A GENERALISED CLASS OF ESTIMATOR OF POPULATION MEAN WITH THE COMBINED EFFECT OF MEASUREMENT ERRORS AND NON-RESPONSE IN SAMPLE SURVEY

Neha Singh[*], Gajendra K. Vishwakarma[1*]
[*]Department of Applied Mathematics, Indian Institute of Technology (ISM) Dhanbad, Dhanbad-826004, Jharkhand, India

ABSTRACT
In this paper, a general method of estimation has been proposed to estimate the combined effect of measurement error and non-response in the estimation of population mean using auxiliary information in simple random sampling. The expressions of mean square errors of the proposed estimators have been derived under large sampling approximation. Mean squared errors of the proposed class of estimators have been compared with the existing corresponding estimators based on the measurement error and non-response, and are found to be more efficient. It is also shown that the estimators envisaged by others are a particular member of the proposed class of estimators. In the end, a simulation study has been carried out to verify the superiority of the proposed estimators. Both theoretical and empirical findings are encouraging and support the soundness of the present study.

KEYWORDS: Measurement errors, non-response, study variable, auxiliary variable, bias, mean square error.

MSC: 62D05.

RESUMEN
En este paper, ha sido propuesto un método general de estimación para estimar el efecto combinado de errores de medición y de no-respuesta en la estimación de la media de población, usando información auxiliar en el muestreo simple aleatorio. Las expresiones del error cuadrático medio de los propuestos estimadores han sido derivados bajo aproximaciones, cuando la muestra es grande. Los errores medios cuadráticos de la clase propuesta de estimadores han sido comparados con los correspondientes estimadores existentes basados en errores de medición y de no-respuesta, y se halló que son más eficientes. También se demuestra que estimadores desarrollados por otros autores son miembros particulares de la clase propuesta. Finalmente , un estudio de simulación ha sido llevado a cabo para verificar la superioridad de los propuestos estimadores. Los hallazgos tanto teóricos como empíricos estimulan y soportan la validez de estudio presentado .

PALABRAS CLAVE: errores de medición, no-respuesta, variable de estudio , variable auxiliar , sesgo, error cuadrático medio.

1. INTRODUCTION

In order to estimate the finite population mean under non-response conditions, Hansen and Hurwitz (1946) suggested a technique of sub-sampling by taking a sub-sample from the non-respondent group with the help of extra efforts and an estimator was developed by combining the information available from the response and non-response groups. Following the work of Hansen and Hurwitz (1946) based on sample mean per unit, several authors adopted this procedure in the context of auxiliary information conventional and alternative ratio and regression type estimation methods. Other researchers including Hansen and Hurwitz (1946), Srinath K.P.(1971), Cochran (1977), Bouza C.N. (1981), Rao (1986), Khare and Srivastava (1993), Okafor and Lee (2000), Singh and Kumar (2008), Singh et al. (2010) and Shabbir and Khan (2013) have studied the effect of non-response on different estimators of population parameters (mean, variance, etc.).

However, the researcher faces another type of non-sampling error which is measurement errors while collecting information from individuals. Basically, measurement errors may be characterized as the difference between the recorded value provided by the respondent and the true value of a variable in the study. Estimating the population parameters using auxiliary information, many authors have addressed the problem in the presence of measurement errors. See Cochran (1968), Fuller (1995), Jackman (1999), Cheng and Van Ness (1994) have discussed the impacts of measurement errors in linear and nonlinear regression modelling in their books on measurement errors while Shalabh (1997), Srivastva and Shalabh (2001) and Manisha and

[1] * nehaismstats@gmail.com, vishwagk@rediffmail.com

Singh (2001, 2002) have studied the impact of measurement errors in ratio and regression method of estimation. Also in the estimation of population mean, Srivastava (1971) has suggested a general procedure using information on known means of p auxiliary variates. Following Srivastava (1971), Allen et al. (2003) have proposed general procedure of estimation of population mean using multi-auxiliary variable under measurement errors. Also, Singh and Karpe (2009) have proposed general procedure of estimation of population variance under measurement errors. Further many authors have studied the combined impact of non-response and measurement error simultaneously in the estimation of population mean. Azeem and Hanif (2015) have proposed a method of estimation of population mean in the presence of measurement errors and non-response. Sharma and Singh (2015), Bouza (2018), Singh et al. (2018) have introduced a method of estimation of population mean in the presence of measurement errors and non–response. Azeem and Hanif (2016) have studied the joint influence of measurement errors and non-response in the estimation of population mean.

Following above literature using auxiliary information we have proposed a general methods of estimation of population mean using auxiliary information considering both the study as well as auxiliary variable are recorded with measurement errors and non-response. In this paper, it is also considered that measurement errors presented in re-contacted interviewed units of samples. The MSEs are calculated with combined effect of responded unit and re-contacted interviewed units to the measurement errors of samples. Since the proposed class of estimator is a generalized class of estimators thus several pre-existing estimators of this literature can be a member of this family and their MSE may also be obtained by choosing suitable constant under the joint influence of measurement errors and non-response.

2. SAMPLE STRUCTURES

Let a sample of size n be drawn from a finite population $U=\left(U_1, U_2,..., U_N\right)$ of size N by using simple random sampling without replacement (SRSWOR) scheme. It is assumed that the population of size N consists of two non-overlapping strata of size N_1 and N_2. Stratum N_1 responding units out of N would respond on the first call and Stratum $N_2\left(N_2 = N_1 - N\right)$ non-responding units out of N would not respond on the first call but would respond to the second call. Following Hansen and Hurwitz (1946) technique, each sample unit belongs to one of the three mutually exclusive groups

a) the n_1 response units supply information to the mail questionnaire,
b) the $\left(n_2 - r\right)$ units who did not provide information to the mail questionnaire and were not contacted again by the personal interview method, and
c) the $r\left(r = n_2/k, k > 1\right)$ units are selected at random without replacement from the n_2 non-respondent units who did not respond to the mail questionnaire but enumerated by personal interview, where k is the inverse sampling ratio.

It is assumed that $\left(x_i^*, y_i^*\right)$ are recorded values instead of $\left(X_i^*, Y_i^*\right)$ true values for i-th $\left(i=1, 2,..., n\right)$ sampling units of two characteristic (x, y).It is assumed that measurement errors observed in study variable ,auxiliary variable of responded unit and re-contracted sampled unit. Let the observational or measurement errors be

$$U_i^* = y_i^* - Y_i^* \text{ and } V_i^* = x_i^* - X_i^* \tag{1}$$

which are stochastic in nature and are uncorrelated with mean zero and variances S_U^2 and S_V^2 respectively. Also $f_1 = \left(\frac{1}{n} - \frac{1}{N}\right)$, $\theta = \frac{W_2\left(k - 1\right)}{n}$ and $W_1 = \frac{N_1}{N}$, $W_2 = \frac{N_2}{N}$ .Let $S_Y^2 = \frac{1}{\left(N - 1\right)}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2$ and $S_X^2 = \frac{1}{\left(N - 1\right)}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$ be the population variances of the study variable and auxiliary variable x and y respectively. $S^2_{Y(1)} = \frac{1}{(N_1 - 1)}\sum_{i=1}^{N_1}(y_i - \bar{Y})^2$ and $S^2_{X(1)} = \frac{1}{(N_1 - 1)}\sum_{i=1}^{N_1}(X_i - \bar{X})^2$ denote the variances of responding part of population group and $S^2_{Y(2)} = \frac{1}{(N_2 - 1)}\sum_{i=1}^{N_2}(Y_i - \bar{Y})^2$ and $S^2_{X(2)} = \frac{1}{(N_2 - 1)}\sum_{i=1}^{N_2}(X_i - \bar{X})^2$ is the variances of the variables y and x respectively for the non-responding part of population. Let $S^2_U = \frac{1}{(N - 1)}\sum_{i=1}^{N}(U_i - \bar{U})^2$ and $S^2_V = \frac{1}{(N - 1)}\sum_{i=1}^{N}(V_i - \bar{V})^2$ is the population variance associated with the measurement errors of variables y and x respectively and $S^2_{U(2)} = \frac{1}{(N_2 - 1)}\sum_{i=1}^{N_2}(U_i - \bar{U})^2$ and $S^2_{V(2)} = \frac{1}{(N_2 - 1)}\sum_{i=1}^{N_2}(V_i - \bar{V})^2$ is the population variance associated with the measurement errors of variables y and x respectively for the non-responding part of population. Also C_Y, C_X are coefficient of variation of study and auxiliary variable and $C_{Y(2)}$, $C_{X(2)}$ are coefficient of variation of study and auxiliary variable of non-responding part of population respectively. $\theta_Y = \frac{S_Y^2 + S_U^2}{S_Y^2}$, $\theta_X = \frac{S_X^2 + S_V^2}{S_X^2}$ are reliability ratio of study and auxiliary variable and $\theta_{Y(2)} = \frac{S_{Y(2)}^2 + S_{U(2)}^2}{S_{Y(2)}^2}$, $\theta_{X(2)} = \frac{S_{X(2)}^2 + S_{V(2)}^2}{S_{X(2)}^2}$ are the reliability ratio of study and auxiliary variable of re-contacted part of population respectively . ρ_{XY} is the correlation coefficient for the respondent population and $\rho_{XY(2)}$ is the correlation coefficient for the non-respondent population respectively.

The following researchers have been proposed the method of estimation of population mean dealing with the problem of non-response. We have obtained the mean squared error in presence of measurement errors and non-response jointly as

1. Hansen and Hurwitz (1946) proposed estimator for estimating population mean, when non-response occur is given by

$$\bar{y}^*_{CR} = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\bar{y}_r \tag{2}$$

The variance of the Hansen and Hurwitz (1946) estimator if non-response as well as measurement errors occur

$$V(\bar{y}^*) = f_1 S_Y^2 + \theta S^2_{Y(2)} + f_1 S_U^2 + \theta S^2_{U(2)} \tag{3}$$

2. Cochran (1977) proposed the following ratio-type estimator of the population mean in the presence of non-response

$$\bar{y}^*_{CR} = \bar{y}^* \frac{\bar{X}}{\bar{x}^*} \tag{4}$$

With measurement errors, the mean square error (MSE) of estimator is

$$\begin{aligned} MSE(\bar{y}^*_{CR}) = {} & f_1\left[S_Y^2 + R^2 S_X^2 - 2R\rho_{YX} S_Y S_X\right] + \theta\left[S^2_{Y(2)} + R^2 S^2_{X(2)} - 2R\rho_{YX(2)} S_{Y(2)} S_{X(2)}\right] \\ & + f_1\left[S_U^2 + R^2 S_V^2\right] + \theta\left[S^2_{U(2)} + R^2 S^2_{V(2)}\right] \end{aligned} \tag{5}$$

3. Rao (1986) proposed the following ratio-type estimator, when non-response conditions occur only on study variable

$$\bar{y}^*_{RR} = \bar{y}^* \frac{\bar{X}}{\bar{x}} \tag{6}$$

The mean square error (MSE) of $\bar{y}^*_{RR}$ is given when there is measurement error present, the mean square error (MSE) of Rao (1986) estimator is given by

$$MSE\left(\bar{y}^*_{RR}\right) = f_1\left[S^2_Y + R^2S^2_X - 2R\rho_{YX}S_YS_X\right] + \theta S^2_{Y(2)} + f_1\left[S^2_U + R^2S^2_V\right] + \theta S^2_{U(2)} \tag{7}$$

4. Singh and Kumar (2008) proposed the following chain-ratio-type estimator of population mean

$$\bar{y}^*_{SKR} = \bar{y}^*\left(\frac{\bar{X}}{\bar{x}^*}\right)\left(\frac{\bar{X}}{\bar{x}}\right) \tag{8}$$

If there is measurement errors occur, the mean square error (MSE) of Singh and Kumar (2008) estimator is given by

$$\begin{aligned} MSE\left(\bar{y}^*_{SKR}\right) = {} & f_1\left[S^2_Y + 4R^2S^2_X - 2R\rho_{YX}S_YS_X\right] + \theta\left[S^2_{Y(2)} + R^2S^2_{X(2)} - 2R\rho_{YX(2)}S_{Y(2)}S_{X(2)}\right] \\ & + f_1\left[S^2_U + 4R^2S^2_V\right] + \theta\left[S^2_{U(2)} + R^2S^2_{V(2)}\right] \end{aligned} \tag{9}$$

5. Azeem and Hanif (2016) developed the following chain-ratio type and exponential type estimators in the presence of measurement errors and non-response, are given by respectively

$$\bar{y}^*_{AH1} = \bar{y}^* \frac{\bar{x}'^*}{\bar{X}} \frac{\bar{x}'^*}{\bar{x}^*} \tag{10}$$

and

$$\bar{y}^*_{AH2} = \bar{y}^* \frac{\bar{x}'^*}{\bar{X}} \exp\left(\frac{\bar{x}'^* - \bar{x}^*}{\bar{x}'^* + \bar{x}^*}\right) \tag{11}$$

and the mean square error of and are given by

$$\begin{aligned} MSE\left(\bar{y}^*_{AH1}\right) = {} & f_1\left[S^2_Y + R^2\left(\frac{N+n}{N-n}\right)^2 S^2_X - 2R\left(\frac{N+n}{N-n}\right)\rho_{YX}S_YS_X\right] \\ & + \theta\left[S^2_{Y(2)} + R^2\left(\frac{N+n}{N-n}\right)^2 S^2_{X(2)} - 2R\left(\frac{N+n}{N-n}\right)\rho_{YX(2)}S_{Y(2)}S_{X(2)}\right] \\ & + f_1\left[S^2_U + R^2\left(\frac{N+n}{N-n}\right)^2 S^2_V\right] + \theta\left[S^2_{U(2)} + R^2\left(\frac{N+n}{N-n}\right)^2 S^2_{V(2)}\right] \end{aligned} \tag{12}$$

$$\begin{aligned} MSE\left(\bar{y}^*_{AH2}\right) = {} & f_1\left[S^2_Y + \frac{1}{4}R^2\left(\frac{N+2n}{N-n}\right)^2 S^2_X - R\left(\frac{N+2n}{N-n}\right)\rho_{YX}S_YS_X\right] \\ & + \theta\left[S^2_{Y(2)} + \frac{1}{4}R^2\left(\frac{N+2n}{N-n}\right)^2 S^2_{X(2)} - R\left(\frac{N+2n}{N-n}\right)\rho_{YX(2)}S_{Y(2)}S_{X(2)}\right] \\ & + f_1\left[S^2_U + \frac{1}{4}R^2\left(\frac{N+2n}{N-n}\right)^2 S^2_V\right] + \theta\left[S^2_{U(2)} + \frac{1}{4}R^2\left(\frac{N+2n}{N-n}\right)^2 S^2_{V(2)}\right] \end{aligned} \tag{13}$$

where $\bar{x}'^* = \dfrac{N\bar{X} - n\bar{x}^*}{N - n}$

6. Singh and Sharma (2015) developed the following estimators and its MSE under measurement errors and non-response is given below

$$\bar{y}^{*}_{PS} = m_1\bar{y}^{*} + m_2\frac{\bar{y}^{*}}{\bar{x}^{*}}\bar{X} \tag{14}$$

$$MSE(\bar{y}^{*}_{PS}) = f_1\left[S_Y^2 + S_U^2\right] + \theta\left[S_{Y(2)}^2 + S_{U(2)}^2\right] - \frac{\left[f_1\left(\rho_{YX}S_YS_X\right) + \theta\left(\rho_{YX(2)}S_{Y(2)}S_{X(2)}\right)\right]^2}{f_1\left[S_X^2 + S_V^2\right] + \theta\left[S_{X(2)}^2 + S_{V(2)}^2\right]} \tag{15}$$

3. SUGGESTED CLASS ESTIMATOR

Following Srivastava (1971), we have proposed a general class of estimator for the estimation of population mean when study and auxiliary variables both occurred with non- response and measurement errors.

$$\mu_r = a\left(\bar{y}^{*}, u\right) \tag{16}$$

where, $u = \bar{x}^{*}/\bar{X}$, is ratio of sample mean to the population mean of the auxiliary variable under non-response and measurement errors, $\bar{Y}$ and $\bar{X}$ are population mean of study and auxiliary variable respectively, while $a\left(\bar{y}^{*}, u\right)$ is a parametric function such that the

i) It is continuous and bounded in R

ii) Its first and second order partial derivatives exist, and are continuous and bounded in R

Now expanding function $a\left(\bar{y}^{*}, u\right)$ at the point $\left(\bar{Y}, 1\right)$ in a second order Taylor's series, we have

$$\begin{aligned}\mu_r &= a(\bar{Y}, 1) + (\bar{y}^{*} - \bar{Y})\left.\frac{\partial a(.)}{\partial \bar{y}}\right|_{(\bar{Y},1)} + (u-1)a^{(1)}(\bar{Y}, 1) \\ &+ \frac{1}{2}\left\{(\bar{y}^{*} - \bar{Y})^2\left.\frac{\partial^2 a(.)}{\partial \bar{y}^2}\right|_{(\bar{y}*,1)} + 2(\bar{y}^{*} - \bar{Y})(u-1)\left.\frac{\partial a^{(1)}(.)}{\partial \bar{y}}\right|_{(\bar{y}*,1)} + (u-1)^2 a^{(2)}(\bar{y}*,1)\right\}\end{aligned} \tag{17}$$

where, $a^{(1)}$ and $a^{(2)}$ are the first and second order partial derivatives of $a(.)$ with respect to u . To obtain the bias and MSE of the proposed class, we consider

$$\varepsilon_0 = \frac{\bar{y}^{*}}{\bar{Y}} - 1,\ \varepsilon_1 = \frac{\bar{x}^{*}}{\bar{X}} - 1 \tag{18}$$

Solving the above equation and by taking expectation of the equation we can have the MSE of the proposed class of estimator at (17) as follows:

$$\begin{aligned}MSE(\mu_r) = \bar{Y}^2\Bigg[&\left\{f_1\left(\frac{C_Y^2}{\theta_Y} + \frac{C_X^2}{\theta_X}a_h^2(1) + 2\rho_{YX}C_XC_Ya_{h(1)}\right)\right\} \\ &+ \theta\left\{\left(\frac{C_{Y(2)}^2}{\theta_{Y(2)}} + \frac{C_{X(2)}^2}{\theta_{X(2)}}a^{*2}_h(1) + 2\rho_{YX(2)}C_{X(2)}C_{Y(2)}a^{*}_{h(1)}\right)\right\}\Bigg]\end{aligned} \tag{19}$$

Differentiating the above equation to minimize the value of constant we have the value of constant for sample and non-responded respectively as:

$$a_{h(1)} = -\frac{\rho_{YX}C_Y}{C_X}\theta_X,\ a^{*}_{h(1)} = -\frac{\rho_{YX(2)}C_{Y(2)}}{C_{X(2)}}\theta_{X(2)} \tag{20}$$

Thus the resultant minimum MSE is

$$MSE(\mu_r)_{\min} = \bar{Y}^2\left[\left\{f_1\left(\frac{C_Y^2}{\theta_Y} - \rho_{YX}^2 C_Y^2\theta_X\right)\right\} + \theta\left\{\left(\frac{C_{Y(2)}^2}{\theta_{Y(2)}}\right) - \rho_{YX(2)}^2 C_{Y(2)}^2\theta_{X(2)}\right\}\right] \tag{21}$$

where, $\bar{Y}^2 f_1\left(\frac{C_Y^2}{\theta_Y} - \rho_{YX}^2 C_Y^2\theta_X\right)$ is the MSE under measurement errors of responded units of sample and $\bar{Y}^2\theta\left\{\left(\frac{C_{Y(2)}^2}{\theta_{Y(2)}}\right) - \rho_{XY(2)}^2 C_{Y(2)}^2\theta_{X(2)}\right\}$ is the MSE under measurement error of re-interviewed non responded units.

It may be noted that the estimators (1), (2), (3) and (6) can be members of suggested class of μ_r. Thus MSE of these exiting estimators can be easily derived by using suitably chosen constant under measurement errors and non-response.

4. EFFICIENCY COMPARISON

In this section we compare the suggested class of estimator with respect to existing estimators $\bar{y}^*$, $\bar{y}_{CR}^*$, $\bar{y}_{RR}^*$, $\bar{y}_{SKR}^*$, $\bar{y}_{AH1}^*$, $\bar{y}_{AH2}^*$ and $\bar{y}_{PS}^*$ and result shown below:

$MSE(\mu_r)_{\min}$ will be more efficient than $V(\bar{y}^*)$ if $MSE(\mu_r)_{\min} \le V(\bar{y}^*)$, which provide

$$\left[f_1\rho_{YX}^2 S_Y^2\theta_X + \theta\,\rho_{YX(2)}^2 S_{Y(2)}^2\theta_{X(2)}\right] \ge 0 \tag{22}$$

$MSE(\mu_r)_{\min}$ is more precise than $MSE(\bar{y}_{CR}^*)$, when $MSE(\mu_r)_{\min} \le MSE(\bar{y}_{CR}^*)$

$$\left[f_1\rho_{YX}^2 S_Y^2\theta_X + \theta\,\rho_{YX(2)}^2 S_{Y(2)}^2\theta_{X(2)}\right] \le f_1\left[2R\rho_{YX}S_Y S_X - R^2S_X^2 - R^2S_V^2\right] + \theta\left[2R\rho_{YX(2)}S_{Y(2)}S_{X(2)} - R^2S_{V(2)}^2 - R^2S_{X(2)}^2\right] \tag{23}$$

$MSE(\mu_r)_{\min}$ is preferable over $MSE(\bar{y}_{RR}^*)$ if $MSE(\mu_r)_{\min} \le MSE(\bar{y}_{RR}^*)$, which converge to

$$\left[f_1\rho_{YX}^2 S_Y^2\theta_X + \theta\,\rho_{YX(2)}^2 S_{Y(2)}^2\theta_{X(2)}\right] \le f_1\left[2R\rho_{YX}S_Y S_X - R^2S_X^2 - R^2S_V^2\right] \tag{24}$$

$MSE(\mu_r)_{\min}$ will dominate $MSE(\bar{y}_{SKR}^*)$ if $MSE(\mu_r)_{\min} \le MSE(\bar{y}_{SKR}^*)$, subsequently we get

$$\left[f_1\rho_{YX}^2 S_Y^2\theta_X + \theta\,\rho_{YX(2)}^2 S_{Y(2)}^2\theta_{X(2)}\right] \le f_1\left[2R\rho_{YX}S_Y S_X - 4R^2S_X^2 - 4R^2S_V^2\right] + \theta\left[2R\rho_{YX(2)}S_{Y(2)}S_{X(2)} - 4R^2S_{V(2)}^2 - 4R^2S_{X(2)}^2\right] \tag{25}$$

$MSE(\mu_r)_{\min}$ is better than $MSE(\bar{y}_{AH1}^*)$ if $MSE(\mu_r)_{\min} \le MSE(\bar{y}_{AH1}^*)$, which gives

$$\begin{aligned}\left[f_1\rho_{YX}^2 S_Y^2\theta_X + \theta\,\rho_{YX(2)}^2 S_{Y(2)}^2\theta_{X(2)}\right] \le\ & f_1\left[2R\left(\frac{N+n}{N-n}\right)\rho_{YX}S_Y S_X - \left(\frac{N+n}{N-n}\right)^2 R^2S_X^2 - R^2\left(\frac{N+n}{N-n}\right)^2 S_V^2\right] \\ & +\theta\left[2R\left(\frac{N+n}{N-n}\right)\rho_{YX(2)}S_{Y(2)}S_{X(2)} - R^2\left(\frac{N+n}{N-n}\right)^2 S_{V(2)}^2 - R^2\left(\frac{N+n}{N-n}\right)^2 S_{X(2)}^2\right]\end{aligned} \tag{26}$$

$MSE(\mu_r)_{\min}$ will be more efficient than if $MSE(\mu_r)_{\min} \le MSE(\bar{y}_{AH2}^*)$, which provide

$$\begin{aligned}\left[f_1\rho_{XY}^2 S_Y^2\theta_X + \theta\,\rho_{YX(2)}^2 S_{Y(2)}^2\theta_{Y(2)}\right] \le\ & f_1\left[R\left(\frac{N+2n}{N-n}\right)\rho_{YX}S_Y S_X - \frac{1}{4}\left(\frac{N+2n}{N-n}\right)^2 R^2S_X^2 - \frac{1}{4}R^2\left(\frac{N+2n}{N-n}\right)^2 S_V^2\right] \\ & +\theta\left[R\left(\frac{N+2n}{N-n}\right)\rho_{YX(2)}S_{Y(2)}S_{X(2)} - \frac{1}{4}R^2\left(\frac{N+2n}{N-n}\right)^2 S_{V(2)}^2 - \frac{1}{4}R^2\left(\frac{N+2n}{N-n}\right)^2 S_{X(2)}^2\right]\end{aligned} \tag{27}$$

$MSE\left(\mu_r\right)_{\min}$ will be more efficient than if $MSE\left(\mu_r\right)_{\min} \leq MSE\left(\bar{y}^*_{PS}\right)$, which provide

$$\left[f_1\rho^2_{YX}S^2_Y\theta_X + \theta\,\rho^2_{YX(2)}S^2_{Y(2)}\theta_{X(2)}\right] \leq \frac{\left[f_1\left(\rho_{YX}S_YS_X\right)+\theta\left(\rho_{YX(2)}S_{Y(2)}S_{X(2)}\right)\right]^2}{f_1\left[S^2_X+S^2_V\right]+\theta\left[S^2_{x(2)}+S^2_{V(2)}\right]} \tag{28}$$

5. COST FUNCTION ANALYSIS

Let the cost function considered to be for the proposed estimator as

$$C = c_0 n + c_1 n_1 + c_2 r \tag{29}$$

Where

c_0 : is the initial cost for set up the survey,

c_1 : is the cost per unit collecting and processing data obtained from n_1 respondent

c_2 : is the cost per unit for collecting information from the sub-sampling units and processing from them.

From above equation (29), the expected cost function can be written as

$$C^* = E(C) = n\left(c_0 + c_1W_1 + c_2\frac{W_2}{k}\right) \tag{30}$$

Let the mean square error of estimator in the presence of non-response is represented as

$$MSE(\mu_r) = \frac{V_1}{n} + \frac{kV_2}{n} + terms \text{ independent of n and k}. \tag{31}$$

Where V_1 and V_2 are the coefficient of $\frac{1}{n}$ and $\frac{k}{n}$ respectively in the expression of the MSE of estimator.

Now for minimizing the mean square error for the fixed cost $C^* \leq C^\bullet$ and to obtain the optimum values of n and k , we define a function by

$$\phi = MSE(\mu_r) + \lambda_i\left\{n\left(c_0 + c_1W_1 + c_2\frac{W_2}{k}\right) - C^\bullet\right\} \tag{32}$$

where λ_i is Lagrange's multiplier. Differentiating ϕ with respect to n and k equating it to zero, we can obtain

$$n = \left\{\frac{V_1 + kV_2}{\lambda\left(c_0 + c_1W_1 + c_2\frac{W_2}{k}\right)}\right\}^{\frac{1}{2}} \tag{33}$$

and

$$\frac{n}{k} = \left(\frac{V_2}{\lambda c_2 W_2}\right)^{\frac{1}{2}} \tag{34}$$

Using the value of n and k from (33) and (34) ,we have

$$k_{opt} = \left(\frac{V_1 c_2 W_2}{\left(c_0 + c_1 W\right) V_2} \right)^{\frac{1}{2}} \tag{35}$$

Substituting the values of n from (33) and k from (35) in (30), respectively, we have

$$\sqrt{\lambda} = \frac{1}{C^{\bullet}} \left[\left(V_1 + kV_2\right)\left(c_0 + c_1 W_1 + c_2 \frac{W_2}{k_{opt}}\right) \right]^{\frac{1}{2}} \tag{36}$$

Thus using optimum value of n and k the optimum mean square error of the proposed estimator can be derived as

$$opt\, MSE\,(\mu_r) = \left[\left(\frac{1}{C^{\bullet}}\right)\left(\sqrt{\left(c_0 + c_1 W_1 + c_2 \frac{W_2}{k_{opt}}\right)\left(V_1 + k_{opt} V_2\right)} \right)^2 - \frac{S_y^2}{n} \right] \tag{37}$$

6. SIMULATION STUDY

A Monte-Carlo simulation study has been carried out for the validation of results for large sample properties of the estimator and its distribution. We have generated the data matrix for the population size *N=5000* dividing into two non-overlapping strata N_1 and N_2 for 1000 replications. A sample of size 1000 units have been drawn from the population and further divided into two sets respondent and non- respondent units. The data matrix on X, Y, U and V have been generated using multivariate normal distribution for four variables with mean vectors $\left(\bar{Y} \quad \bar{X} \quad 0 \quad 0\right)$ and covariance matrix

$$\begin{pmatrix} S_Y^2 & \rho_{YX}\sigma_X\sigma_Y & 0 & 0 \\ \rho_{YX} S_X S_Y & S_X^2 & 0 & 0 \\ 0 & 0 & S_U^2 & 0 \\ 0 & 0 & 0 & S_V^2 \end{pmatrix}$$

Using simulation, two sets of population have been generated by using different parameters namely, population I and population II to validate the result of the proposed estimators.

Details of Population I.

$N = 5000$, $\bar{Y} = 80.00$, $\bar{X} = 72.00$, $S_Y^2 = 25.00$, $S_X^2 = 27.00$, $S_U^2 = 1.08$, $S_V^2 = 1.09$, $\rho_{YX} = 0.85$

Table 1: Parameters obtained by simulated data for population I

N_1	N_2	$S^2_{Y(1)}$	$S^2_{X(1)}$	$S^2_{Y(2)}$	$S^2_{X(2)}$	$S^2_{U(1)}$	$S^2_{V(1)}$	$S^2_{U(2)}$	$S^2_{V(2)}$
3800	1200	5.275	5.586	4.949	5.314	1.031	1.036	1.040	1.086
3900	1100	4.657	4.406	5.035	4.882	1.101	1.081	1.059	1.039
4000	1000	4.557	4.006	5.005	4.222	1.109	1.026	1.406	1.231

Detail of Population II.

$N = 5000$, $\bar{Y} = 110.02$, $\bar{X} = 118.94$, $S_Y^2 = 12.26$, $S_X^2 = 16.05$, $S_U^2 = 1.09$, $S_V^2 = 1.12$, $\rho_{YX} = 0.65$

Table-2: Parameters obtained by simulated data for population II

N_1	N_2	$S^2_{Y(1)}$	$S^2_{X(1)}$	$S^2_{Y(2)}$	$S^2_{X(2)}$	$S^2_{U(1)}$	$S^2_{V(1)}$	$S^2_{U(2)}$	$S^2_{V(2)}$
3800	1200	3.450	4.151	3.467	4.029	1.009	1.062	0.970	1.055
3900	1100	3.3887	4.0079	3.5409	4.0486	1.031	1.0328	1.0365	1.0316
4000	1000	3.4429	3.9859	3.6145	4.1408	1.084	1.0632	1.0121	1.0274

We have computed the percentage relative efficiencies (PREs) of the suggested estimator μ_r and other existing estimators $\bar{y}^*_{CR}$, $\bar{y}^*_{RR}$, $\bar{y}^*_{SKR}$, $\bar{y}^*_{AH1}$, $\bar{y}^*_{AH2}$ and $\bar{y}^*_{PS}$ with respect to usual unbiased Hansen and Hurwitz (1946) estimator $\bar{y}^*$ and PRE given in Table 3 and Table 4.

The percent relative efficiency of an estimator τ with respect to usual unbiased estimator $\bar{y}^*$ is defined by

$$PRE(\tau,\bar{y}^*) = \frac{V(\bar{y}^*)}{M(\tau)}\times 100 \tag{38}$$

where $\tau = \bar{y}^*$, $\bar{y}^*_{CR}$, $\bar{y}^*_{RR}$, $\bar{y}^*_{SKR}$, $\bar{y}^*_{AH1}$, $\bar{y}^*_{AH2}$ and $\bar{y}^*_{PS}$.

Table-3. PREs of the estimators with respect to the Hansen and Hurwitz (1946) estimator $\bar{y}^*$ for population I, when there is non-response and measurement errors occur simultaneously.

N_1	N_2	k	$\bar{y}^*$	$\bar{y}^*_{CR}$	$\bar{y}^*_{RR}$	$\bar{y}^*_{SKR}$	$\bar{y}^*_{AH1}$	$\bar{y}^*_{AH2}$	$\bar{y}^*_{ps}$	μ_r
3800	1200	4	100.00	288.92	147.25	40.19	179.16	295.28	333.05	300.19
		6	100.00	299.33	132.72	49.89	231.10	305.94	327.17	388.50
		8	100.00	293.87	125.62	58.67	282.96	300.61	319.21	472.65
3900	1100	4	100.00	297.46	143.58	41.96	187.79	303.07	334.29	315.10
		6	100.00	298.13	130.76	52.28	245.04	303.34	323.67	408.92
		8	100.00	301.56	123.25	61.34	297.39	306.70	316.75	504.70
4000	1000	4	100.00	290.42	141.54	42.92	192.90	297.53	328.41	326.54
		6	100.00	288.21	129.26	54.12	256.02	294.23	318.12	426.83
		8	100.00	295.60	122.26	64.64	320.03	297.09	312.26	531.62

Table-4. PREs of the estimators with respect to the Hansen and Hurwitz (1946) estimator $\bar{y}^*$ for population II, when there is observed of non-response and measurement errors.

N_1	N_2	k	$\bar{y}^*$	$\bar{y}^*_{CR}$	$\bar{y}^*_{RR}$	$\bar{y}^*_{SKR}$	$\bar{y}^*_{AH1}$	$\bar{y}^*_{AH2}$	$\bar{y}^*_{ps}$	μ_r
3800	1200	4	100.00	146.64	108.56	42.90	140.62	156.83	174.60	215.86
		6	100.00	145.30	107.33	47.95	184.33	166.33	156.23	282.83
		8	100.00	145.47	106.14	54.96	230.74	169.45	156.03	353.51
3900	1100	4	100.00	146.58	109.30	41.96	136.24	167.44	156.84	207.08
		6	100.00	144.31	106.88	49.47	173.51	165.63	169.55	267.29
		8	100.00	146.56	105.34	56.64	214.79	187.77	167.42	332.31
4000	1000	4	100.00	144.08	109.36	39.94	127.35	184.49	175.07	197.05
		6	100.00	145.86	107.35	48.15	166.64	187.17	171.61	255.51
		8	100.00	142.50	106.21	54.39	201.64	181.70	167.43	307.13

7. INTERPRETATIONS OF SIMULATION RESULTS

Table-3 conclude that

i) The suggested class of estimator μ_r is more efficient over other existing estimators for almost all different choice of k and stratum N_1 and N_2 except the $\bar{y}^*_{ps}$ for the value of k =4.

ii) Further, it is to be noted that for all Population the PREs of $\bar{y}^*_{RR}$ and $\bar{y}^*_{ps}$ decrease respectively with the increasing value of k. Also with the increasing the value of k, it can be noted that the PRE of $\bar{y}^*_{CR}$ is not much fluctuated.

iii) For Population I, it shows for $\bar{y}^*_{AH1}$, $\bar{y}^*_{SKR}$ and μ_r PRS's are increasing with the increase non-response rate while the PRE is fluctuating for $\bar{y}^*_{AH2}$.

Table-4 conclude that

i) With increases in the value of k the PREs of suggested estimator μ_r increases

ii) The PREs of different estimators follow same trend as followed in the population 1

iii) The PREs of suggested estimator μ_r is always greater as compared to other existing estimators for different values of k.

8. CONCLUSION

From the simulation analyses, we conclude that our suggested class of estimator for both the populations is more justifiable than other existing estimators of similar nature for the estimation of population mean under non-response and measurements occurrence simultaneously. Since the suggested class of estimator is a wider class of estimators and others pre-existing estimators can be a member of this estimator, and their mean square errors can be obtained under measurement errors and non-response cases. Since the suggested class of estimator is more efficient in the estimation of mean so it can be used for future assessment to study the characteristics of the variable of interest when measurement errors and non-response occur in the survey.

Acknowledgement: The authors are thankful to the editor, Secretary Prof. Carlos N. Bouza and the learned referees for their valuable comments towards the improvement of the paper. Authors are also thankful to Ms. Asmita Sharma, Research Scholar (English Literature), Department of Humanities and Social Science, Indian Institute of Technology (ISM) Dhanbad, India for her kind perusal of the manuscript.

RECEIVED MARCH, 2018

REVISED MAY, 2018

REFERENCES

[1] ALLEN, J., SINGH, H.P. and SMARANDACHE, F. (2003): A family of estimators of population mean using multi-auxiliary information in presence of measurement errors. **International Journal of Social Economics**, 30, 837-849.

[2] AZEEM, M. and HANIF, M. (2015): On estimation of population mean in the presence of measurement error and non-response. **Pakistan Journal of Statistics,** 31, 657-670.

[3] AZEEM, M. and HANIF, M. (2016): Joint influence of measurement error and non-response on estimation of population mean. **Communications in Statistics-Theory and Methods**, 46, 1679-1693.

[4] BOUZA, C.N. (1981): Sobre el problema de la fracción de submuestreo para el caso de las no respuestas. **Trabajos de Estadística y de Investigación Operativa**, 21, 18-24.

[5] BOUZA, C.N. (2018): Effect of measurement error and non response on estimation of population mean. **Revista Investigacion Operacional**, 39, 1, 108-120.

[6] CHENG, C.L and NESS, J.W.V. (1994): On estimating linear relationships when both variables are subject to errors. **Journal of the Royal Statistical Society, Series B (Statistical Methodology)**, 56, 167-183.

[7] COCHRAN, W.G. (1968): Errors of measurement in statistics, **Technometrics,** 10, 637-666.

[8] COCHRAN, W.G. (1977): **Sampling techniques.** 3rd edn. John Wiley & Sons, New York.

[9] FULLER, W. (1995): Estimation in the presence of measurement error. **International Statistical**. **Review,** 63, 2, 121-147.

[10] HANSEN, M.H. and HURWITZ, W.N. (1946): The problem of non-response in sample surveys. **Journal of American Statistical Association,** 41, 517-529.

[11] JACKMAN, S. (1999): Correcting surveys for non-response and measurement error using auxiliary information. **Electoral Studies**, 18, 7–27.

[12] KHARE, B.B. and SRIVASTAVA, S.R. (1993): Estimation of population mean using auxiliary character in presence of non-response. **National Academy of Science Letters India**, 16, 111-114.

[13] MANISHA and SINGH, R.K. (2001): An estimation of population mean in the presence of measurement errors. **Journal of the Indian Society of Agricultural Statistics**, 54, 13-18.

[14] MANISHA and SINGH, R.K. (2002): Role of regression estimator involving measurement errors. **Brazilian Journal of Probability and Statistics**, 16, 39-46.

[15] OKAFOR, F.C. and LEE, H. (2000): Double sampling for ratio and regression estimation with sub-sampling the non-respondent. **Survey Methodology**, 26, 183-188.

[16] RAO, P.S.R.S. (1986): Ratio estimation with sub-sampling the non-respondents. **Survey Methodology**, 12, 217-230.

[17] SHABBIR, J. and KHAN, N.S. (2013): Some modified exponential ratio-type estimators in the presence of non-response under two-phase sampling scheme. **Electronic Journal of Applied Statistical Analysis**, 6, 1-17.

[18] SHALABH (1997): Ratio method of estimation in the presence of measurement errors, **Journal of the Indian Society of Agricultural Statistics,** 50, 150-155.
[19] SINGH, H.P. and KARPE, N. (2009): A class of estimators using auxiliary information for estimating finite population variance in presence of measurement errors. **Communications in Statistics-Theory and Methods**, 38:734–741.
[20] SINGH, H.P. and KUMAR, S. (2008): Estimation of mean in presence of non-response using two-phase sampling scheme. **Statistical Papers**, 51, 559-582.
[21] SINGH, H.P., KUMAR, S. and KOZAK, M. (2010): Improved estimation of finite population mean using sub-sampling to deal with the non-response in two-phase sampling scheme. **Communication in Statistics-Theory Methods,** 39, 791-802.
[22] SINGH, P., SINGH, R. and BOUZA, C. N. (2018): Effect of measurement error and non-response on estimation of population mean. **Revista Investigación Operacional,** 39, 1, 108-120.
[23] SINGH, R. and SHARMA, P. (2015): Method of Estimation in the Presence of Non-response and Measurement Errors Simultaneously. **Journal of Modern Applied Statistical Methods**, 14, 107-114.
[24] SRINATH, K.P. (1971): Multiphase sampling in nonresponse problems. **Journal of the American Statistical Association,** 66, 583-586.
[25] SRIVASTAVA, A.K. and SHALABH. (2001): Effect of measurement errors on the regression method of estimation in survey sampling. **Journal of Statistical Research**, 35, 35-44
[26] SRIVASTAVA, S.K. (1971): A generalized estimator for the mean of a finite population using multi-auxiliary information**. Journal of the American Statistical Association**, 66, 404-407.