

An Improved Estimator of the Population Mean under Double Sampling using Median

R. E.Kurbah¹ and P.Khongji²

¹Department of Statistics, Shillong College,
Shillong, Meghalaya, India
Email : ronaldkurbah@gmail.com

²Basic Sciences & Social Sciences Department,
North-Eastern Hill University, Shillong Meghalaya, India
Email : phrang2000@yahoo.com

ABSTRACT

It is a well-known fact that the utilization of auxiliary information for estimating a parameter of a study variable enhances the efficiency of the estimator. In this paper, we have considered a regression estimator under doubling sampling involving two auxiliary variables which makes use of the sample mean and sample median of the auxiliary variables of both phases. It is found that the proposed estimator is more efficient than the one proposed by Vadlamudi (2017) with one auxiliary variable. The study was tested empirically using simulated data. The study is also generalized for $k(>1)$ variables.

Key Words:Sampling; Auxiliary variable; Regression estimator; Double Sampling;Median estimation.

Mathematics Subject Classification: 62D05

Journal of Economic Literature (JEL) Classification : C83

1. INTRODUCTION

The aim of the paper is to obtain a more efficient regression estimator of the population mean in double sampling, using information on both mean and median of more than one auxiliary variable. Median estimation have been initially studied by Gross(1980), Kuk and Mak(1989), etc. However, the usage of both known population mean and known population median of the auxiliary variable using regression estimator was put forward by R.Lamichhane, S. Singh, N. Diawara (2015). The study of the same estimator was taken further by K.A. Vadlamudi, S.A. Sedory, S. Singh(2017) by introducing the estimator under double sampling scheme. In this paper we further propose the estimator under double sampling scheme with two auxiliary variables.

2. CONCEPTS, DEFINITIONS AND PREVIOUS WORKS

Let us consider a population of size N . Let the variable under study be y having population values $Y_1, Y_2, \dots, Y_N$. Similarly, let the two auxiliary variables under consideration be x and z with population values $X_1, X_2, \dots, X_N$ and $Z_1, Z_2, \dots, Z_N$ respectively. The study variable y is taken to be highly correlated with auxiliary variable x as well as auxiliary variable z , whereas, the two auxiliary variables x and z are to be independent of each other.

Now, as double sampling is to be implemented, firstly, a large preliminary sample (or first phase sample) of size m is drawn from the population of size N using Simple Random Sampling Without Replacement (SRSWOR) measuring the values of x and z respectively according to the population unit numbers selected viz., x_r and z_r ($r=1, 2, \dots, m$).

Next, a smaller sample (or second phase sample) of size n is drawn from the previous first phase sample of size m using Simple Random Sampling Without Replacement (SRSWOR) measuring the values of y , x and z respectively according to the population unit numbers selected viz., y_r , x_r and z_r ($r=1, 2, \dots, n$).

Let M_x and M_z be the population medians of variables x and z respectively. The development in estimation with median as estimators and using the auxiliary information as acknowledged by many researchers was made by Kuk and Mak in 1989. Therefore, in our approach we implemented the same concept by applying the various sample medians in our estimator for a finite population. The study also assume sample asymptotic properties of sample median with a defined asymptotic variance (Gross, 1980).

We further define $\hat{M}_x^*$ and $\hat{M}_z^*$ as the sample medians of x and z respectively based on the first phase samples of size m . Similarly, we define $\hat{M}_x$ and $\hat{M}_z$ as the sample medians of x and z respectively based on the second phase sample of size n .

The estimator proposed by R. Lamichhane, S. Singh, N. Diawara (2015) which makes use of known population mean and known population median of the single auxiliary variable x is given by,

$$\bar{y}_c = \bar{y} + \beta_1(\bar{X} - \bar{x}) + \beta_2(M_x - \hat{M}_x^*) \quad (2.11)$$

where β_1 and β_2 are to be estimated such that the variance of $\bar{y}_c$ is minimum.

The minimum variance of this estimator is given as follows,

$$Min.Var(\bar{y}_c) = Var(\bar{y}) \left[1 - R_{\frac{\bar{x}}{y \bar{x}} \hat{M}_x}^2 \right] \quad (2.12)$$

where,

$$R^2_{\hat{y}_x \hat{M}_x} = \frac{\rho_{\hat{y}_x \hat{M}_x}^2 - 2\rho_{\hat{y}_x \hat{M}_x} \rho_{\hat{y}_x \hat{M}_x} + \rho_{\hat{y}_x \hat{M}_x}^2}{1 - \rho_{\hat{y}_x \hat{M}_x}^2}$$

is the multiple R^2 such that $0 \leq R^2 \leq 1$

Next, K.A. Vadlamudi, S.A. Sedory, S.Singh (2017) modified the above estimator by introducing it under double sampling and proposed the following estimator given by,

$$\bar{y}_{kal} = \bar{y}_n + \beta_1^* (\bar{x}_m - \bar{x}_n) + \beta_2^* (\hat{M}_x^* - \hat{M}_x) \quad (2.13)$$

where β_1^* and β_2^* are unknown partial regression coefficients to be determined such that the variance of the estimator is minimum. The minimum variance of this estimator is given as follows,

$$Min.Var(\bar{y}_{kal}) = \left(\frac{1}{m} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{n} - \frac{1}{m} \right) S_y^2 \left\{ 1 - \rho_{yx}^2 - \frac{(S_x^2 S_{YM_x} - S_{xy} S_{XM_x})^2}{S_y^2 S_x^2 (0.25 S_x^2 - S_{XM_x}^2)} \right\} \quad (2.14)$$

3. PROPOSED ESTIMATOR

The new estimator of the population mean $\bar{Y}$ under double sampling with two auxiliary variables is defined by,

$$\bar{y}_{REK} = \bar{y}_n + \beta_1^* (\bar{x}_m - \bar{x}_n) + \beta_2^* (\hat{M}_x^* - \hat{M}_x) + \beta_3^* (\bar{z}_m - \bar{z}_n) + \beta_4^* (\hat{M}_z^* - \hat{M}_z) \quad (3.1)$$

where $\beta_1^*, \beta_2^*, \beta_3^*$ and β_4^* are unknown partial regression coefficients to be determined such that the variance of the estimator $\bar{y}_{REK}$ is minimum. Since this estimator makes use of the medians of the two auxiliary variables, then, the same assumption as given by R.Lamichhane, S. Singh, N. Diawara (2015) that the distributions of Y , X and Z are non-normal and follow a skewed distribution is also adopted here so that their corresponding means and medians do not coincide.

Now, in order to show that the proposed estimator is unbiased and also to determine its minimum variance we define the following error functions:

$$e_0 = \left(\frac{\bar{y}_n - \bar{Y}}{\bar{Y}} \right), \quad e_1 = \left(\frac{\bar{x}_n - \bar{X}}{\bar{X}} \right), \quad e_2 = \left(\frac{\hat{M}_x - M_x}{M_x} \right) \quad (\text{Based on 2nd phase sample}) \quad (3.2)$$

$$e_3 = \left(\frac{\bar{x}_m - \bar{X}}{\bar{X}} \right), \quad e_4 = \left(\frac{\hat{M}_x^* - M_x}{M_x} \right) \quad (\text{Based on 1st phase sample}) \quad (3.3)$$

$$e_5 = \left(\frac{\bar{z}_n - \bar{Z}}{\bar{Z}} \right), \quad e_6 = \left(\frac{\hat{M}_z^* - M_z}{M_z} \right) \quad (\text{Based on 2nd phase sample}) \quad (3.4)$$

$$e_7 = \left(\frac{\bar{z}_m - \bar{Z}}{\bar{Z}} \right), \quad e_8 = \left(\frac{\hat{M}_z^* - M_z}{M_z} \right) \quad (\text{Based on 2nd phase sample}) \quad (3.5)$$

Here, we obtained their expectations as follows,

$$E(e_0) = E(e_1) = E(e_2) = E(e_5) = E(e_6) = 0$$

$$E(e_3) = E(e_4) = E(e_7) = E(e_8) \approx 0$$

$$E(e_0^2) = \left(\frac{1}{n} - \frac{1}{N} \right) C_y^2, \quad E(e_1^2) = \left(\frac{1}{n} - \frac{1}{N} \right) C_x^2, \quad E(e_2^2) = \left(\frac{1}{n} - \frac{1}{N} \right) \frac{1}{4[f_x(M_x)]^2 M_x^2}$$

$$E(e_3^2) = \left(\frac{1}{m} - \frac{1}{N} \right) C_x^2, \quad E(e_4^2) = \left(\frac{1}{m} - \frac{1}{N} \right) \frac{1}{4[f_x(M_x)]^2 M_x^2}$$

$$E(e_5^2) = \left(\frac{1}{n} - \frac{1}{N} \right) C_z^2, \quad E(e_6^2) = \left(\frac{1}{n} - \frac{1}{N} \right) \frac{1}{4[f_z(M_z)]^2 M_z^2}$$

$$E(e_7^2) = \left(\frac{1}{m} - \frac{1}{N} \right) C_z^2, \quad E(e_8^2) = \left(\frac{1}{m} - \frac{1}{N} \right) \frac{1}{4[f_z(M_z)]^2 M_z^2}$$

$$E(e_0 e_1) = \left(\frac{1}{n} - \frac{1}{N} \right) \rho_{xy} C_x C_y, \quad E(e_0 e_2) = - \left(\frac{1}{n} - \frac{1}{N} \right) \frac{S_{yM_x} [f_x(M_x)]^{-1}}{\bar{Y} M_x}$$

$$E(e_0 e_3) = \left(\frac{1}{m} - \frac{1}{N} \right) \rho_{xy} C_x C_y, \quad E(e_0 e_4) = - \left(\frac{1}{m} - \frac{1}{N} \right) \frac{S_{yM_x} [f_x(M_x)]^{-1}}{\bar{Y} M_x}$$

$$E(e_0 e_5) = \left(\frac{1}{n} - \frac{1}{N} \right) \rho_{yz} C_y C_z, \quad E(e_0 e_6) = - \left(\frac{1}{n} - \frac{1}{N} \right) \frac{S_{yM_z} [f_z(M_z)]^{-1}}{\bar{Y} M_z}$$

$$E(e_0 e_7) = \left(\frac{1}{m} - \frac{1}{N} \right) \rho_{yz} C_y C_z, \quad E(e_0 e_8) = - \left(\frac{1}{m} - \frac{1}{N} \right) \frac{S_{yM_z} [f_z(M_z)]^{-1}}{\bar{Y} M_z}$$

$$E(e_1e_2) = -\left(\frac{1}{n} - \frac{1}{N}\right) \frac{S_{XM_x} [f_x(M_x)]^{-1}}{\overline{XM_x}}, \quad E(e_1e_3) = \left(\frac{1}{m} - \frac{1}{N}\right) C_x^2$$

$$E(e_1e_4) = -\left(\frac{1}{m} - \frac{1}{N}\right) \frac{S_{XM_x} [f_x(M_x)]^{-1}}{\overline{XM_x}},$$

$$E(e_1e_5) = E(e_1e_6) = E(e_1e_7) = E(e_1e_8) = 0,$$

$$E(e_2e_3) = -\left(\frac{1}{m} - \frac{1}{N}\right) \frac{S_{XM_x} [f_x(M_x)]^{-1}}{\overline{XM_x}}, \quad E(e_2e_4) = \left(\frac{1}{m} - \frac{1}{N}\right) \frac{1}{4[f_x(M_x)]^2 M_x^2}$$

$$E(e_2e_5) = E(e_2e_6) = E(e_2e_7) = E(e_2e_8) = 0,$$

$$E(e_3e_4) = -\left(\frac{1}{m} - \frac{1}{N}\right) \frac{S_{XM_x} [f_x(M_x)]^{-1}}{\overline{XM_x}},$$

$$E(e_3e_5) = E(e_3e_6) = E(e_3e_7) = E(e_3e_8) = E(e_4e_5) = E(e_4e_6) = E(e_4e_7) = E(e_4e_8) = 0$$

$$E(e_5e_6) = -\left(\frac{1}{n} - \frac{1}{N}\right) \frac{S_{ZM_z} [f_z(M_z)]^{-1}}{\overline{ZM_z}} \quad E(e_5e_7) = \left(\frac{1}{m} - \frac{1}{N}\right) C_z^2$$

$$E(e_5e_8) = -\left(\frac{1}{m} - \frac{1}{N}\right) \frac{S_{ZM_z} [f_z(M_z)]^{-1}}{\overline{ZM_z}} = E(e_6e_7)$$

$$E(e_6e_8) = \left(\frac{1}{m} - \frac{1}{N}\right) \frac{1}{4[f_z(M_z)]^2 M_z^2} \quad E(e_7e_8) = -\left(\frac{1}{m} - \frac{1}{N}\right) \frac{S_{ZM_z} [f_z(M_z)]^{-1}}{\overline{ZM_z}}$$

R. Lamichhane, S. Singh, N. Diawara (2015) referred the work of Sedory and Singh(2013), who have formulated the following results

$$E[(\bar{y}_n - \bar{Y})(\hat{M}_x - M_x)] = -\left(\frac{1}{n} - \frac{1}{N}\right) \frac{S_{YM_x}}{[f_x(M_x)]},$$

$$E[(\bar{x}_n - \bar{X})(\hat{M}_x - M_x)] = -\left(\frac{1}{n} - \frac{1}{N}\right) \frac{S_{XM_x}}{[f_x(M_x)]}$$

$$E[(\bar{y}_n - \bar{Y})(\bar{x}_n - \bar{X})] = \left(\frac{1}{n} - \frac{1}{N}\right) S_{xy}$$

Further, Sedory and Singh(2013) referred to Gross(1980), who have formulated the following result

$$Var(\hat{M}_x) = \left(\frac{1}{n} - \frac{1}{N} \right) \frac{1}{4[f_x(M_x)]^2}$$

The proposed estimator $\bar{y}_{REK}$ in terms of e_i 's ($i=0,1,2,3,4,5,6,7,8$) reduces to the following form,

$$\begin{aligned} \bar{y}_{REK} = & \bar{Y}(1+e_0) + \beta_1^* \{\bar{X}(1+e_3) - \bar{X}(1+e_1)\} + \beta_2^* \{M_x(1+e_4) - M_x(1+e_2)\} \\ & + \beta_3^* \{\bar{Z}(1+e_7) - \bar{Z}(1+e_5)\} + \beta_4^* \{M_z(1+e_8) - M_z(1+e_6)\} \end{aligned}$$

$$\text{or, } \bar{y}_{REK} = \bar{Y} + \bar{Y}e_0 + \beta_1^* \bar{X}(e_3 - e_1) + \beta_2^* M_x(e_4 - e_2) + \beta_3^* \bar{Z}(e_7 - e_5) + \beta_4^* M_z(e_8 - e_6) \quad (3.6)$$

$$\text{It is clearly seen that, } E(\bar{y}_{REK}) = \bar{Y} \quad (3.7)$$

Hence, the proposed estimator $\bar{y}_{REK}$ is an unbiased estimator of the Population mean $\bar{Y}$.

We now proceed to the derivation of the minimum variance. By the definition of variance, we have

$$\begin{aligned} Var(\bar{y}_{REK}) &= E\{\bar{y}_{REK} - \bar{Y}\}^2 \\ &= E\{ \bar{Y}e_0 + \beta_1^* \bar{X}(e_3 - e_1) + \beta_2^* M_x(e_4 - e_2) + \beta_3^* \bar{Z}(e_7 - e_5) + \beta_4^* M_z(e_8 - e_6) \}^2 \end{aligned} \quad (3.8)$$

On expanding the square of the sum of the terms inside the parenthesis and taking expectation over each term we finally obtained the expression of the variance as given below,

$$\begin{aligned} Var(\bar{y}_{REK}) = & \left(\frac{1}{n} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{n} - \frac{1}{m} \right) \{ \beta_1^{*2} S_x^2 + 0.25 \beta_2^{*2} [f_x(M_x)]^2 + \beta_3^{*2} S_z^2 + 0.25 \beta_4^{*2} [f_z(M_z)]^2 + 2 \beta_1^* \\ & S_{xy} - 2 \beta_2^* S_{yM_x} [f_x(M_x)]^{-1} + 2 \beta_3^* S_{yz} - 2 \beta_4^* S_{yM_z} [f_z(M_z)]^{-1} - 2 \beta_1^* \beta_2^* S_{xM_x} [f_x(M_x)]^{-1} - 2 \beta_3^* \beta_4^* S_{zM_z} [f_z(M_z)]^{-1} \} \end{aligned} \quad (3.9)$$

On differentiating the above variance in (3.13) w.r.t. $\beta_1^*, \beta_2^*, \beta_3^*$ and β_4^* and equating each to zero we obtain their Least squares estimates given in (3.14), (3.15), (3.16) and (3.17) respectively as follows,

$$\beta_1^* = \frac{-S_{xy}}{S_x^2} + \frac{\beta_2^* S_{xM_x} [f_x(M_x)]^{-1}}{S_x^2} \quad (3.10)$$

$$\beta_2^* = \frac{S_x^2 S_{yM_x} - S_{xy} S_{xM_x}}{[f_x(M_x)]^{-1} (0.25 S_x^2 - S_{xM_x}^2)} \quad (3.11)$$

$$\beta_3^* = \frac{-S_{yz}}{S_z^2} + \frac{\beta_4^* S_{ZM_z} [f_z(M_z)]^{-1}}{S_z^2} \quad (3.12)$$

$$\beta_4^* = \frac{S_z^2 S_{YM_z} - S_{yz} S_{ZM_z}}{[f_z(M_z)]^{-1} (0.25 S_z^2 - S_{ZM_z}^2)} \quad (3.13)$$

Substituting the values of these estimates in the expression of the variance given in (3.9) we obtain the minimum variance of the proposed estimator $\bar{y}_{REK}$ given as follows,

$$Min.Var(\bar{y}_{REK}) = \left(\frac{1}{m} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{n} - \frac{1}{m} \right) \left\{ S_y^2 - \frac{S_{xy}^2}{S_x^2} - \frac{S_{yz}^2}{S_z^2} - \frac{(S_x^2 S_{YM_x} - S_{xy} S_{XM_x})^2}{S_x^2 (0.25 S_x^2 - S_{XM_x}^2)} - \frac{(S_z^2 S_{YM_z} - S_{yz} S_{ZM_z})^2}{S_z^2 (0.25 S_z^2 - S_{ZM_z}^2)} \right\} \quad (3.14)$$

$$\Rightarrow Min.Var(\bar{y}_{REK}) = \left(\frac{1}{m} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{n} - \frac{1}{m} \right) S_y^2 \left\{ 1 - \rho_{xy}^2 - \rho_{yz}^2 - \frac{(S_x^2 S_{YM_x} - S_{xy} S_{XM_x})^2}{S_y^2 S_x^2 (0.25 S_x^2 - S_{XM_x}^2)} - \frac{(S_z^2 S_{YM_z} - S_{yz} S_{ZM_z})^2}{S_y^2 S_z^2 (0.25 S_z^2 - S_{ZM_z}^2)} \right\} \quad (3.15)$$

3.1 Comparison of $\bar{y}_{REK}$ with $\bar{y}_{kal}$

We have,

$$Var(\bar{y}_{REK}) - Var(\bar{y}_{kal}) = - \left(\frac{1}{n} - \frac{1}{m} \right) S_y^2 \left\{ \rho_{yz}^2 + \frac{(S_z^2 S_{YM_z} - S_{yz} S_{ZM_z})^2}{S_y^2 S_z^2 (0.25 S_z^2 - S_{ZM_z}^2)} \right\} < 0 \quad (3.1.1)$$

$$\text{Noting that } (0.25 S_z^2 - S_{ZM_z}^2) > 0 \quad (3.1.2)$$

Hence, the new proposed estimator $\bar{y}_{REK}$ is more efficient than the previous estimator $\bar{y}_{kal}$. However, Vadlamudi (2017) who proposed $\bar{y}_{kal}$ found that his estimator is more efficient than the standard ratio and regression estimators under double sampling.

4. EMPIRICAL STUDY

Three sets of data each of size $N=500$ for the study variable y and the two auxiliary variables x and z have been generated using R-software from poisson distribution with parameter $\lambda=140$ until the following satisfactory results are obtained.

$$\rho_{yx} = 0.7174572, \rho_{yz} = 0.6096531, \rho_{xz} = -0.045281$$

As the value of the last population correlation term is non-significant, we have assumed that $\rho_{xz} \approx 0$ so as to satisfy the constraint of independence of the auxiliary variables of the proposed estimator. The values of the various population parameters are given in Table -1.

Table-1: Values of the Population Parameters

$\bar{Y}$	301.7438872
$\bar{X} =$	130.67
$\bar{Z} =$	119.504
$\sigma_y^2 =$	244.8973372
$\sigma_x^2 =$	131.7891
$\sigma_z^2 =$	122.341984
$S_x^2 =$	132.0532064
$S_y^2 =$	245.3881135
$S_z^2 =$	122.5871583
$S_{yx} =$	129.1508323
$S_{yz} =$	105.7382359
$S_{yM_x} =$	-4.55918733
$S_{xM_x} =$	-4.576212425
$S_{yM_z} =$	-3.811299255
$S_{zM_z} =$	-4.355863727
$M_x =$	131
$M_z =$	120
$f_x(M_x) =$	0.03474425
$f_z(M_z) =$	0.036060764

Next, a large preliminary sample of size $m=220$ is drawn using SRSWOR measuring the values of x and z whose sample means and sample medians are given respectively as in Table-2.

Table-2: Sample mean and sample medians of the two auxiliary variables in the first-phase sample

$\bar{X}_m =$	130.6364	$\hat{M}_x^* =$	131
$\bar{Z}_m =$	120.6	$\hat{M}_z^* =$	120

Further, a subsample of size $n=80$ is drawn from the first phase sample of size $m=220$ using SRSWOR measuring the values of y , x and z having the values of sample means and sample medians as in Table-3

Table-3: Sample mean and sample medians of the study variable and the two auxiliary variables in the second-phase sample

$\bar{y}_n =$	301.3174		
$\bar{x}_n =$	129.15	$\bar{M}_x =$	129.5
$\bar{z}_n =$	120.35	$\bar{M}_z =$	120

The values of the partial regression estimates are given in Table-4.

Table-4: Values of the regression coefficients

$\beta_1^* =$	-0.94634646
$\beta_2^* =$	-0.031756926
$\beta_3^* =$	-0.842358948
$\beta_4^* =$	-0.020496717

Therefore, the variances of the various estimators are given in Table-5 for comparison:

Table-5: Comparison of the variance of the proposed estimator with the previous estimators

<u>Variance</u>	<u>Value</u>
$V(\bar{y}_{REK}) =$	0.845471311
$V(\bar{y}_{kal}) =$	1.571211172
$V(\bar{y}_{SRSWOR}) =$	2.576575191
$V(\bar{y}_{rat(d)}) =$	3.433223041
$V(\bar{y}_{reg(d)}) =$	1.571818666

From Table-5 we see that the variance of the proposed estimator possesses the minimum value amongst the other estimators under consideration based on the generated data. Hence, we can conclude that the proposed estimator is more efficient than the previous estimator and the other standard estimators under consideration.

5. GENERALIZATION OF THE PROPOSED ESTIMATOR

It has now been shown that the proposed estimator $\bar{y}_{REK}$ is more efficient than the previously studied estimators simply by introducing an additional auxiliary variable z that is independent of the first auxiliary variable x and being highly correlated to the study variable y . The question that arises now is whether there will be an increase in precision of the proposed estimator if more than two auxiliary variables are utilized i.e. as we increase the number of auxiliary variables. We shall consider this problem below.

Let Y be the variable under study and let $X_1, X_2, \dots, X_k$ be k -independent auxiliary variables which are highly correlated with the study variable. Let us first consider the following table showing the various differences of sample means and sample medians for the k -auxiliary variables involved in the description of the generalized estimator.

Table-6: Layout of the differences of means and medians for k -auxiliary variables

$r:$	1	2	3	...	k
$X_r:$	X_1	X_2	X_3	...	X_k
$\bar{X}_{r(m)} - \bar{X}_{r(n)}:$	$\bar{X}_{1(m)} - \bar{X}_{1(n)}$	$\bar{X}_{2(m)} - \bar{X}_{2(n)}$	$\bar{X}_{3(m)} - \bar{X}_{3(n)}$	...	$\bar{X}_{k(m)} - \bar{X}_{k(n)}$
$\hat{M}_{x_r}^* - \hat{M}_{x_r}:$	$\hat{M}_{x_1}^* - \hat{M}_{x_1}$	$\hat{M}_{x_2}^* - \hat{M}_{x_2}$	$\hat{M}_{x_3}^* - \hat{M}_{x_3}$	...	$\hat{M}_{x_k}^* - \hat{M}_{x_k}$

It is clear from the above Table-6 that the required number of partial regression coefficients is $2k$ and they can be denoted in general as β_{2r-1}^* and β_{2r}^* for $r=1, 2, \dots, k$.

Then, the generalized form of the proposed estimator is given by

$$\begin{aligned} \bar{y}_{REK(G)} = & \bar{y}_n + \beta_1^* (\bar{x}_{1(m)} - \bar{x}_{1(n)}) + \beta_2^* (\hat{M}_{x_1}^* - \hat{M}_{x_1}) + \beta_3^* (\bar{x}_{2(m)} - \bar{x}_{2(n)}) + \beta_4^* (\hat{M}_{x_2}^* - \hat{M}_{x_2}) \\ & + \dots + \beta_{2k-1}^* (\bar{x}_{k(m)} - \bar{x}_{k(n)}) + \beta_{2k}^* (\hat{M}_{x_k}^* - \hat{M}_{x_k}) \end{aligned} \quad (5.1)$$

$$\Rightarrow \bar{y}_{REK(G)} = \bar{y}_n + \sum_{r=1}^k \left\{ \beta_{2r-1}^* (\bar{x}_{k(m)} - \bar{x}_{k(n)}) + \beta_{2r}^* (\hat{M}_{x_r}^* - \hat{M}_{x_r}) \right\} \quad (5.2)$$

Here, $\bar{x}_{r(m)}$ and $\bar{x}_{r(n)}$ are the sample means of auxiliary variable $X_r (r=1, 2, \dots, k)$ based on the first phase and second phase samples respectively. Similarly, $\hat{M}_{x_r}^*$ and $\hat{M}_{x_r}$ are the sample medians of the auxiliary variable X_r based on the first phase and second phase samples respectively. We define the error functions as follows:

$$e_0 = \left(\frac{\bar{y}_n - \bar{Y}}{\bar{Y}} \right) \quad (5.3)$$

For the k -auxiliary variables we define the error functions according to the following table

Table-7: Layout of the various error functions for k -auxiliary variables

r :	1	2	3	...	k
X_r :	X_1	X_2	X_3	...	X_k
e_{4r-3} :	$e_1 = \left(\frac{\bar{x}_{1(n)} - \bar{X}}{\bar{X}} \right)$	$e_5 = \left(\frac{\bar{x}_{2(n)} - \bar{X}}{\bar{X}} \right)$	$e_9 = \left(\frac{\bar{x}_{3(n)} - \bar{X}}{\bar{X}} \right)$	...	$e_{4k-3} = \left(\frac{\bar{x}_{k(n)} - \bar{X}}{\bar{X}} \right)$
e_{4r-2} :	$e_2 = \left(\frac{\hat{M}_{x_1}^* - M_{x_1}}{M_{x_1}} \right)$	$e_6 = \left(\frac{\hat{M}_{x_2}^* - M_{x_2}}{M_{x_2}} \right)$	$e_{10} = \left(\frac{\hat{M}_{x_3}^* - M_{x_3}}{M_{x_3}} \right)$	...	$e_{4k-2} = \left(\frac{\hat{M}_{x_k}^* - M_{x_k}}{M_{x_k}} \right)$
e_{4r-1} :	$e_3 = \left(\frac{\bar{x}_{1(m)} - \bar{X}}{\bar{X}} \right)$	$e_7 = \left(\frac{\bar{x}_{2(m)} - \bar{X}}{\bar{X}} \right)$	$e_{11} = \left(\frac{\bar{x}_{3(m)} - \bar{X}}{\bar{X}} \right)$	...	$e_{4k-1} = \left(\frac{\bar{x}_{k(m)} - \bar{X}}{\bar{X}} \right)$
e_{4r} :	$e_4 = \left(\frac{\hat{M}_{x_1}^* - M_{x_1}}{M_{x_1}} \right)$	$e_8 = \left(\frac{\hat{M}_{x_2}^* - M_{x_2}}{M_{x_2}} \right)$	$e_{12} = \left(\frac{\hat{M}_{x_3}^* - M_{x_3}}{M_{x_3}} \right)$	...	$e_{4k} = \left(\frac{\hat{M}_{x_k}^* - M_{x_k}}{M_{x_k}} \right)$

In general the error terms defined in Table-7 can be expressed as

$$e_{4r-3} = \left(\frac{\bar{x}_{r(n)} - \bar{X}}{\bar{X}} \right) \text{ (Based on second phase sample)} \quad (5.4)$$

$$e_{4r-2} = \left(\frac{\hat{M}_{x_r} - M_{x_r}}{M_{x_r}} \right) \text{ (Based on second phase sample)} \quad (5.5)$$

$$e_{4r-1} = \left(\frac{\bar{x}_{r(m)} - \bar{X}}{\bar{X}} \right) \text{ (Based on first phase sample)} \quad (5.6)$$

$$e_{4r} = \left(\frac{\hat{M}_{x_r}^* - M_{x_r}}{M_{x_r}} \right) \text{ (Based on first phase sample)} \quad (5.7)$$

where for each case above $r=1,2,\dots,k$

It is worth noting that there are a total of $4k$ number of error terms related to the auxiliary variables X_r 's ($r=1,2,\dots,k$)

For $r=1,2,\dots,k$ the various expectations are given below

$$E(e_0) = E(e_{4r-3}) = E(e_{4r-2}) = 0$$

$$E(e_{4r-1}) = E(e_{4r}) \approx 0$$

$$E(e_0^2) = \left(\frac{1}{n} - \frac{1}{N} \right) C_y^2, \quad E(e_{4r-3}^2) = \left(\frac{1}{n} - \frac{1}{N} \right) C_{x_r}^2, \quad E(e_{4r-2}^2) = \left(\frac{1}{n} - \frac{1}{N} \right) \frac{1}{4[f_{x_r}(M_{x_r})]^2 M_{x_r}^2}$$

$$E(e_{4r-1}^2) = \left(\frac{1}{m} - \frac{1}{N} \right) C_{x_r}^2, \quad E(e_{4r}^2) = \left(\frac{1}{m} - \frac{1}{N} \right) \frac{1}{4[f_{x_r}(M_{x_r})]^2 M_{x_r}^2}$$

$$E(e_0 e_{4r-3}) = \left(\frac{1}{n} - \frac{1}{N} \right) \rho_{x_r y} C_{x_r} C_y, \quad E(e_0 e_{4r-2}) = - \left(\frac{1}{n} - \frac{1}{N} \right) \frac{S_{YM_{x_r}} [f_{x_r}(M_{x_r})]^{-1}}{\bar{Y} M_{x_r}}$$

$$E(e_0 e_{4r-1}) = \left(\frac{1}{m} - \frac{1}{N} \right) \rho_{x_r y} C_{x_r} C_y, \quad E(e_0 e_{4r}) = - \left(\frac{1}{m} - \frac{1}{N} \right) \frac{S_{YM_{x_r}} [f_{x_r}(M_{x_r})]^{-1}}{\bar{Y} M_{x_r}}$$

$$E(e_{4r-3} e_{4r-2}) = - \left(\frac{1}{n} - \frac{1}{N} \right) \frac{S_{X_r M_{x_r}} [f_{x_r}(M_{x_r})]^{-1}}{\bar{X}_r M_{x_r}}, \quad E(e_{4r-3} e_{4r-1}) = \left(\frac{1}{m} - \frac{1}{N} \right) C_{x_r}^2$$

$$E(e_{4r-3}e_{4r}) = -\left(\frac{1}{m} - \frac{1}{N}\right) \frac{S_{X_r M_{x_r}} [f_{x_r}(M_{x_r})]^{-1}}{\bar{X}_r M_{x_r}} = E(e_{4r-2}e_{4r-1})$$

$$E(e_{4r-2}e_{4r}) = \left(\frac{1}{m} - \frac{1}{N}\right) \frac{1}{4[f_{x_r}(M_{x_r})]^2 M_{x_r}^2}$$

$$E(e_{4r-1}e_{4r}) = -\left(\frac{1}{m} - \frac{1}{N}\right) \frac{S_{X_r M_{x_r}} [f_{x_r}(M_{x_r})]^{-1}}{\bar{X}_r M_{x_r}}$$

Now, for $r \neq s$ ($=1, 2, \dots, k$) we have

$$E(e_{4r-3}e_{4s-3}) = E(e_{4r-3}e_{4s-2}) = E(e_{4r-3}e_{4s-1}) = E(e_{4r-3}e_{4s}) = 0$$

$$E(e_{4r-2}e_{4s-3}) = E(e_{4r-2}e_{4s-2}) = E(e_{4r-2}e_{4s-1}) = E(e_{4r-2}e_{4s}) = 0 \quad \text{and so on.}$$

This is due to the assumption of independence of the auxiliary variables X_r 's ($r=1, 2, \dots, k$) and thus they are uncorrelated and consequently the expectations of the product terms for $r \neq s$ ($=1, 2, \dots, k$) vanishes.

The generalized proposed estimator will reduce to the following form

$$\bar{y}_{REK(G)} = \bar{Y} + \bar{Y}e_0 + \sum_{r=1}^k \left\{ \beta_{2r-1}^* \bar{X}_r (e_{4r-1} - e_{4r-3}) + \beta_{2r}^* M_{x_r} (e_{4r} - e_{4r-2}) \right\} \quad (5.8)$$

$$\text{It is easy to see that} \quad E(\bar{y}_{REK(G)}) = \bar{Y} \quad (5.9)$$

Thus, the generalized proposed estimator is also an unbiased estimator of the population mean.

Now, by the definition of variance we have

$$\begin{aligned} \text{Var}(\bar{y}_{REK(G)}) &= E\{ \bar{y}_{REK(G)} - \bar{Y} \}^2 \\ &= E \left\{ \bar{Y}e_0 + \sum_{r=1}^k \left\{ \beta_{2r-1}^* \bar{X}_r (e_{4r-1} - e_{4r-3}) + \beta_{2r}^* M_{x_r} (e_{4r} - e_{4r-2}) \right\} \right\}^2 \end{aligned} \quad (5.10)$$

On simplifying the above expression of the variance (5.10) we arrive at the following form

$$\text{Var}(\bar{y}_{REK(G)}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 + \left(\frac{1}{n} - \frac{1}{m}\right) \left\{ 2 \sum_{r=1}^k \beta_{2r-1}^* S_{x_r y} - 2 \sum_{r=1}^k \beta_{2r}^* S_{Y M_{x_r}} [f_{x_r}(M_{x_r})]^{-1} + \sum_{r=1}^k \beta_{2r-1}^{*2} S_{x_r}^2 \right\}$$

$$-2 \sum_{r=1}^k \beta_{2r-1}^* \beta_{2r}^* S_{X_r M_{x_r}} [f_{x_r}(M_{x_r})]^{-1} + 0.25 \sum_{r=1}^k \beta_{2r-1}^{*2} [f_{x_r}(M_{x_r})]^{-2} \} \quad (5.11)$$

Deriving the least squares estimates of the partial regression coefficients in the same manner as before we get

$$\beta_{2r-1}^* = \frac{-S_{x_r y}}{S_{x_r}^2} + \frac{\beta_{2r}^* S_{X_r M_{x_r}} [f_{x_r}(M_{x_r})]^{-1}}{S_{x_r}^2} \quad (5.12)$$

$$\beta_{2r}^* = \frac{S_{x_r}^2 S_{Y M_{x_r}} - S_{x_r y} S_{X_r M_{x_r}}}{[f_{x_r}(M_{x_r})]^{-1} (0.25 S_{x_r}^2 - S_{X_r M_{x_r}}^2)} \quad (5.13)$$

for $r=1, 2, \dots, k$

Substituting these values (5.12) and (5.13) in (5.11) we obtain the expression of the minimum variance of the generalized estimator as

$$Min.Var(\bar{y}_{REK(G)}) = \left(\frac{1}{m} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{n} - \frac{1}{m} \right) \left\{ S_y^2 - \sum_{r=1}^k \left[\frac{S_{x_r y}^2}{S_{x_r}^2} + \frac{(S_{x_r}^2 S_{Y M_{x_r}} - S_{x_r y} S_{X_r M_{x_r}})^2}{S_{x_r}^2 (0.25 S_{x_r}^2 - S_{X_r M_{x_r}}^2)} \right] \right\} \quad (5.14)$$

$$\Rightarrow Min.Var(\bar{y}_{REK(G)}) = \left(\frac{1}{m} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{n} - \frac{1}{m} \right) S_y^2 \left\{ 1 - \sum_{r=1}^k \left[\rho_{x_r y}^2 + \frac{(S_{x_r}^2 S_{Y M_{x_r}} - S_{x_r y} S_{X_r M_{x_r}})^2}{S_y^2 S_{x_r}^2 (0.25 S_{x_r}^2 - S_{X_r M_{x_r}}^2)} \right] \right\} \quad (5.15)$$

It is obvious from the above form of the variance that the generalized estimator is the most efficient as compared to $\bar{y}_{REK}$, $\bar{y}_{kal}$ and the other standard estimators viz., ratio estimator ($\bar{y}_{rat(d)}$), regression estimator ($\bar{y}_{reg(d)}$) and sample mean ($\bar{y}$).

6. CONCLUSION

The proposed estimator $\bar{y}_{REK}$ was developed on the basis of inclusion of an additional auxiliary variable upon the estimator developed by Vadlamudi(2017), which makes use of only one auxiliary

variable. This proposed estimator was found to be more efficient. The result of the empirical study also confirms these findings.

Further, we proceed to generalize the Vadlamudi(2017) estimator by incorporating more auxiliary variables. We define the estimator as $\bar{y}_{REK(G)}$ on the basis of k-auxiliary variables($k>1$). Analytically, we found that the efficiency of $\bar{y}_{REK(G)}$ increases subsequently for $k=1,2,3,\dots$

Hence, the present study also shows that as more auxiliary information is included in the estimation process, the more precise the estimator will be for estimating the population mean.

7. REFERENCES

- Al-Mannai, M. (2014). Two Auxiliary Variables in Ratio Method of Estimation and its Application. *J. Stat. Appl. Pro.* 3, No. 1, 101-108 (2014)
- Cochran, W.G. (1963). *Sampling Techniques*, 3rd Ed., Wiley, New York.
- Gross, S.T. (1980). Median estimation in sample surveys. *Proceedings of the Survey Methods Section (American Statistical Association)*, pp. 181-184.
- Jhajj, H.S., Bhangu, H. K. (2013). Generalized Estimators of Population Median using Auxiliary Information. *International Journal of Scientific & Engineering Research*, Volume 4, Issue 11, November-2013, ISSN 2229-5518.
- Kuk, A. Y. C., Mak, T. K. (1989). Median estimation in the presence of auxiliary information. *J. R. Statist. Soc., B*, 51, pp. 261–269.
- Kumar, S., Chhapparwal, P. (2017). Robust Exponential Ratio and Product Type Estimators for Population Mean using Order Statistics in Simple Random Sampling. *International Journal of Ecological Economics and Statistics* ISSN 0973-7537 (Online) Volume:38, Issue Number:3
- Koyuncu, N. (2014). On improvement in estimating median in two phase sampling using two auxiliary information. *Int. J. Agricult. Stat. Sci.*, Vol. 10, No. 1, pp. 27-33, 2014
- Ladusingh, L. (2018). *Survey Sampling Methods*, PHI Learning Private Limited, Delhi-1100922
- Lamichhane, R., Singh, S., Diawara, N., (2015). Improved estimation of population mean using known median of auxiliary variable. *Communication in Statistics- Simulation and Computation* · November 2015 DOI: 10.1080/03610918.2015.1062102
- Mukhopadhyay, P. (1998). *Theory and Methods of Survey Sampling*. Prentice Hall India.
- Murthy, M.N. (1967). *Sampling Theory and Methods* (2nd Impression 1977), Statistical Publishing Society Calcutta.
- Sedory, S. and Singh, S. (2013). Estimation of mode using auxiliary information. *Communications in Statistics - Simulation and Computation* (In Press, online version is available).
- Sharma, M.K., Brar, S.S., Kaur, H. (2017). Class of Estimators of Population Median Using New Parametric Relationship for Median. *International Journal of Mathematics and Statistics Invention*

(IJMSI) E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759 www.ijmsi.org Volume 5 Issue 2 || February. 2017 || PP-41-48.

Sharma, P., Singh, R.(Unpublished). Efficient Class of Estimators for Population Median Using Auxiliary Information

Singh, D., Chaudhary, F.S.(1986). *Theory and Analysis of Sample Survey Designs*, New Age International (P) Limited, Publishers.(Formerly Wiley Eastern Limited)

Singh, H.P., Singh, S., Puertas, S.M. (2003). Ratio Type Estimators for the Median of finite populations. *Allgemeines Statistisches Archiv* 87, 369-382 (c) Physica-Verlag 2003, ISSN 0002-6018.

Swain, A.K.P.C. *Finite Population Sampling Theory and Methods*. South Asian Publishers, New Delhi.

Vadlamudi, K.A., Sedory, S.A., Singh, S. (2017). A new estimator of mean using double sampling. *STATISTICS IN TRANSITION new series*, December 2017 Vol. 18, No. 4, pp. 271–290