar{e}_x + \frac{1}{8} \bar{e}^2 e_x + \cdots \right]$$

$$= \left[ \bar{Y} + \bar{Y} e_y + \beta (-\bar{Z} e_x) \right] \left[ 1 - \frac{1}{2} \bar{e}_x + \frac{1}{8} \bar{e}^2 e_x + \cdots \right]$$

$$= \left[ \bar{Y} + \bar{Y} e_y + \beta (-\bar{Z} e_x) \right] \left[ \bar{Y} \left[ 1 - \frac{1}{2} \bar{e}_x + \frac{1}{8} \bar{e}^2 e_x \right] - \beta \bar{e}_x \left[ 1 - \frac{1}{2} \bar{e}_x + \frac{1}{8} \bar{e}^2 e_x \right] \right]$$

Taking expectation on both sides of (18),
Applying the definitions of Expectations in section 1 to (19), we have

\[
\begin{align*}
\bar{\gamma}E &= \left[ -\frac{1}{2} \alpha e_x + \frac{1}{8} \alpha^2 e_x^2 + e_y - \frac{1}{2} \alpha e_x e_y - \beta Z \left( e_x - \frac{1}{2} \alpha e_x e_y \right) \right] \\
&= \bar{\gamma} \left[ 0 + \frac{1}{8} \alpha^2 \theta C_x^2 + 0 - \frac{1}{2} \alpha \theta \rho_{yx} C_x C_y \right] - \beta \bar{Z} \left[ 0 - \frac{1}{2} \alpha \theta \rho_{zx} C_x C_z \right] \\
&= \bar{\gamma} \theta \left( \frac{1}{8} \alpha^2 C_x^2 - \frac{1}{2} \alpha \rho_{yx} C_x C_y \right) + \frac{1}{2} \beta \bar{Z} \theta \alpha \rho_{zx} C_x C_z
\end{align*}
\]

Therefore,

\[
\text{Bias}(\bar{\gamma}_{uv}) = \bar{\gamma} \theta \left( \frac{1}{8} \alpha^2 C_x^2 - \frac{1}{2} \alpha \rho_{yx} C_x C_y \right) + \frac{1}{2} \beta \bar{Z} \theta \alpha \rho_{zx} C_x C_z
\]

**Appendix 2. Proof to Theorem 2**

The \( \text{MSE}(\bar{\gamma}_{uv}) = E(\bar{\gamma}_{t1} - \bar{\gamma})^2 \), therefore, squaring both sides of (19), we have:

\[
(\bar{\gamma}_{uv} - \bar{\gamma})^2 = \left[ \bar{\gamma} \left( -\frac{1}{2} \alpha e_x + \frac{1}{8} \alpha^2 e_x^2 + e_y - \frac{1}{2} \alpha e_x e_y \right) - \beta \bar{Z} \left( e_x - \frac{1}{2} \alpha e_x e_y \right) \right]^2
\]

\[
= \bar{\gamma}^2 \left( e_y - \frac{1}{2} \alpha e_x \right)^2 - 2 \beta \bar{Z} e_y \left( e_y - \frac{1}{2} \alpha e_x \right) + \beta^2 \bar{Z}^2 e_y^2
\]

\[
= \bar{\gamma}^2 \left( e_y^2 - 2 \frac{1}{2} \alpha e_x e_y + \frac{1}{4} \alpha^2 e_x^2 \right) - 2 \beta \bar{Z} \left( e_y e_x - \frac{1}{2} \alpha e_x e_y \right) + \beta^2 \bar{Z}^2 e_y^2
\]

\[
E(\bar{\gamma}_{uv} - \bar{\gamma})^2 = \bar{\gamma}^2 \left[ E(e_y^2) - 2 \alpha E(e_y e_x) + \frac{1}{4} \alpha^2 E(e_x^2) \right] - 2 \beta \bar{Z} \left[ E(e_y e_x) - \frac{1}{2} \alpha E(e_x e_y) \right] + \beta^2 \bar{Z}^2 E(e_y^2)
\]

\[
\text{MSE}(\bar{\gamma}_{uv}) = E(\bar{\gamma}_{uv} - \bar{\gamma})^2
\]

\[
= \bar{\gamma}^2 \left[ \theta C_y^2 - \alpha \theta \rho_{yx} C_x C_y + \frac{1}{4} \alpha^2 \theta C_x^2 \right] - 2 \beta \bar{Z} \left[ \theta \rho_{yx} C_y C_x - \frac{1}{2} \alpha \theta \rho_{zx} C_x C_z \right] + \beta^2 \bar{Z}^2 \theta C_x^2
\]

To obtain the optimal value of \( \alpha \) that minimizes the \( \text{MSE}(\bar{\gamma}_{uv}) \), we differentiate (20) with respect to \( \alpha, \beta \) and equate to zero.

\[
\frac{\partial \text{MSE}(\bar{\gamma}_{uv})}{\partial \alpha} = \frac{\partial}{\partial \alpha} \left[ \bar{\gamma}^2 \left[ \theta C_y^2 - \alpha \theta \rho_{yx} C_x C_y + \frac{1}{4} \alpha^2 \theta C_x^2 \right] - 2 \beta \bar{Z} \left[ \theta \rho_{yx} C_y C_x - \frac{1}{2} \alpha \theta \rho_{zx} C_x C_z \right] + \beta^2 \bar{Z}^2 \theta C_x^2 \right] = 0
\]

\[
\Rightarrow \bar{\gamma}^2 \left( \frac{1}{2} \alpha \theta C_x^2 - \theta \rho_{yx} C_x C_y \right) + \beta \bar{Z} \theta \rho_{yx} C_x C_z = 0
\]

\[
\bar{\gamma}^2 \left( \frac{1}{2} \alpha \theta C_x^2 - \theta \rho_{yx} C_x C_y \right) - \beta \bar{Z} \theta \rho_{yx} C_x C_z
\]

\[
\alpha = \frac{\theta \rho_{yx} C_y - \beta \bar{Z} \theta \rho_{yx} C_z}{\bar{\gamma} \theta C_x}
\]

\[
\alpha = \frac{2(\bar{\gamma} \theta \rho_{yx} C_y - \beta \bar{Z} \theta \rho_{yx} C_z)}{\bar{\gamma} \theta C_x}
\]

(21)
Putting (21) into (22a), we have:

\[
\beta = \frac{-\bar{Y}\rho_{yx}\rho_{xz}C_y - \beta\bar{Z}\rho_{xz}C_x - \bar{Y}\rho_{yz}C_y}{ZC_x} = \bar{Y}\rho_{yx}\rho_{xz}C_y - \beta\bar{Z}\rho_{xz}C_x - \bar{Y}\rho_{yz}C_y = \frac{-\bar{Y}(\rho_{yx}\rho_{xz}C_y - \rho_{yz}C_y)}{Z\rho_{xz}C_x - ZC_x}
\]  

(22b)

Substituting the value of (21) and (22b) into Equation (20), gives on simplification using Maple 18, we have:

\[
MSE(\bar{Y}_{uv}) = \frac{\bar{Y}^2C_x^2\theta}{\rho_{xx}^2 - 1}\left(\rho_{\bar{y},x} + \rho_{\bar{z},x} + \rho_{\bar{y},z} - 2\rho_{\bar{y},x}\rho_{\bar{z},x}\rho_{\bar{y},z} - 1\right);
\rho_{xx} \neq \pm 1; \rho_{xy}, \rho_{yx} \neq 1
\]

(23)

We now, substitute the values of \(\alpha\) and \(\beta\) from (21) and (22b) respectively, into (20) to obtain bias of \(\bar{y}_{uv}\) as:

\[
Bias(\bar{y}_{uv}) = \bar{Y}\theta\left(\frac{1}{6}a^2C_x^2 + \frac{1}{2}\alpha\rho_{yx}C_x C_y\right) + \frac{1}{2}\bar{Y}\bar{Z}\alpha\rho_{xz}C_x C_z
\]

\[
= -\frac{1}{2}\frac{C_x^2\theta}{(\rho_{xx}^2 - 1)^2}\left(\rho_{xx}\rho_{yz} - \rho_{yx}\right)^2 ; \rho_{xx} \neq 1
\]

(24)

**Appendix 3. Proof to Theorems 3**

By substituting the definitions for \(\bar{y}, \bar{x}\) and \(\bar{z}\) in section 1 and 2 respectively, into (14), we have:

\[
\bar{y}_{uv} = [(1 + e_y)\bar{Y} + \beta_1((1 - ge_x)\bar{Z} - \bar{Z})] * \exp\left[\alpha_1\left(\frac{(1 - ge_x)\bar{x} - \bar{x}}{\bar{x} + (1 - ge_x)\bar{x}}\right)\right]
\]

\[
= [\bar{Y} + \bar{Y}e_y - \beta_1 ge_x \bar{Z}] \exp\left[\alpha_1\left(\frac{-ge_x\bar{x}}{Z\bar{x} - ge_x\bar{x}}\right)\right]
\]

\[
= [\bar{Y} + \bar{Y}e_y - \beta_1 ge_x \bar{Z}] \exp\left[\alpha_1\left(\frac{-ge_x\bar{x}}{Z - ge_x}\right)\right]
\]

\[
= [\bar{Y} + \bar{Y}e_y - \beta_1 ge_x \bar{Z}] \exp\left[\frac{1}{2}\alpha_1 ge_x \left(1 + \frac{1}{2}ge_x\right)^{-1}\right]
\]

\[
= [\bar{Y} + \bar{Y}e_y - \beta_1 ge_x \bar{Z}] \exp\left[-\frac{1}{2}\alpha_1 ge_x \left(1 - \frac{1}{2}ge_x\right) + \frac{1}{4}g^2e_x^2 - \frac{1}{8}g^3e_x^3 + \cdots\right]
\]

By first order approximation,

\[
= [\bar{Y} + \bar{Y}e_y - \beta_1 ge_x \bar{Z}] \exp\left[-\frac{1}{2}\alpha_1 ge_x\right]
\]

(25)

Expanding the exponent in the expression (25) we have:
Taking expectation on both sides of (26) and subsequent application of the definitions in section 1 gives us the bias.

\[
\text{Bias}(\bar{y}_{tp}) = E(\bar{y}_{tp} - \bar{y}) = \bar{y} \left[ -\frac{1}{2} \alpha_1 g e_x + e_y - \frac{1}{2} \alpha_1 g e_x e_y + \frac{1}{8} \alpha_1^2 g^2 e_x^2 \right] - \beta_1 \bar{Z} \left( g e_x - \frac{1}{2} \alpha_1 g^2 e_x e_y \right)
\]

\[
\bar{y}_{tp} - \bar{y} = \left[ -\frac{1}{2} \alpha_1 g \rho_{yx} C_x C_y + \frac{1}{8} \alpha_1^2 g^2 \theta C_x^2 \right] + \beta_1 \bar{Z} \left[ \frac{1}{4} \alpha_1 g^2 \rho_{yx} C_x C_y \right]
\]

Therefore,

\[
\text{Bias}(\bar{y}_{tp}) = \bar{y} \left[ \frac{1}{8} \alpha_1^2 g^2 \theta C_x^2 - \frac{1}{2} \alpha_1 g \rho_{yx} C_x C_y \right] + \beta_1 \bar{Z} \left[ \frac{1}{4} \alpha_1 g^2 \rho_{yx} C_x C_y \right]
\]

**Appendix 4. Proof to Theorem 4**

In order to obtain the MSE of \( \bar{y}_{tp} \), we square both sides of Equation (26) and take expectations.

\[
(\bar{y}_{tp} - \bar{y})^2 = \bar{y}^2 \left[ -\frac{1}{2} \alpha_1 g e_x + e_y - \frac{1}{2} \alpha_1 g e_x e_y + \frac{1}{8} \alpha_1^2 g^2 e_x^2 \right] - 2 \beta_1 \bar{Z} \left( g e_x - \frac{1}{2} \alpha_1 g^2 e_x e_y \right)^2
\]

\[
= \left[ \bar{y} \left( -\frac{1}{2} \alpha_1 g e_x + e_y \right) - \beta \bar{Z} \left( g e_x \right) \right]^2
\]

\[
= \bar{y}^2 \left[ -\frac{1}{2} \alpha_1 g e_x + e_y \right]^2 - 2 \beta_1 \bar{Y} \bar{Z} g e_x \left[ -\frac{1}{2} \alpha_1 g e_x + e_y \right] + \beta^2 \bar{Z}^2 g^2 e_x^2
\]

Taking expectations on both sides of (27) and substituting the definition of section 1, we have:

\[
E(\bar{y}_{tp} - \bar{y})^2 = \bar{Y}^2 \left( \frac{1}{4} \alpha_1^2 g^2 E(e_x^2) - \alpha_1 g E(e_x e_y) + E(e_y^2) \right) - 2 \beta \bar{Y} \bar{Z} \left[ -\frac{1}{2} \alpha_1 g^2 E(e_x e_y) + g E(e_y e_x) \right] + \beta^2 \bar{Z}^2 g^2 E(e_x^2)
\]

296
\[ \text{Therefore,} \]
\[ \text{MSE}(\tilde{y}_{tv}) = E(\tilde{y}_{tv} - \tilde{y})^2 = \tilde{y}^2 \left( \frac{1}{4} \alpha_1^2 g^2 \theta C_x^2 - \alpha_1 \theta \rho_{yx} C_x C_y + \theta C_y^2 \right) - 2 \beta_1 \tilde{y} \left( g \theta \rho_{yx} C_y C_z - \frac{1}{2} \alpha_1 g^2 \theta \rho_{xx} C_x C_z \right) + \beta_1^2 \tilde{y}^2 g^2 \theta C_z^2 \]  

(28)

To obtain the value of \( \alpha_1, \beta_1 \) that minimizes the MSE, we differentiate Equation (28) partially with respect to \( \alpha_1, \beta_1 \) and equate to zero:

\[ \frac{\partial \text{MSE}(\tilde{y}_{tv})}{\partial \alpha_1} = \frac{\partial}{\partial \alpha_1} \left( \tilde{y}^2 \left( \frac{1}{4} \alpha_1^2 g^2 \theta C_x^2 - \alpha_1 \theta \rho_{yx} C_x C_y + \theta C_y^2 \right) - 2 \beta_1 \tilde{y} \left( g \theta \rho_{yx} C_y C_z - \frac{1}{2} \alpha_1 g^2 \theta \rho_{xx} C_x C_z \right) + \beta_1^2 \tilde{y}^2 g^2 \theta C_z^2 \right) = 0 \]

\[ \tilde{y}^2 \left( \frac{1}{4} \alpha_1 g^2 \theta C_x^2 - \alpha_1 \theta \rho_{yx} C_x C_y + \theta C_y^2 \right) + 2 \beta_1 \tilde{y} \left( g \theta \rho_{yx} C_y C_z - \frac{1}{2} \alpha_1 g^2 \theta \rho_{xx} C_x C_z \right) + \beta_1^2 \tilde{y}^2 g^2 \theta C_z^2 = 0 \]

(29a)

Solving for \( \alpha_1 \) again we have

\[ \alpha_1 = \frac{2(\tilde{y} \theta \rho_{yx} C_y - \beta_1 \tilde{y} g \rho_{xx} C_x)}{\tilde{y} g C_x} \]  

(29b)

Equating 29a) to (29b) and solving for \( \beta_1 \) we have

\[ \beta_1 = \frac{\tilde{y} C_x (\rho_{yx} \rho_{xx} - \rho_{yx})}{Z g C_x (\rho_{xx}^2 - 1)} \]  

(30)

Substituting (27) into (26a) and simplifying we have

\[ \alpha_1 = \frac{2 C_y (\rho_{yx} \rho_{xx} - \rho_{yx})}{g C_x (\rho_{xx}^2 - 1)} \]  

(31)

We finally substitute the expressions for \( \alpha_1, \beta_1 \) in (31) and (30) above into (28) and simplify using maple to obtain the \( \text{MSE}(\tilde{y}_{tv}) \) as:

\[ \text{MSE}(\tilde{y}_{tv}) = \frac{\tilde{y}^2 \theta C_x^2 (\rho_{yx}^2 + \rho_{xx}^2 + \rho_{yx}^2 - 2 \rho_{yx} \rho_{xx} \rho_{yx} - 1)}{\rho_{xx}^2 - 1} \]
Therefore,

\[
MSE(\hat{y}_{tv}) = \bar{Y}^2 \theta C_y^2 \left( \rho_{zz}^2 - 1 \right)^{-1} \left( \rho_{yz}^2 + \rho_{zx}^2 + \rho_{yx}^2 - 2 \rho_{yx} \rho_{zz} \rho_{yz} - 1 \right)
\]

where, \( \rho_{yx}, \rho_{yz}, \rho_{zx} \neq \pm 1 \)

Recall from (15) that:

\[
Bias(\hat{y}_{tv}) = \bar{Y} \left[ \frac{1}{8} \alpha_1 g^2 \theta C_y^2 \left( \rho_{zz}^2 - 1 \right) \right] + \beta_1 \bar{Z} \left[ \frac{1}{2} \alpha_1 g^2 \theta \rho_{zz} C_y C_z \right]
\]

Substituting (31) and (30) into the above expression gives on simplification (using Maple),

\[
Bias(\hat{y}_{tv}) = \frac{1}{2} \frac{\bar{Y} C_y^2 \theta}{(\rho_{zz}^2 - 1)^2} \left( \rho_{zz} \rho_{yz} - \rho_{yx} \right)^2 ; \rho_{xz} \neq 1
\]

© 2021 Uba et al.; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

---

**Peer-review history:**

The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)

https://www.sdiarticle5.com/review-history/79568