Pak. J. Statist.
2017 Vol. 33(4), 277-300

MODIFIED CLASS OF RATIO AND REGRESSION TYPE ESTIMATORS FOR IMPUTING SCRAMBLING RESPONSE

Muhammad Umair Sohail[§], **Javid Shabbir** and **Shakeel Ahmed**
Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan
[§]Corresponding author Email: umairsohailch@gmail.com

ABSTRACT

In this study, we modify the class of ratio and regression type estimators for imputing scrambling response on the lines of Ahmed et al. (2006) and Mohamed et al. (2016). We propose a new class of estimators by using the higher order moments of auxiliary information. Our proposed class of estimators are more efficient as compared to existing estimators. Two numerical studies by using simulated and real life data set at various response rate are also carried out for evaluating the performance of proposed class of estimators.

KEYWORDS

Mean square error, imputation, variance, scrambling response, sensitive variable.

1. INTRODUCTION

In most of surveys, the common problem which encounter was the missing values. Hansen and Hurwitz (1946) first time deal with the missing data and later many studies have been occurred to handle this problem. Rubin (1976) suggested the various imputation methods, which make the data structurally complete. Later Singh and Deo (2003) and Ahmed et al. (2006) suggested many imputation methods to deal with missing data. Usually the non-response is occurred in the socio-economic and political surveys, where researchers are often interested to investigate the sensitive issues of the society.

Initially, Warner (1965) provide the idea of randomized response technique (RRT) to handle the reduced response rate. The main concept behind this idea was to secure the twin objectives: (i) to create the feeling among the respondents that their identity is secured beside their truthful response, (ii) to produce the reliable data to draw the fruitful inference about the population. Later on many studies have been occurred to improve the efficiency of randomized device.

In this research, we consider the scrambling model proposed by Gjestvang and Singh (2009), where the j^{th} respondent has two choices: he can report $Y_j + \eta S_j$ or $Y_j - \gamma S_j$ with probability $p = \frac{\gamma}{(\eta+\gamma)}$ and $(1-p) = \frac{\eta}{(\eta+\gamma)}$ respectively, where Y_j and S_j be the sensitive and scrambling variables respectively, and η and γ are two positive real numbers. A simple random sample of size n is selected from population of N units.

Consider a deck of cards, which has the p proportion of cards that bearing statement: multiplied the scrambling variable S_j with η and add the true value of Y_j, and $(1-p)$ be the proportion of cards that bearing the statement: multiplied the scrambling variable S_j with γ and subtracted it from the true value of Y_j. The j^{th} respondent can report his response as:

$$z_j = p\left(Y_j + \eta S_j\right) + (1-p)\left(Y_j - \gamma S_j\right). \tag{1.1}$$

Let E_3 and V_3 denoted the expected value and variance due to random device. We assume that $E_3(S) = \mu_s$, $V_3(S) = S_s^2$ are known. So $V_3(z_j) = \eta\gamma\left(\mu_s^2 + S_s^2\right)$.

Let r be the number of respondents in subset A in a sample s who provide the response by using the above mentioned randomize response model and $(n-r)$ are those in subset A^c, who refuse to provide the response, so, $s = A \cup A'$. Let the sample mean of the scrambling response is given as: $\bar{z} = \frac{1}{r}\sum_{j=1}^{r} z_j$. Now we have the Lemma 1 as:

Lemma 1:

The variance of sample mean is given as:

$$V(\bar{z}_r) = \left(\frac{1}{r} - \frac{1}{N}\right)S_y^2 + \frac{1}{r}\eta\gamma(\mu_s^2 + S_s^2) \tag{1.2}$$

Proof:

Let E_1 and E_2 denote the expected values for the given n and r, and V_1 and V_2 denoted the variance for the given values of n and r respectively. By the definition of variance, we have:

$$V(\bar{z}_r) = E_1E_2V_3(\bar{z}_r) + E_1V_2E_3(\bar{z}_r) + V_1E_2E_3(\bar{z}_r)$$

$$= E_1E_2V_3\left[\frac{1}{r}\sum_{j=1}^{r} z_j\right] + E_1V_2E_3\left[\frac{1}{r}\sum_{j=1}^{r} z_j\right] + V_1E_2E_3\left[\frac{1}{r}\sum_{j=1}^{r} z_j\right]$$

$$= E_1E_2\left[\frac{1}{r^2}\sum_{j=1}^{r} V_3(z_j)\right] + E_1V_2\left[\frac{1}{r}\sum_{j=1}^{r} E_3(z_j)\right] + V_1E_2\left[\frac{1}{r}\sum_{j=1}^{r} E_3(z_j)\right]$$

$$= E_1E_2\left[\frac{1}{r^2}\sum_{j=1}^{r}\{\eta\gamma(\mu_s^2 + S_s^2)\}\right] + E_1V_2\left[\frac{1}{r}\sum_{j=1}^{r} y_j\right] + V_1E_2\left[\frac{1}{r}\sum_{j=1}^{r} y_j\right]$$

$$= E_1\left[\frac{1}{nr}\sum_{j=1}^{n}\left\{\eta\gamma(\mu_s^2 + S_s^2)\right\}\right] + E_1\left(\frac{1}{r} - \frac{1}{n}\right)s_{y(n)}^2 + V_1\left[\frac{1}{n}\sum_{j=1}^{n} y_j\right]$$

$$= \frac{1}{Nr}\sum_{j=1}^{N}\left\{\eta\gamma(\mu_s^2+S_s^2)\right\}+\left(\frac{1}{r}-\frac{1}{n}\right)S_y^2+\left(\frac{1}{n}-\frac{1}{N}\right)S_y^2$$

$$= \frac{1}{r}\eta\gamma(\mu_s^2+S_s^2)+\left(\frac{1}{r}-\frac{1}{N}\right)S_y^2. \tag{1.3}$$

We prove the lemma 1, so we have the better approximation for results.

2. SOME EXISTING METHODS OF IMPUTATION FOR SCRAMBLING RESPONSE

i) Mohamed et al. (2016)

$$\hat{z}_j = \begin{cases} z_j & if \rightarrow j \in A \\ \dfrac{z_r}{x_r} x_j & if \rightarrow j \in A^c \end{cases}$$

The point estimator is:

$$\bar{y}_{S.R} = \bar{z}_r \frac{\bar{x}_n}{\bar{x}_r}. \tag{2.1}$$

ii) Ahmed et al. (2006)

$$\hat{z}_j = \begin{cases} z_j & if \rightarrow j \in A \\ \left[\dfrac{\bar{z}_r\left[x_i+\dfrac{r}{n-r}\bar{x}_r\right]}{\alpha_1\bar{x}_r+(1-\alpha_1)\bar{x}_n}-\dfrac{r}{n-r}\bar{z}_r\right] & if \rightarrow j \in A^c \end{cases}$$

The point estimator by this model is give as:

$$\bar{y}_{S.A_8} = \frac{\bar{z}_r\bar{x}_n}{\alpha_1\bar{x}_r+(1-\alpha_1)\bar{x}_n}, \tag{2.2}$$

where α_1 is the constant.

iii) Ahmed et al. (2006)

$$\hat{z}_j = \begin{cases} z_j & if \rightarrow j \in A \\ \dfrac{1}{n-r}\left[\dfrac{n\bar{z}_r\bar{X}}{\alpha_2\bar{x}_n+(1-\alpha_2)\bar{X}}-r\bar{z}_r\right] & if \rightarrow j \in A^c \end{cases}$$

The point estimator is defined as:

$$\bar{y}_{S.A_9} = \frac{\bar{z}_r\bar{X}}{\alpha_2\bar{x}_n+(1-\alpha_2)\bar{X}}, \tag{2.3}$$

where α_2 is the constant.

iv) Ahmed et al. (2006)

$$\hat{z}_j = \begin{cases} z_j & if \rightarrow j \in A \\ \dfrac{1}{n-r}\left[\dfrac{n\bar{z}_r \bar{X}}{\alpha_3 \bar{x}_r + (1-\alpha_3)\bar{X}} - r\bar{z}_r\right] & if \rightarrow j \in A^c \end{cases}$$

The point estimator is defined as:

$$\bar{y}_{S.A_{10}} = \frac{\bar{z}_r \bar{X}}{\alpha_3 \bar{x}_r + (1-\alpha_3)\bar{X}}, \tag{2.4}$$

where α_3 is the constant.

v) Ahmed et al. (2006)

$$\hat{z}_j = \begin{cases} z_j & if \rightarrow j \in A \\ \bar{z}_r + \dfrac{nw_1(\bar{X} - \bar{x}_n)}{n-r} & if \rightarrow j \in A^c \end{cases}$$

The point estimator by this model is give as:

$$\bar{y}_{S.A_5} = \bar{z}_r + w_1(\bar{X} - \bar{x}_n), \tag{2.5}$$

where w_1 is the constant.

vi) Ahmed et al. (2006)

$$\hat{z}_j = \begin{cases} z_j & if \rightarrow j \in A \\ \bar{z}_r + \dfrac{nw_2(\bar{X} - \bar{x}_r)}{n-r} & if \rightarrow j \in A^c \end{cases}$$

The point estimator by this model is give as:

$$\bar{y}_{S.A_6} = \bar{z}_r + w_2(\bar{X} - \bar{x}_r), \tag{2.6}$$

where w_2 are the unknown constant.

vii) Mohamed et al. (2016)

$$\hat{z}_j = \begin{cases} z_j & if \rightarrow j \in A \\ \bar{z}_r + \hat{\beta}_s(x_j - \bar{x}) & if \rightarrow j \in A^c \end{cases}$$

The point estimator of $\bar{y}_{S.Re.}$ is given as:

$$\bar{y}_{S.Re.} = \bar{z}_r + \hat{\beta}_s(\bar{x}_n - \bar{x}_r), \tag{2.7}$$

where $\hat{\beta}_s = \dfrac{\sum_{j=1}^{r}(x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^{r}(x_j - \bar{x})^2}$ is the simple regression coefficient.

3. MODIFIED ESTIMATORS

In this section, we modify the ratio and regression type estimators for imputing scrambling response by using the higher order moments of the auxiliary variable. The use of the higher order moments of an auxiliary variable play an important role for the estimation of population parameter of the study variable in survey sampling. The proposed imputation methods are given as:

i) The first proposed method of imputation is

$$\Delta_{Ij} = \begin{cases} z_j & if \rightarrow j \in A \\ \dfrac{\bar{z}_r \left[x_j + \dfrac{r}{n-r}\bar{x}_r \right] \left[\dfrac{n(x_j - \bar{x}_n)^2}{(n-1)} + \dfrac{n\sum_{j \varepsilon A}(x_j - \bar{x}_n)^2}{(n-r)(n-1)} \right]}{\left\{ \alpha_{I1}\bar{x}_r + (1-\alpha_{I1})\bar{x}_n \right\}\left\{ \beta_{I1}s^2_{x(r)} - (1-\beta_{I1})s^2_{x(n)} \right\}} - \dfrac{r}{n-r}\bar{z}_r & if \rightarrow j \in A^c \end{cases} \tag{3.1}$$

where α_{I1} and β_{I1} are the unknown constants whose values are to be determined. The point estimator of population mean under proposed method of imputation, is given as:

$$\bar{y}_{S.SS_I} = \frac{1}{n}\sum_{j=1}^{n}\Delta_{Ij}$$

$$= \frac{1}{n}\left[\sum_{j\varepsilon A} z_j + \frac{\bar{z}_r \left\{ \sum_{j\varepsilon A^c} x_j + \dfrac{r(n-r)}{n-r}\bar{x}_r \right\} \left\{ \dfrac{\sum_{j\varepsilon A^c}(x_j - \bar{x}_n)^2}{(n-1)} + \dfrac{\sum_{j\varepsilon A^c}\sum_{j\varepsilon A}(x_j - \bar{x}_n)^2}{(n-r)(n-1)} \right\}}{\left\{ \alpha_{I1}\bar{x}_r + (1-\alpha_{I1})\bar{x}_n \right\}\left\{ \beta_{I1}s^2_{x(r)} - (1-\beta_{I1})s^2_{x(n)} \right\}} \right]$$

$$-\frac{1}{n}\left[\frac{r(n-r)}{n-r}\bar{z}_r \right]$$

$$= \frac{1}{n}\left[r\bar{z}_r + \frac{\bar{z}_r \{ n\bar{x}_n - r\bar{x}_r + r\bar{x}_r \} \left\{ \dfrac{\sum_{j\varepsilon A^c}(x_j - \bar{x}_n)^2}{(n-1)} + \dfrac{\sum_{j\varepsilon A}(x_j - \bar{x}_n)^2}{(n-1)} \right\}}{\left\{ \alpha_{I1}\bar{x}_r + (1-\alpha_{I1})\bar{x}_n \right\}\left\{ \beta_{I1}s^2_{x(r)} - (1-\beta_{I1})s^2_{x(n)} \right\}} - \frac{r(n-r)}{n-r}\bar{z}_r \right]$$

$$= \frac{1}{n}\left[r\bar{z}_r + \frac{\bar{z}_r n\bar{x}_n \left\{ \frac{\sum_{j\varepsilon A^c}(x_j-\bar{x}_n)^2 + \sum_{j\varepsilon A}(x_i-\bar{x}_n)^2}{n-1} \right\}}{\left\{\alpha_{I1}\bar{x}_r + (1-\alpha_{I1})\bar{x}_n\right\}\left\{\beta_{I1}s^2_{x(r)} - (1-\beta_{I1})s^2_{x(n)}\right\}} - r\bar{z}_r \right]$$

$$= \frac{1}{n}\left[\frac{\bar{z}_r n\bar{x}_n s^2_{x(n)}}{\left\{\alpha_{I1}\bar{x}_r + (1-\alpha_{I1})\bar{x}_n\right\}\left\{\beta_{I1}s^2_{x(r)} - (1-\beta_{I1})s^2_{x(n)}\right\}} \right]$$

$$\bar{y}_{S.SS_I} = \frac{\bar{z}_r \bar{x}_n s^2_{x(n)}}{\left\{\alpha_{I1}\bar{x}_r + (1-\alpha_{I1})\bar{x}_n\right\}\left\{\beta_{I1}s^2_{x(r)} + (1-\beta_{I1})s^2_{x(n)}\right\}}. \quad (3.2)$$

The point estimator given in Equation (3.2) can be used, when we utilized the sample information of the auxiliary variable to imputing the missing value.

ii) The second proposed method of imputation is

$$\Delta_{IIj} = \begin{cases} z_j & if \rightarrow j \in A \\ \frac{1}{n-r}\left[\frac{n\bar{z}_r \bar{X} S^2_x}{\left(\alpha_{II2}\bar{x}_n + (1-\alpha_{II2})\bar{X}\right)\left(\beta_{II2}s^2_{x(n)} + (1-\beta_{II2})S^2_x\right)} - r\bar{z}_r \right] & if \rightarrow j \in A^c \end{cases}$$

(3.3)

where α_{II2} and β_{II2} are the unknown constants whose values are to be determined. The point estimator $\bar{y}_{S.SS_{II}}$ is given as:

$$\bar{y}_{S.SS_{II}} = \frac{1}{n}\sum_{j=1}^{n}\Delta_{IIj}$$

$$= \frac{1}{n}\left[\sum_{j\varepsilon A} z_j + \sum_{j\varepsilon A^c}\frac{1}{n-r}\left\{ \frac{n\bar{z}_r \bar{X} S^2_x}{\left(\alpha_{II2}\bar{x}_n + (1-\alpha_{II2})\bar{X}\right)\left(\beta_{II2}s^2_{x(n)} + (1-\beta_{II2})S^2_x\right)} \right\} \right]$$

$$-\frac{1}{n}\left\{ \sum_{j\varepsilon A^c}\frac{1}{n-r} r\bar{z}_r \right\}$$

$$= \frac{1}{n}\left[r\bar{z}_r + \frac{n-r}{n-r}\left\{ \frac{n\bar{z}_r \bar{X} S^2_x}{\left\{\alpha_{II2}\bar{x}_n + (1-\alpha_{II2})\bar{X}\right\}\left\{\beta_{II2}s^2_{x(n)} + (1-\beta_{II2})S^2_x\right\}} - r\bar{z}_r \right\} \right]$$

$$= \frac{1}{n}\left[r\bar{z}_r + \frac{n\bar{z}_r \bar{X} S_x^2}{\left\{\alpha_{II2}\bar{x}_n + (1-\alpha_{II2})\bar{X}\right\}\left\{\beta_{II2} s_{x(n)}^2 + (1-\beta_{II2}) S_x^2\right\}} - r\bar{z}_r \right]$$

$$\bar{y}_{S.SS_{II}} = \frac{\bar{z}_r \bar{X} S_x^2}{\left\{\alpha_{II2}\bar{x}_n + (1-\alpha_{II2})\bar{X}\right\}\left\{\beta_{II2} s_{x(n)}^2 + (1-\beta_{II2}) S_x^2\right\}}. \qquad (3.4)$$

The point estimator in Equation (3.4) can be used in situation, when population mean and population variance of the auxiliary variable of the auxiliary variable are known.

iii) The third proposed method of imputation

$$\Delta_{IIIj} = \begin{cases} z_j & if \rightarrow j \in A \\ \dfrac{1}{n-r}\left[\dfrac{n\bar{z}_r \bar{X} S_x^2}{\left(\alpha_{III3}\bar{x}_r + (1-\alpha_{III3})\bar{X}\right)\left(\beta_{III3} s_{x(r)}^2 + (1-\beta_{III3}) S_x^2\right)} - r\bar{z}_r \right] & if \rightarrow j \in A^c \end{cases} \qquad (3.5)$$

where α_{III3} and β_{III3} are the unknown constants, whose values are to be determined. The point estimator of $\bar{y}_{S.SS_{III}}$ is given as:

$$\bar{y}_{S.SS_{III}} = \frac{1}{n}\sum_{j=1}^{n}\Delta_{IIIj}$$

$$= \frac{1}{n}\left[\sum_{j\varepsilon A} z_j + \sum_{j\varepsilon A^c} \frac{1}{n-r}\left\{ \frac{n\bar{z}_r \bar{X} S_x^2}{\left\{\alpha_{III3}\bar{x}_r + (1-\alpha_{III3})\bar{X}\right\}\left\{\beta_{III3} s_{x(r)}^2 + (1-\beta_{III3}) S_x^2\right\}} \right\} \right]$$

$$- \frac{1}{n}\left\{ \sum_{j\varepsilon A^c} \frac{1}{n-r} r\bar{z}_r \right\}$$

$$= \frac{1}{n}\left[r\bar{z}_r + \frac{n-r}{n-r}\left\{ \frac{n\bar{z}_r \bar{X} S_x^2}{\left\{\alpha_{III3}\bar{x}_r + (1-\alpha_{III3})\bar{X}\right\}\left\{\beta_{III3} s_{x(r)}^2 + (1-\beta_{III3}) S_x^2\right\}} - r\bar{z}_r \right\} \right]$$

$$= \frac{1}{n}\left[r\bar{z}_r + \frac{n\bar{z}_r \bar{X} S_x^2}{\left\{\alpha_{III3}\bar{x}_r + (1-\alpha_{III3})\bar{X}\right\}\left\{\beta_{III3} s_{x(r)}^2 + (1-\beta_{III3}) S_x^2\right\}} - r\bar{z}_r \right]$$

$$\bar{y}_{S.SS_{III}} = \frac{\bar{z}_r \bar{X} S_x^2}{\left\{\alpha_{III3}\bar{x}_r + (1-\alpha_{III3})\bar{X}\right\}\left\{\beta_{III3} s_{x(r)}^2 + (1-\beta_{III3}) S_x^2\right\}}. \qquad (3.6)$$

The point estimator in Equation (3.6) can be used, when the population mean and population variance are known.

iv) Fourth proposed method of imputation

$$\Delta_{IVj} = \begin{cases} z_j & if \rightarrow j \in A \\ \bar{z}_r + \alpha_{IV4}\left(\dfrac{n\sum_{j=1}^{N} x_j}{N(n-r)} - \sum_{j\varepsilon A} x_j\right) + \beta_{IV4}\left(\dfrac{n\sum_{j=1}^{N}(x_j - \bar{X})^2}{(n-r)(N-1)} - \dfrac{n\sum_{j\varepsilon A}(x_j - \bar{x}_n)^2}{(n-1)}\right) & if \rightarrow j \in A^c \end{cases}$$

(3.7)

where α_{IV4} and β_{IV4} are constants, whose values are to be determined. The point estimator of $\bar{y}_{S.SS_{IV}}$ is given as:

$$\bar{y}_{S.SS_{IV}} = \frac{1}{n}\sum_{j=1}^{n}\Delta_{IVj}$$

$$= \frac{1}{n}\left[\sum_{j\varepsilon A} z_j + \sum_{j\varepsilon A^c}\bar{z}_r + \alpha_{IV4}\left\{\frac{n\sum_{j\varepsilon A^c}\sum_{j=1}^{N} x_j}{N(n-r)} - \sum_{j\varepsilon A^c}\sum_{j\varepsilon A} x_j\right\}\right]$$

$$+\frac{1}{n}\beta_{IV4}\left[\frac{n\sum_{j\varepsilon A^c}\sum_{j=1}^{N}(x_j - \bar{X})^2}{(n-r)(N-1)} - \frac{n\sum_{j\varepsilon A^c}\sum_{j\varepsilon A}(x_j - \bar{x}_n)^2}{(n-1)}\right]$$

$$= \frac{1}{n}\left[r\bar{z}_r + (n-r)\bar{z}_r + \alpha_{IV4}\left(n\bar{X} - n\bar{x}_n\right) + \beta_{IV4}\left\{\frac{n\sum_{j=1}^{N}(x_j - \bar{X})^2}{(N-1)} - \frac{n\sum_{j\varepsilon s}(x_j - \bar{x}_n)^2}{(n-1)}\right\}\right]$$

$$= \frac{1}{n}\left[n\bar{z}_r + \alpha_{IV4}\left(n\bar{X} - n\bar{x}_n\right) + \beta_{IV4}\{nS_x^2 - ns_{x(n)}^2\}\right]$$

$$\bar{y}_{S.SS_{IV}} = \bar{z}_r + \alpha_{IV4}\left(\bar{X} - \bar{x}_n\right) + \beta_{IV4}\left(S_x^2 - s_{x(n)}^2\right) \quad (3.8)$$

The point estimator in Equation (3.8) is the regression imputation method, is used when the population parameters $(\bar{X}, S_x^2)$ are known.

v) Fifth proposed method of imputation

$$\Delta_{Vj} = \begin{cases} z_j & if \rightarrow j \in A \\ \bar{z}_r + \alpha_{V5}\left(\dfrac{n\sum_{j=1}^{N} x_j}{N(n-r)} - \dfrac{n\sum_{j\varepsilon A} x_j}{r(n-r)}\right) + \beta_{V5}\left(\dfrac{n\sum_{j=1}^{N}(x_i - \bar{X})^2}{(n-r)(N-1)} - \dfrac{n\sum_{j\varepsilon A}(x_i - \bar{x}_r)^2}{(n-r)(r-1)}\right) & if \rightarrow j \in A^c \end{cases}$$

(3.9)

where α_{V5} and β_{V5} are the constant values whose are to be determine. The point estimator of population mean is given as:

$$\bar{y}_{S.SS_V} = \frac{1}{n}\sum_{j=1}^{n}\Delta_{Vj}$$

$$= \frac{1}{n}\left[\sum_{j\varepsilon A} z_j + \sum_{j\varepsilon A^c}\bar{z}_r + \alpha_{V5}\left\{\frac{n\sum_{j\varepsilon A^c}\sum_{j=1}^{N} x_j}{N(n-r)} - \frac{n\sum_{j\varepsilon A^c}\sum_{j\varepsilon A} x_j}{r(n-r)}\right\}\right]$$

$$+ \frac{1}{n}\beta_{V5}\left[\frac{n\sum_{j\varepsilon A^c}\sum_{j=1}^{N}(x_j - \bar{X})^2}{(n-r)(N-1)} - \frac{n\sum_{j\varepsilon A^c}\sum_{j\varepsilon A}(x_j - \bar{x}_r)^2}{(n-r)(r-1)}\right]$$

$$= \frac{1}{n}\left[r\bar{z}_r + (n-r)\bar{z}_r + \alpha_{V5}\left(n\bar{X} - n\bar{x}_r\right) + \beta_{V5}\left\{\frac{n\sum_{j=1}^{N}(x_j - \bar{X})^2}{(N-1)} - \frac{n\sum_{j\varepsilon A}(x_i - \bar{x}_r)^2}{(r-1)}\right\}\right]$$

$$= \frac{1}{n}\left[n\bar{z}_r + \alpha_{V5}\left(n\bar{X} - n\bar{x}_r\right) + \beta_{V5}\{nS_x^2 - ns_{x(r)}^2\}\right]$$

$$\bar{y}_{S.SS_V} = \bar{z}_r + \alpha_{V5}\left(\bar{X} - \bar{x}_r\right) + \beta_{V5}\left(S_x^2 - s_{x(r)}^2\right). \tag{3.10}$$

Expression given in Equation (3.10), can be used when population mean and variance are known.

4. LARGE SAMPLE APPROXIMATION

For evaluating the bias and mean square error of the estimators, we define some useful results as:

Let

$$e_0=\frac{\bar{z}_r}{\bar{Y}}-1, e_1=\frac{\bar{x}_r}{\bar{X}}-1, e_2=\frac{\bar{x}_n}{\bar{X}}-1, e_3=\frac{s^2_{x(r)}}{S^2_x}-1, e_4=\frac{s^2_{x(n)}}{S^2_x}-1, e_5=\frac{s^2_{xz(r)}}{S^2_{xy}}-1,$$

$$E(e_i)=0 \quad \text{for} \quad i=0,1,2,3,4,5$$

To first order of approximation, we have

$$E\left(e_0^2\right)=\frac{1}{\bar{Y}^2 r}\eta\gamma(\mu_s^2+S_s^2)+\theta_{r,N}C_y^2, E\left(e_1^2\right)=\theta_{r,N}C_x^2, E\left(e_2^2\right)=\theta_{n,N}C_x^2,$$

$$E\left(e_3^2\right)=\theta_{r,N}\left(\lambda_{04}-1\right), E\left(e_4^2\right)=\theta_{n,N}\left(\lambda_{04}-1\right), E\left(e_0e_1\right)=\theta_{r,N}\rho_{xy}C_xC_y,$$

$$E\left(e_0e_2\right)=\theta_{n,N}\rho_{xy}C_xC_y, E\left(e_0e_3\right)=\theta_{r,N}C_y\lambda_{12}, E\left(e_0e_4\right)$$
$$=\theta_{n,N}C_y\lambda_{12}, E\left(e_1e_2\right)=\theta_{n,N}C_x{}^2,$$

$$E\left(e_1e_3\right)=\theta_{r,N}C_x\lambda_{03}, E\left(e_1e_4\right)=\theta_{n,N}C_x\lambda_{03}, E\left(e_2e_3\right)$$
$$=\theta_{n,N}C_x\lambda_{03}, E\left(e_2e_4\right)=\theta_{n,N}C_x\lambda_{03},$$

$$E\left(e_3e_4\right)=\theta_{n,N}\left(\lambda_{04}-1\right), E\left(e_1e_5\right)=\theta_{r,N}\rho_{xy}^{-1}C_x\lambda_{12}, E\left(e_2e_5\right)=\theta_{n,N}\rho_{xy}^{-1}C_x\lambda_{12},$$

where

$$\bar{\tau}=\frac{1}{N}\sum_{j=1}^{N}\tau_j, C_\tau^2=\frac{\sigma_\tau^2}{\bar{\tau}^2}, \rho_{\tau\psi}=\frac{S_{\tau\psi}}{S_\tau S_\psi}, R=\frac{\bar{Y}}{\bar{X}}, \theta_r=\frac{1}{r}, \theta_{r,N}=\left(\frac{1}{r}-\frac{1}{N}\right),$$

$$\theta_{n,N}=\left(\frac{1}{n}-\frac{1}{N}\right), \theta_{r,n}=\left(\frac{1}{r}-\frac{1}{n}\right), S_{\tau\psi}=\frac{1}{N-1}\sum_{j=1}^{N}(\tau_j-\bar{\tau})(\psi_j-\bar{\psi}),$$

$$\lambda_{ab}=\frac{\mu_{ab}}{\mu_{20}^{a/2}\mu_{02}^{b/2}}, \mu_{ab}=\frac{1}{N-1}\sum_{j=1}^{N}(y_j-\bar{Y})^a(x_j-\bar{X})^b,$$

where $\tau=x,y$ and $\psi=x,y.$

5. PROPERTIES OF IMPUTATION METHODS

In this section, we define the properties of existing and proposed imputation methods under scrambling response as:

i) From Equation (2.1), the bias and mean square error of $\bar{y}_{S.R}$ is given by:

$$Bias(\bar{y}_{S.R})\cong\theta_{r,n}\bar{Y}\left[C_x^2-\rho_{xy}C_xC_y\right] \tag{5.1}$$

and

$$MSE(\bar{y}_{S.R})\cong\theta_{n,N}S_y^2+\theta_r\eta\gamma(\mu_s^2+S_s^2)+\theta_{r,n}(S_y^2+R^2S_x^2-2RS_{xy}). \tag{5.2}$$

ii) After simplifying the Equation (2.2), The bias and mean square error is given as:

$$Bias(\bar{y})_{S.A_8} \cong \theta_{r,N}\bar{Y}\left(\alpha_1^2 C_x^2 - \alpha_1\rho_{xy}C_xC_y\right) \tag{5.3}$$

and

$$MSE(\bar{y})_{S.A_8} \cong \bar{Y}^2\left(\theta_{r,N}C_y^2 + \theta_{r,n}\alpha_1^2C_x^2 - 2\theta_{r,n}\alpha_1\rho_{xy}C_xC_y\right) + \theta_r\eta\gamma(\mu_s^2 + S_s^2)$$

where $\alpha_1 = \rho_{xy}\dfrac{C_y}{C_x}$,

$$MSE(\bar{y})_{S.A_{8(min.)}} \cong \theta_{r,N}S_y^2 - \theta_{r,n}\frac{S_{xy}^2}{S_x^2} + \theta_r\eta\gamma(\mu_s^2 + S_s^2). \tag{5.4}$$

iii) By solving the Equation (2.3), bias and mean square can be approximated as:

$$Bias(\bar{y})_{S.A_9} \cong \theta_{n,N}\bar{Y}(\alpha_2^2C_x^2 - \alpha_2\rho_{xy}C_xC_y) \tag{5.5}$$

and

$$MSE(\bar{y})_{S.A_9} \cong \theta_{r,N}C_y^2 + \theta_{n,N}\alpha_2^2C_x^2 - 2\theta_{n,N}\alpha_2\rho_{xy}C_xC_y + \theta_r\eta\gamma(\mu_s^2 + S_s^2)$$

where $\alpha_2 = \rho_{xy}\dfrac{C_y}{C_x}$, so

$$MSE(\bar{y})_{S.A_{9(min.)}} \cong \theta_{r,N}S_y^2 - \theta_{n,N}\frac{S_{xy}^2}{S_x^2} + \theta_r\eta\gamma(\mu_s^2 + S_s^2). \tag{5.6}$$

iv) By solving the Equation (2.4), bias and mean square can be approximated as:

$$Bias(\bar{y})_{S.A_{10}} \cong \theta_{r,N}\bar{Y}(\alpha_3^2C_x^2 - \alpha_3\rho C_xC_y) \tag{5.7}$$

and

$$MSE(\bar{y})_{S.A_{10}} \cong \theta_{r,N}C_y^2 + \theta_{r,N}\alpha_3^2C_x^2 - 2\theta_{r,N}\alpha_3\rho C_xC_y + \theta_r\eta\gamma(\mu_s^2 + S_s^2)$$

where $\alpha_3 = \rho_{xy}\dfrac{C_y}{C_x}$, so

$$MSE(\bar{y})_{S.A_{10(min.)}} \cong \theta_{r,N}S_y^2\left\{1 - \frac{S_{xy}^2}{S_x^2}\right\} + \theta_r\eta\gamma(\mu_s^2 + S_s^2). \tag{5.8}$$

v) For Equation (2.5), the mean square error can be computed as:

$$V(\bar{y})_{S.A_5} \cong \theta_{r,N}S_y^2 + \theta_{n,N}w_1^2S_x^2 - 2\theta_{n,N}w_1S_{xy} + \theta_r\eta\gamma(\mu_s^2 + S_s^2) \tag{5.9}$$

where $w_1 = \dfrac{S_{xy}}{S_x^2}$, so

$$V(\bar{y})_{S.A_{5(min.)}} \cong \theta_{r,N}S_y^2 - \theta_{n,N}\rho_{xy}^2 + \theta_r\eta\gamma(\mu_s^2 + S_s^2). \tag{5.10}$$

vi) For Equation (2.6), the mean square error can be computed as:

$$V(\bar{y})_{S.A_6} \cong \theta_{r,N}(S_y^2 + w_2^2 S_x^2 - 2w_2 S_{xy}) + \theta_r \eta\gamma(\mu_s^2 + S_s^2) \quad (5.11)$$

where $w_2 = \dfrac{S_{xy}}{S_x^2}$, so

$$V(\bar{y})_{S.A_{6(min.)}} \cong \theta_{r,N} S^2(1-\rho_{xy}^2) + \theta_r \eta\gamma(\mu_s^2 + S_s^2). \quad (5.12)$$

vii) The bias and mean square error of $\bar{y}_{S.Re.}$ are given as by solving Equation (2.7):

$$Bias \quad (5.13)$$

and

$$MSE(\bar{y}_{S.Re.})_{(min.)} \cong \theta_{n,N} S_y^2 + \theta_r \eta\gamma(\mu_s^2 + S_s^2) + +\theta_{r,n} S_y^2\left[1-\rho_{xy}^2\right]. \quad (5.14)$$

5.1 Modified Imputation Methods

The Equation (3.2) in term of error can be written as:

$$\bar{y}_{S.SS_I} = \frac{\bar{Y}(1+e_0)\bar{X}(1+e_2)S_x^2(1+e_4)}{\left\{\alpha_{I1}\bar{X}(1+e_1)+(1-\alpha_{I1})\bar{X}(1+e_2)\right\}\left\{\beta_{I1}S_x^2(1+e_3)+(1-\beta_{I1})S_x^2(1+e_4)\right\}}$$

$$= \bar{Y}(1+e_0)(1+e_2)(1+e_4)\{1+\alpha_{I1}e_1+(1-\alpha_{I1})e_2\}^{-1}(1+\beta_{I1}e_3+(1-\beta_{I1})e_4)^{-1}$$

$$= \bar{Y}(1+e_0)(1+e_2)(1+e_4)\left\{1-(\alpha_{I1}e_1+(1-\alpha_{I1})e_2)+(\alpha_{I1}e_1+(1-\alpha_{I1})e_2)^2-\cdots\right\}$$
$$\left\{1-(\beta_{I1}e_3+(1-\beta_{I1})e_4)+(\beta_{I1}e_3+(1-\beta_{I1})e_4)^2-\cdots\right\}$$

$$= \bar{Y}(1+e_0)(1+e_2)(1+e_4)\left\{1-(\alpha_{I1}e_1+(1-\alpha_{I1})e_2)+(\alpha_{I1}e_1+(1-\alpha_{I1})e_2)^2-\cdots\right\}$$
$$\left\{1-(\beta_{I1}e_3+\beta_{I1}e_4)+(\beta_{I1}e_3+(1-\beta_{I1})e_4)^2-\cdots\right\}$$

After Simplification, we have

$$= \bar{Y}\left\{1-\beta_{I1}e_3-(1-\beta_{I1})e_4+\beta_{I1}^2e_3^2+(1-\beta_{I1})^2e_4^2+2\beta_{I1}(1-\beta_{I1})e_3e_4-\alpha_{I1}e_1\right.$$
$$+\alpha_{I1}\beta_{I1}e_1e_3+\alpha_{I1}(1-\beta_{I1})e_1e_4-(1-\alpha_{I1})e_2+\beta_{I1}(1-\alpha_{I1})e_2e_3$$
$$+(1-\alpha_{I1})(1-\beta_{I1})e_2e_4+\alpha_{I1}^2e_1^2+(1-\alpha_{I1})^2e_2^2+2\alpha_{I1}(1-\alpha_{I1})e_1e_2$$
$$+e_4-\beta_{I1}e_3e_4-(1-\beta_{I1})e_4^2-\alpha_{I1}e_1e_4-(1-\alpha_{I1})e_2e_4+e_2-\beta_{I1}e_2e_3$$
$$-(1-\beta_{I1})e_2e_4-\alpha_{I1}e_1e_2-(1-\alpha_{I1})e_2^2+e_2e_4+e_0-\beta_{I1}e_0e_3-(1-\beta_{I1})e_0e_4$$
$$\left.-\alpha_{I1}e_0e_1-(1-\alpha_{I1})e_0e_2+e_0e_4\right\}$$

After simplification

$$Bias(\bar{y})_{S.SS_I} \cong \theta_{n,N}\bar{Y}E\left(\alpha_{I1}^2e_2^2-e_0e_2\right)$$
$$+\theta_{r,N}\bar{Y}E\left\{\alpha_{I1}\beta_{I1}e_1e_3-\beta_{I1}e_0e_3+\beta_{I1}^2e_3^2+\alpha_{I1}^2e_1^2+e_0e_1\right\}$$

$$\cong \theta_{n,N}\bar{Y}(\alpha_{I1}^2 C_x^2 - \rho xy C_x C_y)$$
$$+\theta_{r,N}\bar{Y}\left\{\alpha_{I1}\beta_{I1}C_x\lambda_{03} - \beta_{I1}C_y^2\lambda_{12} + \beta_{I1}^2(\lambda_{04}-1) + \alpha_{I1}^2 C_x^2 + \rho_{xy}C_x C_y\right\} \quad (5.15)$$

For mean square error, it can be describe as

$$MSE(\bar{y})_{S.SS_I} \cong \bar{Y}^2 E\left\{e_0 + \alpha_{I1}(e_2 - e_1) + \beta_{I1}(e_4 - e_3)\right\}^2$$
$$\cong \bar{Y}^2 E\left\{e_0^2 + \alpha_{I1}^2(e_2 - e_1)^2 + \beta_{I1}^2(e_4 - e_3)^2\right.$$
$$\left.+2\alpha_{I1}\, e_0(e_2 - e_1) + 2\beta_{I1}e_0(e_4 - e_3) + 2\alpha_{I1}\beta_{I1}(e_2 - e_1)(e_4 - e_3)\right\}$$

After simplification, we have

$$MSE(\bar{y})_{S.SS_I} \cong \theta_{n,N}S_y^2 + \theta_r\eta\gamma(\mu_s^2 + S_s^2) + \theta_{r,n}\bar{Y}^2\left\{C_y^2 + \alpha_{I1}^2 C_x^2 + \beta_{I1}^2(\lambda_{04}-1)\right.$$
$$\left.-2\alpha_{I1}\rho_{xy}C_x C_y - 2\beta_{I1}C_y\lambda_{12} + 2\alpha_{I1}\beta_{I1}C_x\lambda_{03}\right\} \quad (5.16)$$

The optimum values of the unknown constants are determined by minimizing the mean squared error. The optimum value for α_{I1} and β_{I1} is given as

$$\alpha_{I1} = \frac{S_{xy}(\lambda_{04}-1) - S_y\lambda_{12}S_x\lambda_{03}}{RS_x^2(\lambda_{04}-1-\lambda_{03}^2)} \quad \text{and} \quad \beta_{I1} = \frac{S_y\lambda_{12}S_x - S_{xy}\lambda_{03}}{S_x\bar{Y}(\lambda_{04}-1-\lambda_{03}^2)}, \quad (5.17)$$

After substituting the optimum value of α_{II2} and β_{II2} in Equation (5.15), we have

$$MSE(\bar{y})_{S.SS_I(min.)} \cong \theta_{n,N}S_y^2 + \theta_r\eta\gamma(\mu_s^2 + S_s^2)$$
$$+\theta_{r,n}S_y^2\left\{1 - \frac{\lambda_{12}^2 + \rho_{xy}^2(\lambda_{04}-1) - 2\rho_{xy}\lambda_{12}\lambda_{03}}{\lambda_{04}-1-\lambda_{03}^2}\right\}. \quad (5.18)$$

The Equation (3.4) in term of error can be describe as

$$\bar{y}_{S.SS_{II}} = \frac{\bar{Y}(1+e_0)\bar{X}S_x^2}{\left\{\alpha_{II2}\bar{X}(1+e_2) + (1-\alpha_{II2})\bar{X}\right\}\left\{\beta_{II2}S_x^2(1+e_4) + (1-\beta_{II2})S_x^2\right\}}$$

$$= \bar{Y}(1+e_0)(1+\alpha_{II2}e_2)^{-1}(1+\beta_{II2}e_4)^{-1}$$

and to first order approximation

$$\bar{y}_{S.SS_{II}} = \bar{Y}(1+e_0)(1-\alpha_{II2}e_2 + \alpha_{II2}^2 e_2^2 - \cdots)^{-1}(1-\beta_{II2}e_4 + \alpha_{II2}^2 e_4^2 - \cdots)^{-1}$$

$$Bias(\bar{y}_{S.SS_{II}}) \cong \bar{Y}\theta_{n,N}\left\{\alpha_{II2}^2 C_x^2 + \beta_{II2}^2(\lambda_{04}-1)\right.$$
$$\left.+\alpha_{II2}\beta_{II2}C_x\lambda_{03} - \alpha_{II2}\rho_{xy}C_x C_y - \beta_{II2}C_y\lambda_{12}\right\} \quad (5.19)$$

and

$$MSE(\bar{y})_{S.SS_{II}} \cong \bar{Y}^2 E(e_0 - \alpha_{II2}e_2 - \beta_{II2}e_4)^2$$

$$\cong \bar{Y}^2 E\left(e_0^2 + \alpha_{II2}^2 e_2^2 + \beta_{II2}^2 e_4^2 - 2\alpha_{II2}e_0e_2 - 2\beta_{II2}e_0e_4 + 2\alpha_{II2}\beta_{II2}e_2e_4\right)$$

After simplification, we have

$$MSE(\bar{y})_{S.SS_{II}} \cong \theta_r \eta\gamma(\mu_s^2 + S_s^2) + \theta_{r,N}S_y^2 + \theta_{n,N}\bar{Y}^2 \left\{\alpha_{II2}^2 C_x^2 + \beta_{II2}^2(\lambda_{04}-1) - 2\alpha_{II2}\rho_{xy}C_xC_y - 2\beta_{II2}C_y\lambda_{12} + 2\alpha_{II2}\beta_{II2}C_x\lambda_{03}\right\} \quad (5.20)$$

and the optimum values of α_{II2} and β_{II2} are given as

$$\alpha_{II2} = \frac{S_{xy}(\lambda_{04}-1) - S_yS_x\lambda_{12}\lambda_{03}}{RS_x^2(\lambda_{04}-1-\lambda_{03}^2)} \quad \text{and} \quad \beta_{II2} = \frac{S_yS_x\lambda_{12} - S_{xy}\lambda_{03}}{\bar{Y}S_x(\lambda_{04}-1-\lambda_{03}^2)}, \quad (5.21)$$

Putting he optimum values of α_{II2} and β_{II2} from Equation (5.21) put in (5.20), so we have

$$MSE(\bar{y})_{S.SS_{II}(min.)} \cong \theta_{n,N}S_y^2 + \theta_r\eta\gamma\left(\mu_s^2 + S_s^2\right) - \theta_{r,N}S_y^2\left\{\frac{\lambda_{12}^2 + \rho_{xy}^2(\lambda_{04}-1) - 2\rho_{xy}\lambda_{12}\lambda_{03}}{\lambda_{04}-1-\lambda_{03}^2}\right\}. \quad (5.22)$$

The Equation (3.6) in term of error can be defined as

$$\bar{y}_{S.SS_{III}} = \frac{\bar{Y}(1+e_0)\bar{X}S_x^2}{\left\{\alpha_{III3}\bar{X}(1+e_1) + (1-\alpha_{III3})\bar{X}\right\}\left\{\beta_{III3}S_x^2(1+e_3) + (1-\beta_{III3})S_x^2\right\}}$$

$$\bar{y}_{S.SS_{III}} = \bar{Y}(1+e_0)(1+\alpha_{III3}e_1)^{-1}(1+\beta_{III3}e_3)^{-1}$$

and for first order approximation

$$\bar{y}_{S.SS_{III}} = \bar{Y}(1+e_0)\left(1-\alpha_{III3}e_1 + \alpha_{III3}^2e_1^2 - \cdots\right)^{-1}\left(1-\beta_{III3}e_3 + \alpha_{III3}^2e_3^2 - \cdots\right)^{-1}$$

$$Bias(\bar{y}_{S.SS_{III}}) \cong \theta_{r,N}\bar{Y}\left\{\alpha_{III3}^2C_x^2 + \beta_{III3}^2(\lambda_{04}-1) + \alpha_{III3}\beta_{III3}C_x\lambda_{03} - \alpha_{III3}\rho_{xy}C_xC_y - \beta_{III3}C_y\lambda_{12}\right\} \quad (5.23)$$

and

$$MSE(\bar{y})_{S.SS_{III}} \cong \bar{Y}^2E(e_0 - \alpha_{III3}e_1 - \beta_{III3}e_3)^2$$

$$\cong \bar{Y}^2E(e_0^2 + \alpha_{III3}^2e_1^2 + \beta_{III3}^2e_3^2 - 2\alpha_{III3}e_0e_1 - 2\beta_{III3}e_0e_3 + 2\alpha_{III3}\beta_{III3}e_1e_3)$$

After simplification, we have

$$MSE(\bar{y})_{S.SS_{III}} \cong \theta_r \eta\gamma(\mu_s^2 + S_s^2) + \theta_{r,N}\bar{Y}^2 \left\{C_y^2 + \alpha_{III3}^2 C_x^2 + \beta_{III3}^2(\lambda_{04}-1) \right.$$
$$\left. -2\alpha_{III3}\rho_{xy}C_xC_y - 2\beta_{III3}C_y\lambda_{12} + 2\alpha_{III3}\beta_{III}C_x\lambda_{03}\right\} \quad (5.24)$$

and the optimum values of α_{II2} and β_{II2} are given as

$$\alpha_{III3} = \frac{S_{xy}(\lambda_{04}-1) - S_yS_x\lambda_{12}\lambda_{03}}{RS_x^2(\lambda_{04}-1-\lambda_{03}^2)} \quad \text{and} \quad \beta_{III3} = \frac{S_yS_x\lambda_{12} - S_{xy}\lambda_{03}}{\bar{Y}S_x(\lambda_{04}-1-\lambda_{03}^2)}, \quad (5.25)$$

substituting the Equation (5.25) in (5.24), so

$$MSE(\bar{y})_{S.SS_{III}(min.)} \cong \theta_r\eta\gamma(\mu_s^2 + S_s^2) + \theta_{r,N}S_y^2\left\{1 - \frac{\lambda_{12}^2 + \rho_{xy}^2(\lambda_{04}-1) - 2\rho_{xy}\lambda_{12}\lambda_{03}}{\lambda_{04}-1-\lambda_{03}^2}\right\}. \quad (5.26)$$

The variance of $(\bar{y})_{S.SS_{IV}}$ in terms of error is defined as

$$V(\bar{y})_{S.SS_{IV}} \cong E\left[\bar{Y}^2e_0 - \alpha_{IV4}\bar{X}e_2 - \beta_{IV4}S_x^2e_4\right]^2$$
$$\cong E\left(\bar{Y}^2e_0^2 + \alpha_{IV4}^2\bar{X}^2e_2^2 + \beta_{IV4}^2S_x^4e_4^2 - 2\alpha_{IV4}\overline{YX}e_0e_2 \right.$$
$$\left. -2\beta_{IV4}\bar{Y}S_x^2e_0e_4 + 2\alpha_{IV4}\beta_{IV4}\bar{X}S_x^2e_1e_3\right)$$

After simplification, we have

$$V(\bar{y})_{S.SS_{IV}} \cong \theta_r\eta\gamma(\mu_s^2 + S_s^2) + \theta_{r,N}S_y^2 + \theta_{n,N}\left\{\alpha_{IV4}S_x^2 + \beta_{IV4}^2S_x^4(\lambda_{04}-1)\right.$$
$$\left. -2\alpha_{IV4}S_{xy} - 2\beta_{IV4}S_x^2S_y\lambda_{12} + 2\alpha_{IV4}\beta_{IV4}S_x^3\lambda_{03}\right\} \quad (5.27)$$

optimum values of α_{IV4} and β_{IV4} are given as

$$\alpha_{IV4} = \frac{S_{xy}(\lambda_{04}-1) - S_yS_x\lambda_{12}\lambda_{03}}{S_x^2(\lambda_{04}-1-\lambda_{03}^2)} \quad \text{and} \quad \beta_{IV4} = \frac{S_yS_x\lambda_{12} - S_{xy}\lambda_{03}}{S_x^3(\lambda_{04}-1-\lambda_{03}^2)}, \quad (5.28)$$

After simplification, we have

$$V(\bar{y})_{S.SS_{IV(min.)}} \cong \theta_r\eta\gamma(\mu_s^2 + S_s^2) + \theta_{r,N}S_y^2 - \theta_{n,N}S_y^2\left\{\rho_{xy}^2 + \frac{(\lambda_{12} - \rho_{xy}\lambda_{03})^2}{\lambda_{04}-1-\lambda_{03}^2}\right\}. \quad (5.29)$$

The variance of $\bar{y}_{S.SS_V}$ in term of error is defined as

$$V(\bar{y})_{S.SS_V} \cong E\left[\bar{Y}^2e_0 + \alpha_{V5}\left\{\bar{X} - \bar{X}(1+e_1)\right\} + \beta_{V5}\left\{S_x^2 - S_x^2(1+e_3)\right\}\right]^2$$
$$\cong E\left[\bar{Y}^2e_0 - \alpha_{V5}\bar{X}e_1 - \beta_{V5}S_x^2e_3\right]^2$$

$$\cong E\left[\begin{array}{r}\bar{Y}^2 e_0^2 + \alpha_{V5}^2 \bar{X}^2 e_1^2 + \beta_{V5}^2 S_x^4 e_3^2 - 2\alpha_{V5}\bar{Y}\bar{X}e_0 e_1 \\ -2\beta_{V5}\bar{Y}S_x^2 e_0 e_3 + 2\alpha_{V5}\beta_{V5}\bar{X}S_x^2 e_1 e_3\end{array}\right]$$

After simplification, we have

$$V(\bar{y})_{S.SS_V} \cong \theta_r \eta\gamma(\mu_s^2 + S_s^2) + \theta_{r,N}\left\{S_y^2 + \alpha_{V5}S_x^2 + \beta^2 S_x^2(\lambda_{04} - 1) - 2\alpha_{V5}S_{xy} - 2\beta_{V5}S_x^2 S_y \lambda_{12} + 2\alpha_{V5}\beta_{V5}S_x^3\lambda_{03}\right\} \quad (5.30)$$

optimum values of α_{V5} and β_{V5} are given as

$$\alpha_{V5} = \frac{S_{xy}(\lambda_{04}-1) - S_y S_x \lambda_{12}\lambda_{03}}{S_x^2(\lambda_{04} - 1 - \lambda_{03}^2)} \quad \text{and} \quad \beta_{V5} = \frac{S_y S_x \lambda_{12} - S_{xy}\lambda_{03}}{S_x^3(\lambda_{04} - 1 - \lambda_{03}^2)}, \quad (5.31)$$

After simplification, we have

$$V(\bar{y})_{S.SS_{V(min.)}} \cong \theta_r \eta\gamma(\mu_s^2 + S_s^2) + \theta_{r,N} S_y^2 \left\{1 - \rho_{xy}^2 - \frac{(\lambda_{12} - \rho_{xy}\lambda_{03})^2}{\lambda_{04} - 1 - \lambda_{03}^2}\right\}. \quad (5.32)$$

5.2 Theoretical Comparison

In this section, we considered the theoretical comparison of the proposed method with the mean imputation method is given as

i) $V(\bar{y}) - \left[MSE(\bar{y})_{S.SS_{I(min.)}} \text{ or } MSE(\bar{y})_{S.SS_{III(min.)}}\right] > 0$

$$\lambda_{12}^2 + \rho_{xy}(\lambda_{04} - 1) - 2\rho_{xy}\lambda_{12}\lambda_{03} > 0 \quad (5.33)$$

ii) $V(\bar{y}) - MSE(\bar{y})_{S.SS_{II(min.)}} > 0$

$$\theta_{nN} + \theta_{rN}\frac{\lambda_{12}^2 + \rho_{xy}(\lambda_{04} - 1) - 2\rho_{xy}\lambda_{12}\lambda_{03}}{\lambda_{04} - 1 - \lambda_{03}^2} > 0 \quad (5.34)$$

iv) $V(\bar{y}) - \left[MSE(\bar{y})_{S.SS_{IV(min.)}} \quad \text{or} \quad MSE(\bar{y})_{S.SS_{V(min.)}}\right] > 0$

$$\rho_{xy}^2 + \frac{(\lambda_{12} - \rho xy\lambda_{03})^2}{\lambda_{04} - 1 - \lambda_{03}^2} > 0 \quad (5.35)$$

Our study plan is perform better as compare to the mean imputation method, if the above mentioncondition are satisfied. Further their performance can be illustrated numerically in the next section.

6. APPLICATION

The numerically comparison of the proposed imputation method is considered by used the simulated and real life data set as:

6.1 Empirical Study

For the empirical study, we consider the following steps as follow:

Step 1:

We generate 1000 random numbers using bivariate normal distribution. A simple random sample of size 100 is selected from the population of 1000 units and suppose that 70 out of 100 samples can provide the response and we want to impute the values for the remaining units who does not provide the response.

$$\bar{Y} = 2.6968, \bar{X} = 9.9391, S_y^2 = 0.75109, S_x^2 = 8.7257001, S_{xy} = 1.93743, \rho_{xy} = 0.7499, \gamma = 0.02,, \beta_{IV4} = \beta_{V5} = -0.0131$$

$$\eta = 0.04, \alpha_1 = \alpha_2 = \alpha_3 = 0.8363, \beta_1 = \beta_2 = 0.22204, \alpha_{I1} = \alpha_{II2} = \alpha_{III3} = 0.8901, \alpha_{IV4} = \alpha_{V5} = -0.0146,$$

$$\beta_{I1} = \beta_{II2} = \beta_{III3} = 0.2363.$$

Step 2:

We repeat the process of selection of the sample 50000 times. Thus we obtain the 50000 values of the $\bar{y}_k$.

Step 3:

For bias and mean square error, we used the follow expression:

$$Bias(\bar{y}_k) = \frac{\sum_{g=1}^{50000} \bar{y}_{kg}}{50000} - \bar{Y} \tag{6.1}$$

and for mean square error, we have

$$MSE(\bar{y}_k) = \frac{\sum_{g=1}^{50000} (\bar{y}_{kg} - \bar{Y})^2}{50000} \tag{6.2}$$

For accessing the performance of existing and proposed estimators with respect to mean, we consider the following expression as:

$$P.R.E(i) = \frac{MSE(\bar{y})_{S.M}}{M.S.E(\bar{y}_k)} \times 100 \tag{6.3}$$

where

$i = 1, 2, 3, \cdots, 13$ and

$k = S.R, S.Re, S.A_6, S.SS_V, S.A_8, S.SS_I, S.A_9, S.SS_{II}, S.A_5, S.SS_{IV}, S.A_{10}, S.SS_{III}$.

Table 1
Bias, MSE and Percentage Relative Efficiency (i).

Estimators	Bias	MSE	Efficiency %
Existing Imputation Methods			
Mean Imputation Method			
$\bar{y}_{S.M}$	-0.00072	0.017979	100.0000
Ratio Imputation Methods			
$\bar{y}_{S.R}$	0.000226	0.012201	147.3478
$\bar{y}_{S.A_8}$	-0.001100	0.012000	150.1932
$\bar{y}_{S.A_9}$	-0.000971	0.013649	131.7162
$\bar{y}_{S.A_{10}}$	-0.000625	0.007644	235.1844
Regression Imputation Methods			
$\bar{y}_{S.SS_5}$	-0.001096	0.013787	130.4055
$\bar{y}_{S.SS_6}$	-0.000959	0.013652	131.6901
$\bar{y}_{S.Re.}$	-0.000619	0.007680	234.0948
Proposed Imputation Methods			
Ratio Imputation Methods			
$\bar{y}_{S.SS_I}$	-0.000604	0.011877	151.3747
$\bar{y}_{S.SS_{II}}$	-0.000960	0.013580	132.3855
$\bar{y}_{S.SS_{III}}$	-0.000842	0.007486	240.1635
Regression Imputation Methods			
$\bar{y}_{S.SS_{IV}}$	-0.001189	0.013590	132.2883
$\bar{y}_{S.SS_V}$	-0.001991	0.007600	237.8263

In Table 1, we easily understand that, the bias, mean square error and percentage relative efficiencies of existing and proposed estimators. In column 2, the value of bias of all the estimators is expressed in front of their respective estimator. The value of bias of the $\bar{y}_{S.M}$, is minimum as compare to all other estimators. In column 2 and 3, the mean square error and percentage relative efficiencies are expressed. The mean square error of the ratio estimator is 0.012201 (147.3478 %), for $\bar{y}_{S.A_8}$ is 0.012000 (150.1982 %), for $\bar{y}_{S.A_9}$ is 0.013649 (131.7162.96 %), for $\bar{y}_{S.A_{10}}$ the value is 0.007644 (235.1844 %) and for the $\bar{y}_{S.A_5}$, the value of mean square error is 0.013787 (130.4055 %), for the estimator $\bar{y}_{S.A_6}$ 0.013652(131.6901 %) and for regression estimator the value is 0.007680 (234.0948 %). The value of mean square error for the proposed imputation methods is, for $\bar{y}_{S.SS_I}$ is 0.011877 (151.3747 %), for $\bar{y}_{S.SS_{II}}$ is 0.013580 (132.3855 %), for $\bar{y}_{S.SS_{III}}$

the value is 0.007486(240.1635 %), for $\bar{y}_{S.SS_{IV}}$ is 0.013590 (132.2883 %) and for $\bar{y}_{S.SS_V}$ is 0.007600 (237.8263 %). The quantity inside the braces shows the amount of percentage relative efficiency relative to mean imputation method. If correlation between the both the study and the auxiliary variable is high than our proposed class of imputation methods can perform better as compare to other existing estimator. The fifth method imputation can perform outstanding as compare to other methods, but other methods of imputation are also perform relatively better as compare to mean imputation method.

For the other values of the parameters, another simulation study is considered under the procedure which was described above at the varying response. The values of the parameter are given as

$$\bar{Y} = 2.6368, \bar{X} = 9.9311, S_y^2 = 0.7517915,$$
$$S_x^2 = 8.725733, S_{xy} = 1.9375, \rho_{xy} = 0.7565,$$
$$\gamma = 0.02, \eta = 0.04, \alpha_1 = \alpha_2 = \alpha_3 = 0.2220, \beta_1 = \beta_2 = 0.2220,$$
$$\alpha_{I1} = \alpha_{II2} = \alpha_{III3} = 0.8362, \alpha_{IV4} = \alpha_{V5} = 0.89001,$$
$$\beta_{I1} = \beta_{II2} = \beta_{III3} = -0.0145, \beta_{IV4} = \beta_{V5} = -0.0145$$

Table 2
Bias, MSE and Percentage Relative Efficiency(i) for Various Response Rate

	r	Bias	MSE	Efficiency%
Existing Imputation Method				
Mean Imputation Method				
$\bar{y}_{S.M}$	20	-0.0012	0.0371	100.0000
		Ratio Imputation Method		
$\bar{y}_{S.R}$		0.0036	0.0198	187.8650
$\bar{y}_{S.A_8}$		0.0015	0.0192	193.4406
$\bar{y}_{S.A_9}$		-0.0016	0.0327	113.7307
$\bar{y}_{S.A_{10}}$		0.0012	0.0155	239.0239
		Regression Imputation Method		
$\bar{y}_{S.A_5}$		0.0066	0.0207	179.4520
$\bar{y}_{S.A_6}$		-0.0014	0.0326	113.0605
$\bar{y}_{S.Re}$		0.0013	0.0155	240.2402

	r	**Bias**	**MSE**	**Efficiency%**
Proposed Imputation Method				
Ratio Imputation Method				
$\bar{y}_{S.I}$		0.0018	0.0189	196.8858
$\bar{y}_{S.II}$		-0.0014	0.0327	113.7349
$\bar{y}_{S.III}$		0.0018	0.0152	244.6133
Regression Imputation Method				
$\bar{y}_{S.IV}$		-0.0009	0.0328	113.2036
$\bar{y}_{S.V}$		0.0012	0.0151	246.0172
Existing Imputation Method				
Mean Imputation Method				
$\bar{y}_{S.M}$	40	0.0036	0.0189	100.0000
Ratio Imputation Method				
$\bar{y}_{S.R}$		-0.0023	0.0115	164.8983
$\bar{y}_{S.A_8}$		-0.0018	0.0114	165.4402
$\bar{y}_{S.A_9}$		0.0029	0.0145	130.2225
$\bar{y}_{S.A_{10}}$		-0.0024	0.0075	251.7009
Regression Imputation Method				
$\bar{y}_{S.A_5}$		0.0002	0.0116	163.0143
$\bar{y}_{S.A_6}$		0.0031	0.0146	129.7876
$\bar{y}_{S.Re}$		-0.0023	0.0076	248.8106
Proposed Imputation Method				
Ratio Imputation Method				
$\bar{y}_{S.I}$		-0.0017	0.0112	168.4670
$\bar{y}_{S.II}$		0.0029	0.0145	130.8509
$\bar{y}_{S.III}$		-0.0023	0.0073	259.6287
Regression Imputation Method				
$\bar{y}_{S.IV}$		0.0029	0.0146	129.8388
$\bar{y}_{S.V}$		-0.0020	0.0074	256.9631

	r	Bias	MSE	Efficiency%
Existing Imputation Method				
Mean Imputation Method				
$\bar{y}_{S.M}$	70	0.0008	0.0098	100.0000
Ratio Imputation Method				
$\bar{y}_{S.R}$		0.0019	0.0083	118.9766
$\bar{y}_{S.A_8}$		0.0016	0.0081	120.6622
$\bar{y}_{S.A_9}$		0.0037	0.0059	167.0542
$\bar{y}_{S.A_{10}}$		0.0045	0.0044	225.2390
Regression Imputation Method				
$\bar{y}_{S.A_5}$		0.0022	0.0082	119.6638
$\bar{y}_{S.A_6}$		0.0038	0.0059	167.0073
$\bar{y}_{S.Re}$		0.0045	0.0044	224.9431
Proposed Imputation Method				
Ratio Imputation Method				
$\bar{y}_{S.I}$		0.0014	0.0081	121.2890
$\bar{y}_{S.II}$		0.0040	0.0058	169.9483
$\bar{y}_{S.III}$		0.0047	0.0043	229.4913
Regression Imputation Method				
$\bar{y}_{S.IV}$		0.0041	0.0057	170.8463
$\bar{y}_{S.V}$		0.0040	0.0043	229.3880

In Table 2, we carried out the simulation study over the various response rate to access out the performance of the existence estimators. It is noticed that our proposed class of estimators is perform better as compare than their counterpart.

6.2 Real Life Application

For the real life application, we use the data set of FEV.DAT which was attach with the text Rosner (2015). For this research work we consider the *FEVstatus* and *age* of the children as the study and auxiliary variable respectively. From the population of 654 children, we select the sample of size 65 units for this study. The data description is given in Table 2.

Table 3
Discriptive Statitsics of Variables (FEV and Age).

Variable	**Corr.**	**Mini.**	1st Qu.	**Median**	**Mean**	3st Qu.	**Max.**	**St.Dev.**
FEV	0.756	0.791	1.981	2.548	2.637	3.118	5.793	0.867
Age		3.000	8.000	10.000	9.931	12.000	19.000	2.954

In Table 3, the correlation value between between the study and auxiliary variable is 0.756. The minimum and maximum values of the study and the auxiliary variable is 0.791 and 5.793, 3 and 19 respectively. The mean and standard deviation of the both variables is 2.637 and 0.867, 0.867 and 2.964 respectively.

For accessing the performance of existing and proposed imputation methods, we use the expression which is given in Equation (6.3) with different values of $\mu_s^2, S_s^2, \eta, \gamma$ and r.

$$\mu_s^2 = 0.100, 0.010, 0.900, 0.090, 0.000. S_s^2 = 0.020, 0.200, 0.400, 0.040, 0.5.$$

$$\eta = 0.090, 0.120, 0.150, 0.180, 0.210. \gamma = 0.100, 0.080, 0.060, 0.040 \, and \, 0.020.$$

In Table 4 (see Appendix): we shows the $PRE(i)$ over the different values of μ_s^2, S_s^2, η and γ at varying response r $(23.08\% - 92.31\%)$ by using the real life data set. The maximum $P.R.E(i)$ of the k estimator is 176.272%, 130.247%, 233.373%, 252.865%, 181.797%, 190.818%,113.839%, 114.757 %,113.839%, 114.757%, 233.373% and 252.865% respectively. In entire combination of the μ_s^2, S_s^2, η and γ with r our proposed estimator can perform better as compare to the existing related estimators.

7. CONCLUSION

Based on the results of Table 1,2 and 4 , we observed that, our proposed imputation methods can perform better as compare to all of the existing imputation methods. Furthermore, we notice that the estimators, (namely $\bar{y}_{S.SS_{III}}$, $\bar{y}_{S.SS_V}$) have the higher level of efficiency as compare to other proposed imputation methods. Overall by the use of higher order moments of the auxiliary variable would lead to improve the efficiency of imputation methods.

In Table 4, we see that by the use of higher order moments of the auxiliary variable over the different values of μ_s^2, S_s^2, γ and η provide more precise results as compare to the existing estimators. From our small scale study, we concluded that the proposed estimators can perform better as compare to the existing estimator.

We see by comparison of Table 2 and Table 4, the behavior of proposed class of estimators in empirical study and in real life data set is quite similar but we see some of the fluctuation in results of empirical study and real life data because in simulation study our results are based on some kind of imaginations but in real life these imaginations may or may not be acceptable, so their is slightly variability exists. As a whole, our proposed class of estimators under simulated and empirical study can performed better as compare to existing estimators.

ACKNOWLEDGEMENT

The authors are thankful to the referees for their suggestions, which led to the improvement of this article.

REFERENCES

1. Ahmed, M., Al-Titi, O., Al-Rawi, Z. and Abu-Dayyeh, W. (2006). Estimation of a population mean using different imputation methods. *Statistics in Transition*, 7, 6: 1247, 1264.
2. Gjestvang, C.R. and Singh, S. (2009). An improved randomized response model: estimation of mean. *Journal of Applied Statistics*, 36(12), 1361-1367.
3. Hansen, M.H. and Hurwitz, W.N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41(236), 517-529.
4. Mohamed, C., Sedory, S. and Singh, S. (2016). Comparison of Different Imputing Methods for Scrambled Responses. *Handbook of Statistics*, 34, 471-495.
5. Rosner, B. (2015). *Fundamentals of Biostatistics*, Nelson Education.
6. Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.
7. Singh, S. and Deo, B. (2003). Imputation by power transformation. *Statistical Papers*, 44(4), 555-579.
8. Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309), 63-69.

APPENDIX

Table 4: P.R.E(i) **Values under Different Situations**

r	RR	μ_s^2	S_s^2	η	γ	P.R.E (1)	P.R.E (2)	P.R.E (3)	P.R.E (4)	P.R.E (5)	P.R.E (6)	P.R.E (7)	P.R.E (8)	P.R.E (9)	P.R.E(10)	P.R.E(11)	P.R.E(12)
15	0.049	0.100	0.020	0.090	0.100	176.246	130.239	233.312	252.790	181.768	190.784	113.836	114.753	113.836	114.753	233.312	252.790
30	0.024	0.100	0.020	0.090	0.1000	144.948	119.966	233.301	252.776	147.591	151.765	133.141	135.741	133.141	135.741	233.301	252.776
45	0.016	0.100	0.020	0.090	0.100	122.183	110.796	233.290	252.762	123.273	124.953	161.959	167.974	161.959	167.974	233.290	252.762
60	0.011	0.100	0.020	0.090	0.100	104.880	102.561	233.278	252.747	105.085	105.395	209.617	223.795	209.617	223.795	233.278	252.747
15	0.049	0.010	0.200	0.120	0.080	176.076	130.189	232.918	252.300	181.579	190.565	113.816	114.732	113.816	114.732	232.918	252.300
30	0.024	0.010	0.200	0.120	0.080	144.863	119.935	232.897	252.275	147.500	151.663	133.084	135.678	133.084	135.678	232.897	252.275
45	0.016	0.010	0.200	0.120	0.080	122.147	110.780	232.876	252.248	123.234	124.912	161.825	167.822	161.825	167.822	232.876	252.248
60	0.011	0.010	0.200	0.120	0.060	104.873	102.558	232.854	252.220	105.077	105.387	209.303	223.417	209.303	223.417	232.854	252.220
15	0.050	0.900	0.400	0.150	0.060	174.354	129.681	228.968	247.410	179.677	188.351	113.612	114.513	113.612	114.513	228.968	247.410
30	0.024	0.900	0.400	0.150	0.060	144.005	119.618	228.857	247.273	146.575	150.630	132.502	135.038	132.502	135.038	228.857	247.273
45	0.016	0.900	0.400	0.150	0.060	121.780	110.618	228.741	247.129	122.846	124.490	160.475	166.287	160.475	166.287	228.741	247.129
60	0.012	0.900	0.400	0.150	0.060	104.802	102.521	228.619	246.978	105.003	105.308	206.161	219.638	206.161	219.638	228.619	246.978
15	0.049	0.090	0.040	0.180	0.040	176.272	130.247	233.373	252.865	181.797	190.818	113.839	114.757	113.839	114.757	233.373	252.865
30	0.024	0.090	0.040	0.180	0.040	144.961	119.971	233.364	252.854	147.605	151.781	133.150	135.751	133.150	135.751	233.364	252.854
45	0.016	0.090	0.040	0.180	0.040	122.189	110.799	233.354	252.841	123.278	124.960	161.979	167.997	161.979	167.997	233.354	252.841
60	0.011	0.090	0.040	0.180	0.040	104.882	102.562	233.343	252.828	105.086	105.396	209.665	223.854	209.665	223.854	233.343	252.828
15	0.049	0.000	0.500	0.210	0.020	176.060	130.185	232.883	252.256	181.563	190.545	113.814	114.730	113.814	114.730	232.883	252.256
30	0.024	0.000	0.500	0.210	0.020	144.856	119.932	232.861	252.23	147.492	151.654	133.079	135.673	133.079	135.673	232.861	252.230
45	0.016	0.000	0.500	0.210	0.020	122.144	110.779	232.839	252.202	123.231	124.908	161.813	167.808	161.813	167.808	232.839	252.202
60	0.011	0.000	0.500	0.210	0.020	104.873	102.558	232.816	252.173	105.077	105.386	209.275	223.384	209.275	223.384	232.816	252.173