Revised Research Article

AN ALTERNATIVE HYBRID ESTIMATOR OF FINITE POPULATION MEAN IN SIMPLE RANDOM SAMPLING

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*. (2015) 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,

Key words: 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 (1991) 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.* (2015), 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 (2003, 2005), 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 (2004, and 2006) 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 (2013) 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 (1981) 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 (Singh and Espejo, 2007; 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.* (2015) 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, $\bar{y}$, defined 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 $\bar{y}$ and $N\bar{y}$ are consistent estimates of the population mean and the mean total, respectively. Searls (1964) proposed a modified version of the sample mean estimator, $\bar{y}$, where k 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 (Pradhan, 2005). 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. (*2014) 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 (1937) made use of the regression method of estimation to estimate the average area of the leaves on a plant. Cochran (1940, 1942) 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 (1957) 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 (1993). These improved ratio estimators, though, biased, are more efficient than the classical ratio estimator.

Rashid *et al.* (2015) 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.* (2014), and Jhajj and Lata (2014) 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 (2013) estimator. This new estimator was validated by using ten different real datasets. Similar exponential ratio type estimators were developed by Singh and Vishwakarma (2007), Vishwakarma and Kumar (2014), and, Singh and Khahid (2015) with application in two phase sampling.

On the other hand, Kumar *et al.* (2017) 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.* (2013), 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 (2015) 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.* (2014) 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.* (2016) 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.* (2014) 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 (2017) 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 (2018) suggested a more generalized square root transformed ratio type estimator and exponential ratio type estimator similar to Gupta *et al.* (2017) estimator except that it combined two ratio estimators in the linear combination and two ratio estimator in the exponential component. Meanwhile, Riaz *et al.* (2014) had constructed regression-cum-ratio/product exponential type estimator by combining the concept of Bahl and Tujeja (1991), and the regression estimator.

On the other hand, many authors have used median, coefficient of kurtosis, coefficient of skewness, deciles, quartiles deviation, etc. (Yadav *et al.*, 2017, Subramani, 2013., Subramani and Kumarapandiyan, 2012., Singh and Espejo, 2003., Sosidia and Dwivedi, 1981., Yan and Tian, 2010., Jeelani *et al.* 2013., Shittu and Adepoju 2014., Subzar *et al.* 2017., Raja *et al.* 2017., Irfan *et al.* 2018).

Abid *et al.* (2016) 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.* (2013), Misra *et al.* (2017) 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 (2004), Subzar *et al.* (2017) developed a new class of more efficient ratio type 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.* (2013) 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 (2007) 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.* (2020) 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.* (2021) 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 (2021) 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 (2021) 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.* (2021) provided an improved version of the Bahl and Tuteja (19991) ratio estimator, Etebong *et al.* (2021) 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*. (2015) 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 (1980) to ascertain whether or not the efficiency of the proposed estimator is improved.

3. Preliminaries and notations

Consider a finite population, $U = \{U_1, U_2, \dots , U_N\}$. 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; $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$: population variance of study variable;
$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$: population variance of x variable; $S_z^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Z_i - \bar{Z})^2$: 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 ρ_1 ; $\rho_{yx} = \frac{S_{yx}}{S_y S_x}$: correlation coefficient between y and x denoted by ρ_2; $\rho_{xz} = \frac{S_{xz}}{S_x S_z}$: correlation coefficient between x and z denoted by ρ_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 $E(e_i) = 0 \; for \; (i = x, y, z)$;$E(e_y^2) = \theta C_y^2$; $E(e_x^2) = \theta C_x^2$; $E(e_z^2) = \theta C_z^2$; $E(e_x e_y) = \theta \rho_{yx} C_x C_y$;
$E(e_x e_z) = \theta \rho_{xz} C_x C_z$; $E(e_y e_z) = \theta \rho_{yz} C_y C_z$ where, $\theta = \frac{1}{n} - \frac{1}{N}$

4. Existing Ratio-Exponential Estimators in Simple Random Sampling

(i) Classical regression estimator

Cochran (1942) 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:

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

The MSE of the classical regression estimator is given by:

$$MSE(\bar{y}_{lr}) = \bar{Y}^2 C_y^2 \theta\left(1 - \rho_{y,x}^2\right) \tag{2}$$

(ii) Exponential ratio-cum-ratio estimator

Rashid *et al.* (2015) used two transformed auxiliary variables to develop this estimator under single phase sampling which is an improvement on Bahl and Tuteja (1991) estimator. It is given as:

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

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

$$Bias(\bar{y}_{RAH}) = \frac{\bar{Y}\theta}{8}\{g^2[C_z^2(2abK_{xz} + b^2) + a^2C_x^2] - 4g(aK_{yx}C_x^2 + bK_{yz}C_z^2)\} \quad (4)$$

$$MSE(\bar{y}_{RAH}) = \bar{Y}^2\theta C_y^2 \frac{[1 - (\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{xz}\rho_{yz})]}{(1 - \rho_{xz}^2)} \quad (5)$$

where, $K_{y,x} = \rho_{y,x}\frac{C_y}{C_x}$; $\theta = \frac{1}{n} - \frac{1}{N} = \frac{N-n}{Nn}$.

(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:

$$\bar{y}_{EE} = \theta_1\bar{y} + \theta_2(\bar{X} - \bar{x})\exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right) \quad (5)$$

where: θ_1 and θ_2 are suitably chosen scalars, such that $\theta_1 > 0$ and $-\infty < \theta_2 < \infty$. Its MSE and bias are given as:

$$B(\bar{y}_{EE}) = \bar{Y}\left[(\theta_1 - 1) + \theta_2 K_{y,x}\theta\frac{C_x^2}{2}\right] \quad (6)$$

$$MSE(\bar{y}_{EE}) = \bar{Y}^2[1 + \theta_1^2\gamma_1 - 2\theta_1 - 2\theta_1\theta_2 m\gamma_2 - 2\theta_2 m\gamma_3 + \theta_2^2 m^2\gamma_4] \quad (7)$$

where: $\gamma_1 = 1 + \theta C_y^2$; $\gamma_2 = C_x^2\theta\left(K_{yx} - \frac{1}{2}\right)$; $\gamma_3 = \theta\frac{C_x^2}{2}$; $\gamma_4 = \theta C_x^2$; $m = \frac{\bar{X}}{\bar{Y}}$; $\theta_1 = \frac{\gamma_4 + \gamma_2\gamma_3}{\gamma_1\gamma_4 - \gamma_2^2}$; $\theta_2 = \frac{R(\gamma_2 + \gamma_1\gamma_3)}{\gamma_1\gamma_4 - \gamma_2^2}$; $R = \frac{\bar{Y}}{\bar{X}}$

(iv) **Exponential regression-ratio/product estimator**

Riaz *et al.* (2014) developed the regression-ratio/product exponential estimator by combining the concept of Bahl and Tuteja (1991) exponential type estimator and the classical regression estimator. It is given as:

$$\bar{y}_{RNH} = [\bar{y} + a_1(\bar{Z} - \bar{z})]exp\left[\gamma\frac{\bar{X} - \bar{x}}{\bar{X} + (b_1 - 1)\bar{x}}\right] \quad (8)$$

where a_1, b_1 are real positive constants and γ may take the values -1 and 1.

$$Bias(\bar{y}_{RNH}) = a_1\gamma\theta\bar{Z}\rho_{x,z}\frac{C_xC_z}{b_1} + \frac{\theta\gamma\bar{Y}C_x^2}{2b_1^2}[1 + 2(b_1 - 1) - 2b_1K_{yx}] \quad (9)$$

where:

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

$$MSE(\bar{y}_{RNH}) = \frac{\bar{Y}^2 C_y^2 \theta}{\left(1-\rho_{x,z}^2\right)}\left[1-\gamma^2\rho_{x,z}^2-\rho_{y,z}^2-\gamma^2\rho_{y,x}^2+2\gamma^2\rho_{y,z}\rho_{x,z}\rho_{y,x}\right. \quad (10)$$

5. Proposed Estimator

The proposed alternative hybrid regression-cum-ratio exponential type estimators of finite population mean modifies Rashid *et al.* (2015) 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:

$$\bar{y}_{uv} = [\bar{y}+\beta(\bar{Z}-\bar{z})]exp\left[\alpha\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right)\right] \quad (11)$$

where, α and β suitably chosen constants such that $MSE(\bar{y}_{uv})$ is minimized.

Theorem 1:

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

$$Bias(\bar{y}_{uv}) = \bar{Y}\theta\left(\frac{1}{8}\alpha^2 C_x^2 - \frac{1}{2}\alpha\rho_{y,x}C_x C_y\right)+\frac{1}{2}\beta\bar{Z}\alpha\theta\rho_{x,z}C_x C_z \quad (12)$$

Theorem 2:

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

$$MSE(\bar{y}_{uv}) = \bar{Y}^2\theta C_y^2\,(\rho_{xz}^2-1)^{-1}\left(\rho_{yz}^2+\rho_{xz}^2+\rho_{yx}^2-2\rho_{yx}\rho_{xz}\rho_{yz}-1\right) \quad (13)$$

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

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

$$\bar{y}_{tv} = [\bar{y}+\beta_1(\bar{z}^*-\bar{Z})]\exp\left[\alpha_1\left(\frac{\bar{x}^*-\bar{X}}{\bar{X}+\bar{x}^*}\right)\right] \quad (14)$$

where, α_1 and β_1 are suitably chosen constants that minimize MSE of $\bar{y}_{tv}$. $\bar{x}^*$ and $\bar{z}^*$ are transformed auxiliary variables such that $\bar{x}^* = (1-ge_x)\bar{X}$ and $\bar{z}^* = (1-ge_z)\bar{Z}$ and $g = \frac{n}{N-n}$;

Theorem 3:

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

$$Bias(\bar{y}_{tv}) = \bar{Y}\left[\frac{1}{8}\alpha_1^2 g^2\theta C_x^2 - \frac{1}{2}\alpha_1 g\theta\rho_{yx}C_x C_y\right]+\beta_1\bar{Z}\left[\frac{1}{2}\alpha_1 g^2\theta\rho_{xz}C_x C_z\right] \quad (15)$$

Theorem 4:

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

$$MSE(\bar{y}_{tv}) = \bar{Y}^2\theta C_y^2\,(\rho_{xz}^2-1)^{-1}\left(\rho_{yz}^2+\rho_{xz}^2+\rho_{yx}^2-2\rho_{yx}\rho_{xz}\rho_{yz}-1\right) \quad (16)$$

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, $Bias(\bar{y}_{uv})$ is -1 times the bias of the alternative hybrid estimator when the auxiliary variables are transformed, $Bias(\bar{y}_{tv})$. i.e. $Bias(\bar{y}_{uv}) = -Bias(\bar{y}_{tv})$

Corollary 2:

The MSE of $\bar{y}_{tv}$ 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. $MSE(\bar{y}_{uv}) = MSE(\bar{y}_{tv})$

6. Efficiency Comparison of Estimators

This study employs the Coefficient of Variation (CV) to compare performance of estimators considered in this study. Bowerman (2001) 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:

$$CV = \frac{\sqrt{Var(X)}}{\bar{X}} \times 100\%$$

For estimators that are biased, the Coefficient of Variation is given as:

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

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]

$N = 1000$; $\bar{Y} = 100.0786$; $n_1 = 10, n_2 = 25$, $n_3 = 50, n_4 = 100$; $\bar{X} = 25.90251$;

$\bar{x}_1 = 25.42673$; $\bar{x}_2 = 26.29295$; $\bar{x}_3 = 24.82662$; $\bar{x}_4 = 27.15513$; $\bar{Z} = 75.27509$;

$\bar{z}_1 = 59.3777$; $\bar{z}_2 = 82.72477$; $\bar{z}_3 = 70.25747$; $\bar{z}_4 = 89.85487$; $C_y = 0.71745$;

$C_x = 0.55689$; $C_z = 0.56869$; $\rho_{yx} = -0.0158$; $\rho_{yz} = 0.009504$; $\rho_{xz} = -0.0036$

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:

$N = 40;\ \bar{Y} = 6.1532;\ n_1 = 10, n_2 = 25,\ n_3 = 35, \bar{X} = 4.4659; \bar{Z} = 4.3583; \bar{x}_1 = 4.3847$

$;\ \bar{x}_2 = 4.6830;\ \bar{x}_3 = 4.6589;\ \bar{z}_1 = 4.5185; \bar{z}_2 = 4.296; \bar{z}_3 = 4.5191;\ C_y = 0.7977;$
$C_x = 0.8531; C_z = 0.9601; \rho_{yx} = -0.02495;\ \rho_{yz} = -0.01002; \rho_{xz} = 0.160939$

We denote the four correlation coefficient partitions by the following:
(i) ρ_{HLL} is the region, ($0.7 < \rho_1 < 1$ and $0 < \rho_2, \rho_3 < 0.5$)
(ii) ρ_{HHH} is the region, ($0.7 < \rho_1, \rho_2, \rho_3 < 1$)
(iii) ρ_{LLL} is the region, ($0 < \rho_1, \rho_2, \rho_3 < 0.5$)
(iv) ρ_{LLL}^{-} is the region ($-0.5 < \rho_1, \rho_2, \rho_3 < 0$)
where, $\rho_1 = \rho_{yz}$, $\rho_2 = \rho_{yx}$ 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 $n \to \infty$

S/No	$Corr.Coeff.$	$Estimator$	Sample Size (n) 10	25	50	100
1	ρ_{HLL}	$\bar{y}_{uv}$	147.4198	120.4119	122.4200	116.8018
		$\bar{y}_{rah}$	153.4693	117.4461	121.6723	120.2516
		$\bar{y}_{rnh}$	194.1192	117.2039	119.5628	114.2322
		$\bar{y}_{tv}$	145.7691	118.6578	123.0296	121.5947
2	ρ_{HHH}	$\bar{y}_{uv}$	146.7985	117.5898	122.7491	122.7091
		$\bar{y}_{rah}$	146.7881	117.5749	122.7485	122.7097
		$\bar{y}_{rnh}$	135.2068	118.1773	120.0823	113.3764
		$\bar{y}_{tv}$	146.7992	117.5802	122.7495	122.7099
3	ρ_{LLL}	$\bar{y}_{uv}$	103.7209	130.4219	146.8077	121.2824
		$\bar{y}_{rah}$	103.0653	129.4315	146.2017	121.7589
		$\bar{y}_{rnh}$	95.81516	133.7055	146.0153	120.2461
		$\bar{y}_{tv}$	104.8436	128.1368	146.3138	122.1731
4	ρ_{LLL}^{-}	$\bar{y}_{uv}$	144.9099	129.0238	127.1195	126.0852
		$\bar{y}_{rah}$	146.3571	128.8966	127.1003	126.0653
		$\bar{y}_{rnh}$	134.8345	127.6844	126.6791	125.6746
		$\bar{y}_{tv}$	149.7035	129.3264	127.3368	126.2446

Table 2: Mean Square Error of Population I as $n \to \infty$

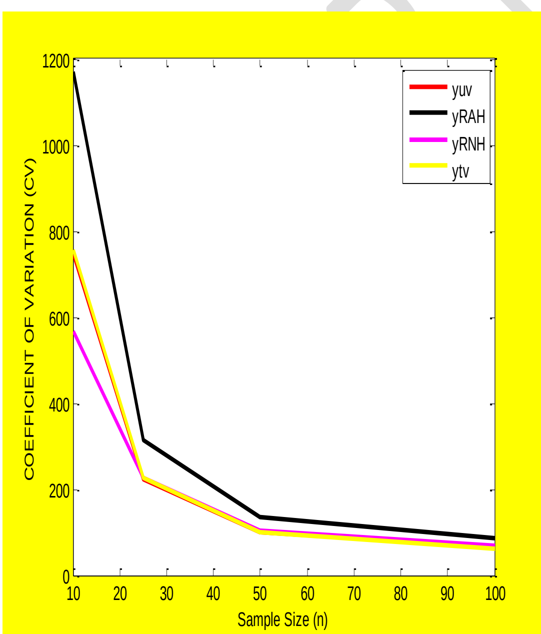
S/No	$Corr.Coeff.$	$Estimator$	Sample Size (n) 10	25	50	100
1	ρ_{HLL}	$\bar{y}_{uv}$	121.1768	7.1448	1.5136	0.5845
		$\bar{y}_{rah}$	323.3008	13.5262	2.6982	1.0518

		$\bar{y}_{rnh}$	121.1768	7.1448	1.5136	0.5845
		$\bar{y}_{tv}$	121.1768	7.1448	1.5136	0.5845
2	ρ_{HHH}	$\bar{y}_{uv}$	83.9331	6.3376	1.5063	0.5703
		$\bar{y}_{rah}$	10765.58	2187.5446	2498.7742	3414.1341
		$\bar{y}_{rnh}$	83.9331	6.3376	1.5063	0.5703
		$\bar{y}_{tv}$	83.9331	6.3376	1.5063	0.5703
3	ρ_{LLL}	$\bar{y}_{uv}$	223.4875	192.9899	69.9694	40.5525
		$\bar{y}_{rah}$	258.637	196.1521	71.8245	40.5574
		$\bar{y}_{rnh}$	223.4875	192.9899	69.9694	40.5525
		$\bar{y}_{tv}$	223.4875	192.9899	69.9694	40.5525
4	$\rho_{\bar{L}LL}$	$\bar{y}_{uv}$	329.9688	148.6753	84.50811	47.8513
		$\bar{y}_{rah}$	344.059	148.6887	85.1384	47.8785
		$\bar{y}_{rnh}$	329.9688	148.6753	84.5081	47.8513
		$\bar{y}_{tv}$	329.9688	148.6753	84.5081	47.8513

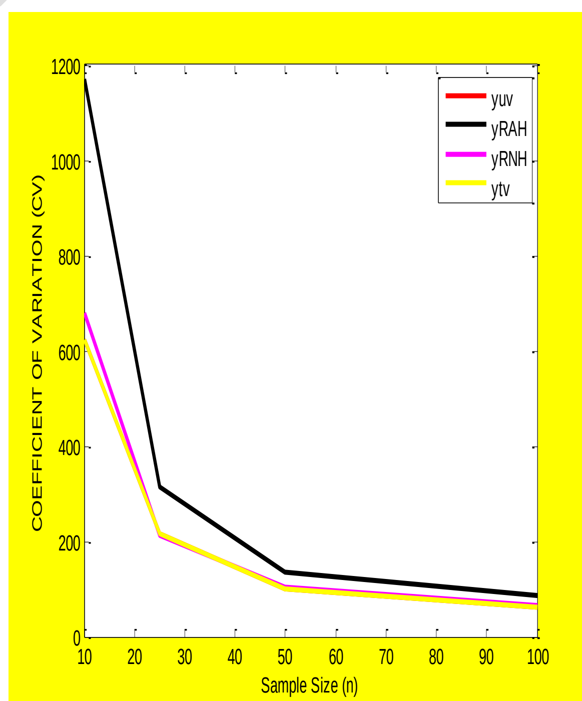
Table 3: Coefficient of Variation of Population I as $n \to \infty$

S/No	$Corr.Coeff.$	$Estimator$	$Sample\ Size\ (n)$			
			10	25	50	100
		$\bar{y}_{uv}$	746.7133	221.9855	100.4975726	65.4559

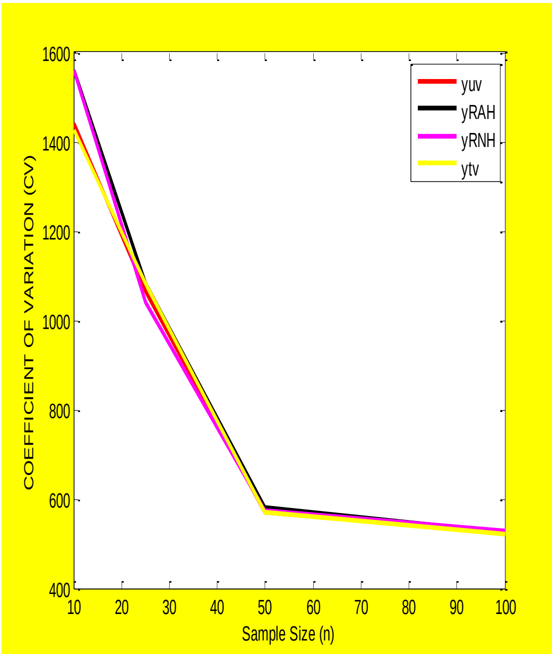
1	ρ_{HLL}	$\bar{y}_{rah}$	1171.607	313.1472	135.0028942	85.2867
		$\bar{y}_{rnh}$	567.0758	228.0614	102.8992177	66.9284
		$\bar{y}_{tv}$	755.1692	225.2671	99.9996388	62.8759
2	ρ_{HHH}	$\bar{y}_{uv}$	624.087	214.0891	99.9858889	61.5448
		$\bar{y}_{rah}$	7068.512	3977.9885	4072.369351	4761.6916
		$\bar{y}_{rnh}$	677.5915	213.0246	102.2063491	66.6109
		$\bar{y}_{tv}$	624.0837	214.1064	99.9855416	61.5444
3	ρ_{LLL}	$\bar{y}_{uv}$	1441.319	1065.1644	569.7773019	525.0625
		$\bar{y}_{rah}$	1560.389	1082.0729	579.6739417	523.0391
		$\bar{y}_{rnh}$	1560.243	1039.0064	572.8695164	529.5877
		$\bar{y}_{tv}$	1425.886	1084.1602	571.7005986	521.2346
4	$\rho_{\bar{L}LL}$	$\bar{y}_{uv}$	1253.540	945.0388	723.1641231	548.6341
		$\bar{y}_{rah}$	1267.368	946.0143	725.9656075	548.8764
		$\bar{y}_{rnh}$	1347.210	954.9521	725.6784863	550.4262
		$\bar{y}_{tv}$	1213.402	942.8273	721.9302	547.9411



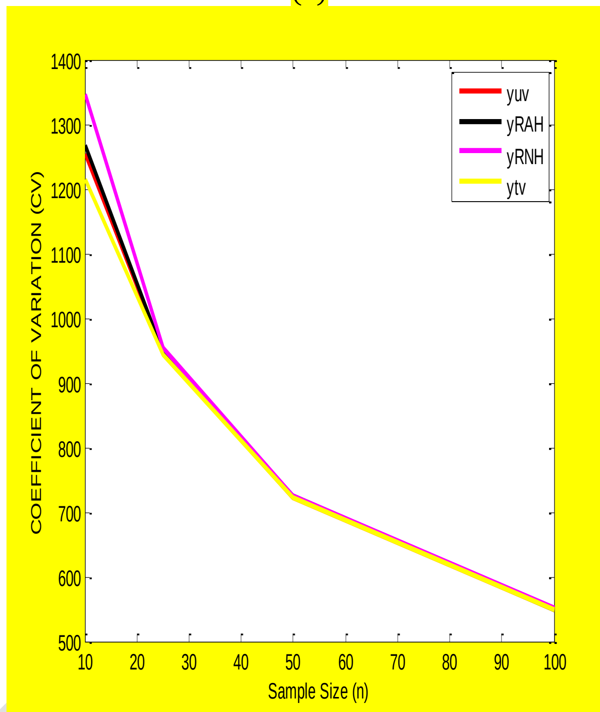
(a)



(b)



(c)



(d)

Figure 1: Graph showing Coefficient of Variation of Population I (a) ρ_{HLL} (b) ρ_{HHH} (c) ρ_{LLL} (d) $\rho_{\bar{L}LL}$

Table 4 : Estimated Mean of Population II

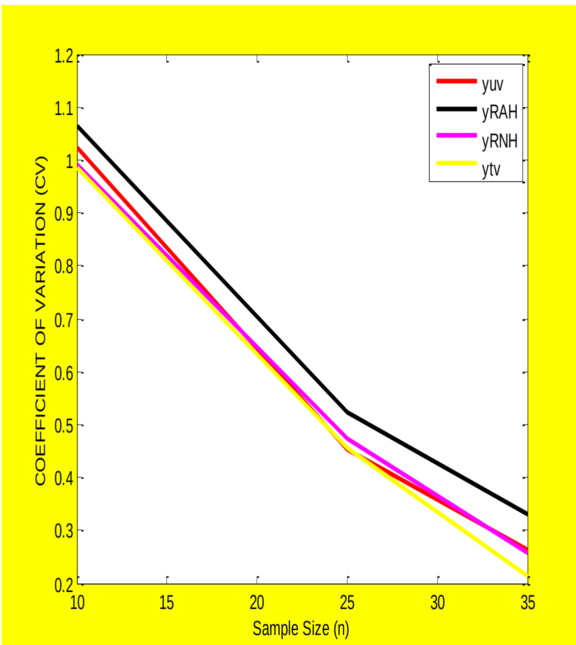
S/No	$Corr.Coeff.$	$Estimator$	$Sample\ Size\ (n)$ 10	25	35
1	ρ_{HLL}	$\bar{y}_{uv}$	6.3383	5.9181	5.2240
		$\bar{y}_{rah}$	6.6081	5.8879	4.3158
		$\bar{y}_{rnh}$	6.5481	5.6662	5.3222
		$\bar{y}_{tv}$	6.5898	5.8737	6.4415
2	ρ_{HHH}	$\bar{y}_{uv}$	6.6134	5.8965	4.3062
		$\bar{y}_{rah}$	6.6110	5.8967	4.3016
		$\bar{y}_{rnh}$	6.6056	5.7568	5.3540
		$\bar{y}_{tv}$	6.6111	5.8965	4.3146
3	ρ_{LLL}	$\bar{y}_{uv}$	6.0836	6.2053	4.6485
		$\bar{y}_{rah}$	6.0651	6.2137	4.3686
		$\bar{y}_{rnh}$	6.0330	6.1856	5.0513
		$\bar{y}_{tv}$	6.0673	6.2146	4.8663
4	$\rho_{\bar{L}LL}$	$\bar{y}_{uv}$	6.5918	5.8971	4.3470
		$\bar{y}_{rah}$	6.6082	5.8963	4.4203
		$\bar{y}_{rnh}$	6.5364	5.8905	4.2470
		$\bar{y}_{tv}$	6.6101	5.8952	4.2828

Table 5: Mean Square Error of Population II

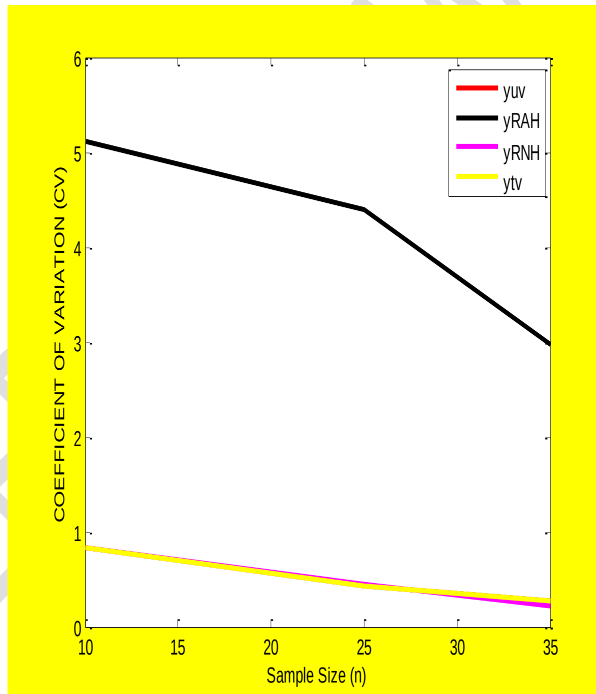
S/No	$Corr.Coeff.$	$Estimator$	$Sample\ Size\ (n)$		
			10	25	35
1	ρ_{HLL}	$\bar{y}_{uv}$	1.0242	0.4526	0.2618
		$\bar{y}_{rah}$	1.0629	0.5239	0.3295
		$\bar{y}_{rnh}$	0.9914	0.4727	0.2571
		$\bar{y}_{tv}$	0.9851	0.4561	0.2124
2	ρ_{HHH}	$\bar{y}_{uv}$	0.0031	0.0006	0.0001
		$\bar{y}_{rah}$	0.1146	0.0673	0.0163
		$\bar{y}_{rnh}$	0.0031	0.0006	0.0001
		$\bar{y}_{tv}$	0.0031	0.0006	0.0001
3	ρ_{LLL}	$\bar{y}_{uv}$	0.0254	0.0073	0.0021
		$\bar{y}_{rah}$	0.0295	0.0078	0.0021
		$\bar{y}_{rnh}$	0.0254	0.0073	0.0021
		$\bar{y}_{tv}$	0.0254	0.0073	0.0021
4	$\rho_{\bar{L}LL}$	$\bar{y}_{uv}$	0.0259	0.0061	0.0026
		$\bar{y}_{rah}$	0.0262	0.0061	0.0026
		$\bar{y}_{rnh}$	0.0259	0.0061	0.0026
		$\bar{y}_{tv}$	0.0259	0.0061	0.0026

Table 6: Coefficient of Variation of Population II

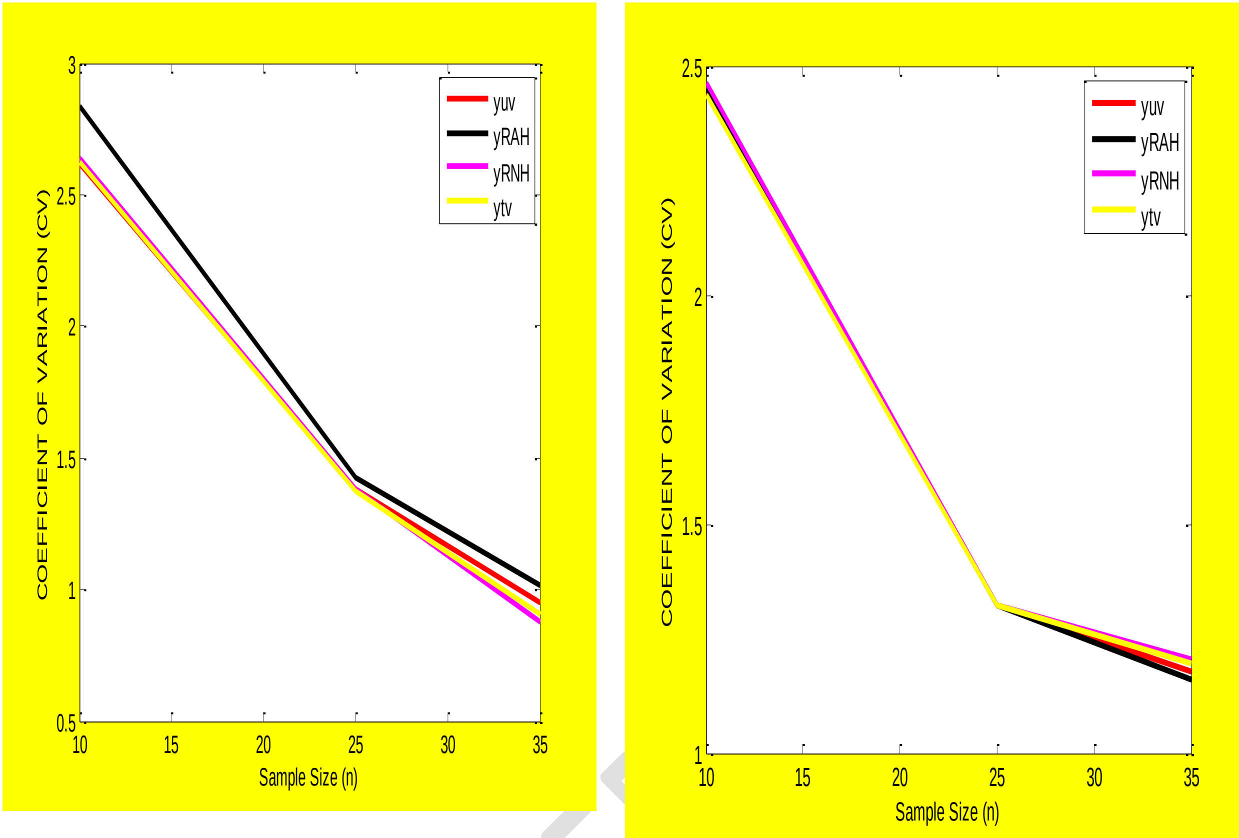
S/No	$Corr.Coeff.$	$Estimator$	$Sample\ Size\ (n)$		
			10	25	35
1	ρ_{HLL}	$\bar{y}_{uv}$	1.0242	0.4526	0.2618
		$\bar{y}_{rah}$	1.0629	0.5239	0.3295
		$\bar{y}_{rnh}$	0.9914	0.4727	0.2571
		$\bar{y}_{tv}$	0.9851	0.4561	0.2124
2	ρ_{HHH}	$\bar{y}_{uv}$	0.8289	0.4316	0.2638
		$\bar{y}_{rah}$	5.1207	4.3994	2.9681
		$\bar{y}_{rnh}$	0.8299	0.4421	0.2121
		$\bar{y}_{tv}$	0.8292	0.4316	0.2633
3	ρ_{LLL}	$\bar{y}_{uv}$	2.6196	1.3772	0.9518
		$\bar{y}_{rah}$	2.8336	1.4252	1.0129
		$\bar{y}_{rnh}$	2.6415	1.3816	0.8759
		$\bar{y}_{tv}$	2.6266	1.3751	0.9092
4	$\rho_{\bar{L}LL}$	$\bar{y}_{uv}$	2.4431	1.3207	1.1775
		$\bar{y}_{rah}$	2.4510	1.3209	1.1581
		$\bar{y}_{rnh}$	2.4638	1.3222	1.2053
		$\bar{y}_{tv}$	2.4363	1.3211	1.1952


COEFFICIENT OF VARIATION (CV)
Sample Size (n)
yuv
yRAH
yRNH
ytv
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
10
15
20
25
30
35

(a)


COEFFICIENT OF VARIATION (CV)
Sample Size (n)
yuv
yRAH
yRNH
ytv
6
5
4
3
2
1
0
10
15
20
25
30
35

(b)



(c) (d)

Figure 2: Graph showing Coefficient of Variation of Population II (a) ρ_{HLL} (b) ρ_{HHH} (c) ρ_{LLL} (d) $\rho_{\bar{L}LL}$

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) ρ_{HLL} (ii) ρ_{HHH} (iii) ρ_{LLL} and (iv) $\rho_{\bar{L}LL}$

.From Tables 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 ρ_{HLL} and underestimated the mean at ρ_{LLL}.

In Table 2, the MSE of $\bar{y}_{RAH}$ is consistently higher for all sample sizes at ρ_{HLL}, ρ_{HHH} and ρ_{LLL}, but fairly stable when $n = 50$ at $\rho_{\bar{L}LL}$. 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 ρ_{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 Figure 1(a). In case of the real data presented in Table 6 and depicted by

Figure 2(a), the proposed estimators are more efficient and precise than the other estimators on the entire parameter space.

Considering the second parameter subspace, ρ_{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 Figure 1(b)). On the other hand, the proposed estimators are more efficient for small samples (Table 6 and Figure 2b) but one of the competing estimators appeared more efficient as sample size increases for the real data set.

In the third region, ρ_{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 Figure 1(c)). Considering the real data, the proposed estimators are more efficient than other estimators for small samples (see Table 6 and Figure 2c).

The fourth experiment considered the region, $\rho_{\bar{L}LL}$. 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 Figure 1d). In case of real data, The proposed are more efficient for small samples as shown in Table 6 and depicted in Figue 2d. Both $\bar{y}_{tv}$ and $\bar{y}_{RAH}$ utilize two auxiliary variables that are transformed according to Srivenkataramana (1980) yet $\bar{y}_{tv}$ 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, (Pradhan, 2005). The transformation of the auxiliary variables has also shown improvement on efficiency of $\bar{y}_{tv}$ as compared to that of $\bar{y}_{uv}$.

8. Concluding Remark

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, ρ_{HLL}.
- All correlation coefficients are high and positive, ρ_{HHH}.
- All correlation coefficients are low and positive, ρ_{LLL}.
- All correlation coefficients are low and negative, $\rho_{\bar{L}L}$.

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, $\bar{y}_{uv}$, and $\bar{y}_{tv}$, 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.

REFERENCES

Abid, M. Nasir, A., Haz, Z. N., and Zhengyan, L. (2016). Enhancing the Mean Ratio Estimators for Estimating Population Mean using Non-Conventional Location Parameter. *Revista Colombiana de Estadistica*, 39(1): 63-79.

Akkus, O. (2017). A Comparison of Exponential Type Mean Estimators in Two Phase Sampling. *American Journal of Mathematics and Statistics.* 7(5): 209-225.

Bahl, S., and Tuteja, R. K. (1991). Ratio and Product Type Exponential Estimator. *Information and optimization sciences*, 12: 159-163.

Bowerman, B. L., O'Connell, R. T. and Hand, M. L. (2001). Business Statistics in Practice. 2^{nd} Edition, McGraw-Hill, Irwin, 856pp.

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.

Cochran, W. G. (1942). Sampling Theory When the Sampling Units are Unequal Sizes. *Journal American Statistical Association*, 37: 199-212.

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.

Etebong, P. C., Ekpoese, E. D., and Inyang, E. j. (2021). Improving the Efficiency of Estimators of Population Mean in Systematic Sampling using Inverse Exponentiation. *International Journal of Statistics and Applied Mathematics,* 6(3): 87-96.

Gupta, S., Zatezalo, T., and Shabbir, J. (2017). A General Mixture of the Mean of a Sensitive Variable in the Presence of Non-Sensitive Auxiliary Information. *Statistics and Applications*, 15(1&2): 27-36.

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.

Hasan, M. Z., Sultan, M., Fatema, K., Hossain, M. A., and Hossain, M. M. (2020). A New Regression Type Estimator and its Application in Survey Sampling. *Open Journal of Statistics*, 10(6): 1010-1019.

Hussain, S., Sharma, M., and Jeelani, M. I. B. (2021). Optimum Exponential Ratio Type Estimators for Estimating Population Mean. *Journal of Statistics Application and Probability Letters*, 2: 73-82.

Irfan, M., Javed, M., Abid, M., and Lin, Z. (2018). Improved Ratio Type Estimators of Population Mean based on Median of a Study Variable and an Auxiliary Variable. *Harcettepe Journal of Mathematics and Statistics,* 47(3): 659-673.

Jabbar, M., Shabbir, J., Ahmed, Z., and Rehman, Z. (2014). Ratio Type Estimator for the Estimation of Finite Population Variance under Two-Stage Sampling. *Research Journal of Applied Sciences, Engineering and Technology*, 7(19): 4095-4099.

Jeelani, M. C., Maqbool, S., and Mir, S. A. (2013). Modified Ratio Estimators of Population Mean Using Linear Combination of Coefficient of Skewness and Quartile Deviation. *International Journal of Modern Mathematical Sciences*, 6(3): 174-183.

Jhajj, H. S., and Lata, K. (2014). A Family of Difference-Cum-Exponential Type Estimators for Estimating the Population Variance Using Auxiliary Information in Sample Surveys. *Advanced Trends in Mathematics*, 1: 15-21.

Jitthavech, J., and Lorchirachoonkul, V. (2013). Estimators in simple random sampling: Searls approach. *Songklanakarin Journal of Science and Technology*. 35(6): 749-760.

Kadilar, C., and Cingi, H. (2004). Ratio Estimators in Simple Random Sampling, *Applied Mathematics and Computation*, 151: 893-902.

Kadilar, C., and Cingi, H. (2006). An Improvement in Estimating the Population mean by using the Correlation Coefficient. *Hacettepe Journal of Mathematics and Statistics*, 35(1): 103-109.

Kumar, A., Singh, A. K., Singh, P., and Singh, V. K. (2017). A Class of Exponential Chain Type Estimator for Population Mean with Imputation of Missing Data Under Double Sampling Scheme. *Journal of Statistics and Probability*, 6(3): 479-485.

Lu, J., Yan, Z. and Peng, X. (2014). A New Exponential Ratio-Type Estimator with Linear Combination of Two Auxiliary Variables, *PLoS ONE*, 9(12): 1-10.

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.

Misra, S., Kumar, B., Yadav, S. K., Yadav, D. K., and Shukla, A. K. (2017). An Improved Ratio Type Predictive Estimator for Estimating Finite Population Mean Using Auxiliary Information. *International Journal of Engineering Sciences and Research Technology*. 6(6): 524-530.

Ogzul, N., and Cingi, H. (2013). A New Class of Exponential Regression-Cum-Ratio Estimator in Two-Phase Sampling. *Hacettepe Journal of Mathematics and Statistics*, 43(1): 131-140.

Onyeka, A. C., Nlebedim, V. U., and Izunobi, C. H. (2013). Estimation of Population Ratio in Simple Random Sampling Using Variable Transformation. *Global Journal of Science Frontier Research Mathematics and Decision Sciences*, 13(4): 56-65.

Pradhan, B. K. (2005). A Chain Regression Estimator in Two Phase Sampling Using Multi-auxiliary Information. *Bulletin of the Malaysian Mathematical Sciences Society*, 28(1): 81-86.

Raja, T. A., Subzar, M., Maqbool, S., and Hakak, A. A. (2017). Enhancing the Mean Ratio Estimator for Estimating Population Mean using Conventional Parameters. *International Journal of Mathematics and Statistics Invention*, 5(1): 58-61

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.

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.

Robson, D. S. (1957). Application of Multivariate Polykays to the Theory of Unbiased Ratio Type Estimators. *Journal of American Statistical Association*, 52: 511-522.

Saini, M., and Kumar, A. (2015). Exponential Type Product Estimator for Finite Population Mean with Information on Auxiliary Attribute. 10(1): 106-113.

Sen, A. R. (1993). Some Early Developments in Ratio Estimation. *Biometric Journal*, 3-13.

Shabbir, J., Ahmed, S., Sanaullah, A., and Onyango, R. (2021). Measuring Performance of Ratio-Exponential-Log Type General Class of Estimators using Two Auxiliary Variables. *Mathematical Problems in Engineering*, 2021: Article ID 5245421, 12 Pages, https//doi.org/10.1155/2021/5245421.

Shabbir, J., Haq, A., and Gupta, S. (2014). A New Difference-cum-Exponential Type Estimator of Finite Population Mean in Simple Random Sampling, *Revista Colombiana de Estadistica*, 37 (1): 197-209.

Shittu, O. I., and Adepoju, K. A. (2014). On the Efficiency of of Modified Ratio Estimator based on Linear Combination of Kurtosis, Median and Quantile Deviation. *International Journal of Mathematical Sciences*, 103-107.

Singh, G. N., and Khahid, M. (2015). Exponential Chain Dual to Ratio and Regression Type Estimators of Population Mean in Two-Phase Sampling. *Statistica*, 75(4): 279-389.

Singh, H. P., and Tailor, R. (2003). Use of Known Coefficient in Estimating the Finite Population Means. *Statistics in Transition*, 6(4): 555-560.

Singh, H. P., and Tailor, R. (2005). Estimation of Finite Population Mean using known Correlation Coefficient between Auxiliary Characters. *Statistica, 65:* 407-418.

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.

Sisodia, B. V. S. and Dwivedi, V. K. (1981). A Class of Ratio-cum-Product Type Estimator in Double Sampling. *Biometric Journal*, 24(4): 419-424.

Srivenkataramana, T. (1980). A Dual to ratio Estimator in Sample Surveys. *Biometrika*, 67: 199-204.

Subramani, J. (2013). Generalized Modified Ratio Estimator of Finite Population Mean. *Journal of Modern Applied Statistical Methods*, 12(2):121-155.

Subramani, J., and Kumarapandiyan, G. (2012). Estimation of Mean using Coefficient of Variation and Median of an Auxiliary Variable. *International Journal of Probability and Statistics*, 1(4): 111-118.

Subzar, M., Maqbool, S., and Raja, T. A. (2017). A Class of Efficient Ratio Type Estimators for the Estimation of Population Mean Using the Auxiliary Information in Survey Sampling. *International Journal of Engineering Development and research*, 5(2): 1756-1761.

Vishwakarma, G. K., and Kumar, M. (2014). An Improved Class of Chain Ratio-Product Type Estimators in Two-Phase Sampling using Two Auxiliary Variables. *Journal of Probability and Statistics*, 2014: 1-6.

Vishwakarma, G. K., and Zeeshan, S. M. (2021). Generalized Ratio-cum-Product Estimator for Finite Population Mean under Two-Phase Sampling Scheme. *Journal of Modern Applied Statistical Methods*, 19(1): Article 7.

Watson, D. J. (1937). The Estimation of Leaf Area in Field Crops. *The Journal of Agricultural Science*, 27(3): 474-483.

Yadav, D. K., D., and Misra, S. (2017). An Efficient Estimator for Estimating Finite Population Mean Using Known Median of the Study Variable. *International Journal of Engineering Sciences and research Technology*, 6(6): 503-509.

Yadav, S. K., and Kadilar, C. (2013). Efficient Family of Exponential Estimators for the Population Mean. *Hacettepe Journal of Mathematics and Statistics*, 42(6): 671-677.

Yadav, S. K., Misra, S., Kumar, R., Verma, S. and Kumar, S. (2016). An Efficient Dual to Ratio and Product Estimators of Population Variance in Sample Surveys, *3*2016: 178-188.

Yan, Z., and Tian, B. (2010). Ratio Method to the Mean Estimation using Coefficient of Skewness of Auxiliary Variable. *Information Computing and Applications*, 106(2): 103-110.

Zaman, T., and Kadilar, C. (2021). Exponential Ratio and Product Type Estimators of the Mean in Stratified and Two-Phase Sampling. *AIMS Mathematics*, 6(5): 4265-4279.

Appendix 1: Proof to Theorem 1

Consider the expression in (11) above,

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

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

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

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

$$= \left[\bar{Y} + \bar{Y}e_y + \beta(-\bar{Z}e_z)\right] exp\left[\alpha\left(\frac{-\bar{X}e_x}{2\bar{X} + \bar{X}e_x}\right)\right]$$

$$= \left[\bar{Y} + \bar{Y}e_y + \beta(-\bar{Z}e_z)\right] exp\left[\alpha\left(\frac{-\bar{X}e_x}{(2 + e_x)\bar{X}}\right)\right]$$

$$= \left[\bar{Y} + \bar{Y}e_y + \beta(-\bar{Z}e_z)\right] exp\left[\alpha\left(\frac{-e_x}{(2 + e_x)}\right)\right]$$

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

$$= \left[\bar{Y} + \bar{Y}e_y + \beta(-\bar{Z}e_z)\right] exp\left[-\frac{1}{2}\alpha e_x\left(1 - \frac{1}{2}e_x + \frac{1}{4}e_x^2 - \cdots\right)\right]$$

$$= \left[\bar{Y} + \bar{Y}e_y + \beta(-\bar{Z}e_z)\right] exp\left[-\frac{1}{2}\alpha e_x + \frac{1}{4}\alpha e_x^2 - \frac{1}{8}\alpha e_x^3 + \cdots\right]$$

By First order approximation principle, we have:

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

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

$$= \left[\bar{Y} + \bar{Y}e_y + \beta(-\bar{Z}e_z)\right]\left[1 - \frac{1}{2}\alpha e_x + \frac{1}{2!}.\frac{1}{2^2}\alpha^2 e_x^2 - \frac{1}{3!}.\frac{1}{2^3}\alpha^3 e_x^3 + \cdots\right]$$

$$= \left[\bar{Y} + \bar{Y}e_y + \beta(-\bar{Z}e_z)\right]\left[1 - \frac{1}{2}\alpha e_x + \frac{1}{8}\alpha^2 e_x^2 - \frac{1}{48}\alpha^3 e_x^3 + \cdots\right]$$

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

$$= \bar{Y}\left[1 - \frac{1}{2}\alpha e_x + \frac{1}{8}\alpha^2 e_x^2\right] + \bar{Y}e_y\left[1 - \frac{1}{2}\alpha e_x + \frac{1}{8}\alpha^2 e_x^2\right] - \beta\bar{Z}e_z\left[1 - \frac{1}{2}\alpha e_x + \frac{1}{8}\alpha^2 e_x^2\right]$$

$$= \bar{Y}\left[1 - \frac{1}{2}\alpha e_x + \frac{1}{8}\alpha^2 e_x^2\right] + \bar{Y}\left[e_y - \frac{1}{2}\alpha e_x e_y + \frac{1}{8}\alpha^2 e_x^2 e_y\right] - \beta\bar{Z}\left[e_z - \frac{1}{2}\alpha e_x e_z + \frac{1}{8}\alpha^2 e_x^2 e_z\right]$$

$$= \bar{Y}\left[1 - \frac{1}{2}\alpha e_x + \frac{1}{8}\alpha^2 e_x^2\right] + \bar{Y}\left[e_y - \frac{1}{2}\alpha e_x e_y + \frac{1}{8}\alpha^2 e_x^2 e_y\right] - \beta\bar{Z}\left[e_z - \frac{1}{2}\alpha e_x e_z + \frac{1}{8}\alpha^2 e_x^2 e_z\right]$$

$$= \bar{Y}\left[1 - \frac{1}{2}\alpha e_x + \frac{1}{8}\alpha^2 e_x^2 + e_y - \frac{1}{2}\alpha e_x e_y + \frac{1}{8}\alpha^2 e_x^2 e_y\right] - \beta\bar{Z}\left[e_z - \frac{1}{2}\alpha e_x e_z + \frac{1}{8}\alpha^2 e_x^2 e_z\right]$$

$$= \bar{Y} + \bar{Y}\left[-\frac{1}{2}\alpha e_x + \frac{1}{8}\alpha^2 e_x^2 + e_y - \frac{1}{2}\alpha e_x e_y\right] - \beta\bar{Z}\left[e_z - \frac{1}{2}\alpha e_x e_z\right]$$

$$\bar{y}_{uv} - \bar{Y} = \bar{Y}\left[-\frac{1}{2}\alpha e_x + \frac{1}{8}\alpha^2 e_x^2 + e_y - \frac{1}{2}\alpha e_x e_y\right] - \beta\bar{Z}\left[e_z - \frac{1}{2}\alpha e_x e_z\right] \tag{18}$$

Taking expectation on both sides of (18),

$$Bias(\bar{y}_{uv}) \quad = E(\bar{y}_{uv} - \bar{Y}) = \bar{Y}E\left[-\frac{1}{2}\alpha e_x + \frac{1}{8}\alpha^2 e_x^2 + e_y - \frac{1}{2}\alpha e_x e_y\right] - \beta\bar{Z}E\left[e_z - \frac{1}{2}\alpha e_x e_z\right]$$

$$= \bar{Y}\left[-\frac{1}{2}\alpha E(e_x) + \frac{1}{8}\alpha^2 E(e_x^2) + E(e_y) - \frac{1}{2}\alpha E(e_x e_y)\right]$$

$$-\beta\bar{Z}\left[E(e_z) - \frac{1}{2}\alpha E(e_x e_z)\right] \qquad (19)$$

Applying the definitions of Expectations in section 1 to (19), we have

$$= \bar{Y}\left[0 + \frac{1}{8}\alpha^2\theta C_x^2 + 0 - \frac{1}{2}\alpha\theta\rho_{yx}C_xC_y\right] - \beta\bar{Z}\left[0 - \frac{1}{2}\alpha\theta\rho_{xz}C_xC_z\right]$$

$$= \bar{Y}\theta\left(\frac{1}{8}\alpha^2 C_x^2 - \frac{1}{2}\alpha\rho_{yx}C_xC_y\right) + \frac{1}{2}\beta\bar{Z}\alpha\theta\rho_{xz}C_xC_z$$

Therefore,

$$Bias(\bar{y}_{uv}) = \bar{Y}\theta\left(\frac{1}{8}\alpha^2 C_x^2 - \frac{1}{2}\alpha\rho_{yx}C_xC_y\right) + \frac{1}{2}\beta\bar{Z}\alpha\theta\rho_{xz}C_xC_z$$

Appendix 2: Proof to Theorem 2

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

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

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

$$= \bar{Y}^2\left(e_y - \frac{1}{2}\alpha e_x\right)^2 - 2\,\beta\bar{Z}e_z\left(e_y - \frac{1}{2}\alpha e_x\right) + \beta^2\bar{Z}^2 e_z^2$$

$$= \bar{Y}^2\left(e_y^2 - 2.\frac{1}{2}\alpha e_x e_y + \frac{1}{4}\alpha^2 e_x^2\right)^2 - 2\,\beta\bar{Y}\bar{Z}\left(e_y e_z - \frac{1}{2}\alpha e_x e_z\right) + \beta^2\bar{Z}^2 e_z^2$$

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

$$-2\,\beta\bar{Y}\bar{Z}\left[E(e_y e_z) - \frac{1}{2}\alpha E(e_x e_z)\right] + \beta^2\bar{Z}^2 E(e_z^2)$$

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

$$= \bar{Y}^2\left[\theta C_y^2 - \alpha\theta\rho_{y,x}C_xC_y + \frac{1}{4}\alpha^2\theta C_x^2\right]^2$$

$$-2\,\beta\bar{Y}\bar{Z}\left[\theta\rho_{y,z}C_yC_z-\frac{1}{2}\alpha\theta\rho_{x,z}C_xC_z\right]+\beta^2\bar{Z}^2\theta C_z^2 \tag{20}$$

To obtain the optimal value of α that minimizes the $MSE(\bar{y}_{uv})$, we differentiate (20) with respect to α,β and equate to zero.

$$\frac{\partial MSE(\bar{y}_{uv})}{\partial\alpha}=\frac{\partial}{\partial\alpha}\left\{\bar{Y}^2\left[\theta C_y^2-\alpha\theta\rho_{yx}C_xC_y+\frac{1}{4}\alpha^2\theta C_x^2\right]^2-2\,\beta\bar{Y}\bar{Z}\left[\theta\rho_{yz}C_yC_z-\frac{1}{2}\alpha\theta\rho_{xz}C_xC_z\right]+\beta^2\bar{Z}^2\theta C_z^2\right\}=0$$

$$\Rightarrow\bar{Y}^2\left(\frac{1}{2}\alpha\,\theta C_x^2-\theta\rho_{yx}C_xC_y\right)+\ \beta\bar{Y}\bar{Z}\,\theta\rho_{xz}C_xC_z=0$$

$$\bar{Y}^2\left(\frac{\alpha}{2}\,\theta C_x^2-\theta\rho_{yx}C_xC_y\right)=-\ \beta\bar{Y}\bar{Z}\,\theta\rho_{xz}C_xC_z$$

$$\frac{\alpha}{2}\ =\frac{\bar{Y}\rho_{yx}C_xC_y-\beta\bar{Z}\,\rho_{xz}C_xC_z}{\bar{Y}C_x^2}$$

$$\alpha\ =\frac{2\left(\bar{Y}\rho_{yx}C_y-\beta\bar{Z}\,\rho_{xz}C_z\right)}{\bar{Y}C_x} \tag{21}$$

$$\frac{\partial MSE(\bar{y}_{uv})}{\partial\,\beta}=2\bar{Y}\bar{Z}\left(\frac{1}{2}\alpha\theta\rho_{xz}C_xC_z-\theta\rho_{yz}C_yC_z\right)+2\beta\bar{Z}^2C_z^2\theta=0$$

$$\bar{Y}\bar{Z}\left(\frac{1}{2}\alpha\theta\rho_{xz}C_xC_z-\theta\rho_{yz}C_yC_z\right)+\beta\bar{Z}^2C_z^2\theta=0$$

$$\beta\bar{Z}C_z\theta=\bar{Y}\left(\theta\rho_{yz}C_y-\frac{1}{2}\alpha\theta\rho_{xz}C_x\right)$$

$$\beta=-\frac{1}{2}\frac{\bar{Y}\left(\alpha\rho_{xz}C_x-2\rho_{yz}C_y\right)}{\bar{Z}C_z} \tag{22a}$$

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

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

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

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

$$\beta=\frac{\bar{Y}\left(\rho_{yx}\rho_{xz}C_y-\rho_{yz}C_y\right)}{\bar{Z}\rho_{xz}^2C_z-\bar{Z}C_z} \tag{22b}$$

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

$$MSE(\bar{y}_{uv})=\frac{\bar{Y}^2C_y^2\theta}{\rho_{x,z}^2-1}\left(\rho_{y,x}^2+\rho_{x,z}^2+\rho_{y,z}^2-2\rho_{y,x}\rho_{x,z}\rho_{y,z}-1\right);\ \rho_{x,z}\neq\pm1;\rho_{y,x},\rho_{y,z},\neq 1 \tag{23}$$

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

$$Bias(\bar{y}_{uv}) = \bar{Y}\theta\left(\frac{1}{8}\alpha^2 C_x^2 - \frac{1}{2}\alpha\rho_{y,x}C_x C_y\right) + \frac{1}{2}\beta\bar{Z}\alpha\theta\rho_{x,z}C_x C_z$$

$$= -\frac{1}{2}\frac{\bar{Y}C_y^2\theta}{(\rho_{xz}^2 - 1)^2}\left(\rho_{xz}\rho_{yz} - \rho_{yx}\right)^2;\ \ \rho_{xz} \neq 1 \qquad (24)$$

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:

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

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

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

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

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

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

By first order approximation,

$$= \left[\bar{Y} + \bar{Y}e_y - \beta_1 ge_z\bar{Z}\right]\exp\left[-\frac{1}{2}\alpha_1 ge_x\right] \qquad (25)$$

Expanding the exponent in the expression (25) we have:

$$= \left[\bar{Y} + \bar{Y}e_y - \beta_1 ge_z\bar{Z}\right]\left[1 - \frac{1}{2}\alpha_1 ge_x + \frac{1}{2!}\left(\frac{1}{2}\alpha_1 ge_x\right)^2 - \frac{1}{3!}\left(\frac{1}{2}\alpha_1 ge_x\right)^3 + \cdots\right.$$

$$= \left[\bar{Y} + \bar{Y}e_y - \beta_1 ge_z\bar{Z}\right]\left[1 - \frac{1}{2}\alpha_1 ge_x + \frac{1}{8}\alpha_1^2 g^2e_x^2 - \frac{1}{48}\alpha_1^3 g^3e_x^3 + \cdots\right]$$

$$= \left[\bar{Y} + \bar{Y}e_y - \beta_1 ge_z\bar{Z}\right]\left[1 - \frac{1}{2}\alpha_1 ge_x + \frac{1}{8}\alpha_1^2 g^2e_x^2\right]$$

$$= \bar{Y}\left[1 - \frac{1}{2}\alpha_1 ge_x + \frac{1}{8}\alpha_1^2 g^2e_x^2\right] + \bar{Y}e_y\left[1 - \frac{1}{2}\alpha_1 ge_x + \frac{1}{8}\alpha_1^2 g^2e_x^2\right] - \beta_1 ge_z\bar{Z}\left[1 - \frac{1}{2}\alpha_1 ge_x + \frac{1}{8}\alpha_1^2 g^2e_x^2\right]$$

$$= \bar{Y} + \bar{Y}\left[-\frac{1}{2}\alpha_1 g e_x + \frac{1}{8}\alpha_1^2 g^2 e_x^2\right] + \bar{Y}\left[e_y - \frac{1}{2}\alpha_1 g e_x e_y + \frac{1}{8}\alpha_1^2 g^2 e_x^2 e_y\right]$$
$$-\beta_1 \bar{Z}\left[g e_z - \frac{1}{2}\alpha_1 g^2 e_x e_z + \frac{1}{8}\alpha_1^2 g^3 e_x^2 e_z\right]$$
$$\bar{y}_{tv} - \bar{Y} = \bar{Y}\left[-\frac{1}{2}\alpha_1 g e_x - e_y - \frac{1}{2}\alpha_1 g e_x e_y + \frac{1}{8}\alpha_1^2 g^2 e_x^2\right]$$
$$-\beta_1 \bar{Z}\left[g e_z - \frac{1}{2}\alpha_1 g^2 e_x e_z\right] \qquad (26)$$

Taking expectation on both sides of (26) and subsequent application of the definitions in section 1 gives us the bias.

$$Bias(\bar{y}_{tv}) = E(\bar{y}_{tv} - \bar{Y})$$
$$= \bar{Y}\left[-\frac{1}{2}\alpha_1 g E(e_x) + E(e_y) - \frac{1}{2}\alpha_1 g E(e_x e_y)\right.$$
$$\left. + \frac{1}{8}\alpha_1^2 g^2 E(e_x^2)\right] - \beta_1 \bar{Z}\left[g E(e_z) - \frac{1}{2}\alpha_1 g^2 E(e_x e_z)\right]$$
$$= \bar{Y}\left[0 - 0 - \frac{1}{2}\alpha_1 g \theta \rho_{yx} C_x C_y + \frac{1}{8}\alpha_1^2 g^2 \theta C_x^2\right]$$
$$-\beta_1 \bar{Z}\left[0 - \frac{1}{2}\alpha_1 g^2 \theta \rho_{xz} C_x C_z\right]$$
$$= \bar{Y}\left[-\frac{1}{2}\alpha_1 g \theta \rho_{yx} C_x C_y + \frac{1}{8}\alpha_1^2 g^2 \theta C_x^2\right] + \beta_1 \bar{Z}\left[\frac{1}{2}\alpha_1 g^2 \theta \rho_{xz} C_x C_z\right]$$

Therefore,

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

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.

$$(\bar{y}_{tv} - \bar{Y})^2 = \left\{\bar{Y}\left(-\frac{1}{2}\alpha_1 g e_x + e_y - \frac{1}{2}\alpha_1 g e_x e_y + \frac{1}{8}\alpha_1^2 g^2 e_x^2\right)\right.$$
$$\left. -\beta_1 \bar{Z}\left(g e_z - \frac{1}{2}\alpha_1 g^2 e_x e_z\right)\right\}^2$$
$$= \left[\bar{Y}\left(-\frac{1}{2}\alpha_1 g e_x + e_y\right) - \beta \bar{Z}(g e_z)\right]^2$$
$$= \bar{Y}^2\left(-\frac{1}{2}\alpha_1 g e_x + e_y\right)^2 - 2\,\beta_1 \bar{Y}\bar{Z} g e_z\left(-\frac{1}{2}\alpha_1 g e_x + e_y\right)$$
$$+\beta^2 \bar{Z}^2 g^2 e_z^2$$
$$= \bar{Y}^2\left[\frac{1}{4}\alpha_1^2 g^2 e_x^2 - 2\left(\frac{1}{2}\right)\alpha_1 g e_x e_y + e_y^2\right]$$
$$-2\,\beta \bar{Y}\bar{Z}\left[-\frac{1}{2}\alpha_1 g^2 e_x e_z + g e_y e_z\right]$$
$$+\beta^2 \bar{Z}^2 g^2 e_z^2 \qquad (27)$$

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

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

$$+\beta^2\bar{Z}^2g^2E(e_z^2)$$

$$=\bar{Y}^2\left(\frac{1}{4}\alpha_1^2g^2\theta C_x^2-g\theta\rho_{yx}C_xC_y+\theta C_y^2\right)$$

$$-2\,\beta\bar{Y}\bar{Z}\left(g\theta\rho_{yz}C_yC_z-\frac{1}{2}\alpha_1g^2\theta\rho_{xz}C_xC_z\right)$$

$$+\beta^2\bar{Z}^2g^2\theta C_z^2$$

Therefore,

$$MSE(\bar{y}_{tv})=E\left(\bar{y}_{tp}-\bar{Y}\right)^2\ =\bar{Y}^2\left(\tfrac{1}{4}\alpha_1^2g^2\theta C_x^2-\alpha_1g\theta\rho_{yx}C_xC_y+\theta C_y^2\right)$$

$$-2\,\beta_1\bar{Y}\bar{Z}\left(g\theta\rho_{yz}C_yC_z-\frac{1}{2}\alpha_1g^2\theta\rho_{xz}C_xC_z\right)$$

$$+\beta_1^2\bar{Z}^2g^2\theta C_z^2 \qquad (28)$$

To obtain the value of α_1, β_1 that minimizes the MSE, we differentiate Equation (28) partially with respect to α_1,β_1 and equate to zero:

$$\frac{\partial MSE(\bar{y}_{tv})}{\partial\alpha_1}=\frac{\partial}{\partial\alpha_1}\Big\{\bar{Y}^2\left(\frac{1}{4}\alpha_1^2g^2\theta C_x^2-\alpha_1g\theta\rho_{yx}C_xC_y+\theta C_y^2\right)$$

$$-2\,\beta_1\bar{Y}\bar{Z}\left(\frac{1}{2}\alpha_1g^2\theta\rho_{xz}C_xC_z-g\theta\rho_{yz}C_yC_z\right)+\beta_1^2\bar{Z}^2g^2\theta C_z^2\Big\}\ =0$$

$$\bar{Y}^2\left(2.\frac{1}{4}\alpha_1g^2\theta C_x^2-g\theta\rho_{yx}C_xC_y\right)+2\beta_1\bar{Y}\bar{Z}\left(\frac{1}{2}g^2\theta\rho_{xz}C_xC_z\right)=0$$

$$\bar{Y}^2\left(\frac{1}{2}\alpha_1g^2\theta C_x^2-g\theta\rho_{yx}C_xC_y\right)+\beta_1\bar{Y}\bar{Z}g^2\theta\rho_{xz}C_xC_z=0$$

$$\frac{1}{2}\alpha_1g^2C_x=\frac{\bar{Y}g\rho_{yx}C_y-\beta_1\bar{Z}g^2\rho_{xz}C_z}{\bar{Y}}$$

$$\alpha_1=\frac{2(\bar{Y}g\rho_{yx}C_y-\beta_1\bar{Z}g^2\rho_{xz}C_z)}{\bar{Y}g^2C_x}$$

Therefore,

$$\alpha_1=\frac{2(\bar{Y}\rho_{yx}C_y-\beta_1\bar{Z}g\rho_{xz}C_z)}{\bar{Y}gC_x} \qquad (29a)$$

$$\frac{\partial MSE(\bar{y}_{tv})}{\partial\beta_1}=-2\bar{Y}\bar{Z}\left(\rho_{yz}C_yC_z-\frac{1}{2}\alpha_1g^2\theta\rho_{xz}C_xC_zg\theta\right)+2\,\beta_1\bar{Z}^2g^2\theta C_z^2\ =0$$

$$-\bar{Y}\bar{Z}\left(-\frac{1}{2}\alpha_1g^2\theta\rho_{xz}C_xC_z+g\theta\rho_{yz}C_yC_z\right)-\ \beta_1\bar{Z}^2g^2\theta C_z^2\ =0$$

Solving for α_1 again we have

$$\alpha_1=\frac{2(\bar{Y}\rho_{xz}C_y+\ \beta_1\bar{Z}gC_z)}{\bar{Y}g\rho_{xz}C_x} \qquad (29b)$$

Equating 29a) to (29b) and solving for β_1 we have

$$\beta_1=\frac{\bar{Y}C_y(\rho_{yx}\rho_{xz}-\rho_{yz})}{\bar{Z}gC_z(\rho_{xz}^2-1)} \qquad (30)$$

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

$$\alpha_1=\frac{2C_y(\rho_{yx}\rho_{xz}-\rho_{yz})}{gC_z(\rho_{xz}^2-1)} \qquad (31)$$

We finally substitute the expressions for α_1 and β_1 in (31) and (30) above into (28) and simplify using maple to obtain the $MSE(\bar{y}_{tv})$ as:

$$MSE(\bar{y}_{tv}) = \frac{\bar{Y}^2\theta C_y^2\left(\rho_{yz}^2 + \rho_{xz}^2 + \rho_{yx}^2 - 2\rho_{yx}\rho_{xz}\rho_{yz} - 1\right)}{\rho_{xz}^2 - 1}$$

Therefore,

$$MSE(\bar{y}_{tv}) = \bar{Y}^2\theta C_y^2\ (\rho_{xz}^2 - 1)^{-1}\left(\rho_{yz}^2 + \rho_{xz}^2 + \rho_{yx}^2 - 2\rho_{yx}\rho_{xz}\rho_{yz} - 1\right) \qquad (32)$$

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

Recall from (15) that:

$$Bias(\bar{y}_{tv}) = \bar{Y}\left[\frac{1}{8}\alpha_1^2 g^2\theta C_x^2 - \frac{1}{2}\alpha_1 g\theta\rho_{yx}C_xC_y\right] + \beta_1\bar{Z}\left[\frac{1}{2}\alpha_1 g^2\theta\rho_{xz}C_xC_z\right]$$

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

$$Bias(\bar{y}_{tv}) = \frac{1}{2}\frac{\bar{Y}C_y^2\theta}{(\rho_{xz}^2 - 1)^2}\left(\rho_{xz}\rho_{yz} - \rho_{yx}\right)^2;\ \ \rho_{xz} \neq 1 \qquad (33)$$