Estimation of the Population Mean in the Presence of Non-Response and Measurement Errors

Namita Srivastava[1], Akrashi Bhardwaj[2]

Department of Statistics, St. John's College, Agra, U.P, India
Email: *dmamita.sjc[at]gmail.com*
ORCID-ID-0000-0001-8695-9148

Department of Statistics, St. John's College Agra, U.P, India
Corresponding Author Email: *akrashibhardwaj08[at]gmail.com*
ORCID-ID- 0009-0009-4631-9029

Abstract: ***To improve the efficiency of the estimators of population mean, auxiliary information is widely used in literature at the estimation stage. This paper proposes a new class of estimator designed for situations where non-response and measurement error are present, affecting both the study variable and the auxiliary variables. The expressions for the bias and the mean square error (MSE) of this proposed class of estimator are derived, considering the simultaneous occurrence of non-response and response errors, up to the first order of approximation. It was demonstrated that the proposed class of estimators offers superior efficiency compared to traditional methods. when non-response and response errors are present together. To validate these theoretical findings and provide practical insights into its performance, a comprehensive numerical study and simulation study was conducted on real and artificial populations.***

Keywords: Population mean, Study variable, Auxiliary variable, Mean squared error, Measurement errors, Non-response

1. Introduction

In survey sampling when estimating the population mean while constructing the estimators, it is generally assumed that the sampled observations are free from any kind of errors. However, the presence of non-sampling errors is a common phenomenon in surveys and can lead to invalid and inaccurate results if not properly accounted for. Over the past several decades, statisticians were interested in the problem of estimating the parameters of interest in presence of non-sampling errors like nonresponse errors and response errors or measurement errors. If measurement errors are very small and we can neglect it, then the statistical inferences based on observed data continue to remain valid. On the contrary, when they are not appreciably small and negligible, the inferences may not be simply invalid and inaccurate but may often lead to unexpected, undesirable and unfortunate consequences. Some important sources of measurement errors in survey data are discussed in Cochran (1968). Shalabh (1997), Singh and Karpe (2008, 2010), Kumar, Singh, and Smarandache (2011), Kumar, Singh, Sawan, and Chauhan (2011) and Sharma and Singh (2013) studied the properties of some estimators of population parameters under measurement error.

Shalabh (1997) developed the ratio-type estimator when there is measurement error on study and auxiliary variables with known population mean. Other researchers who suggested modified methodology to incorporate the impact of measurement errors are Shalabh (1997), Manisha and Singh (2002), Allen, Singh and Smarandache (2003), Singh and Karpe (2008, 2009, 2010), and Sharma and Singh (2013), etc.

The error due to non-response can be reduced by contacting again the nonresponding units and adjusting the estimate accordingly as suggested and worked out by Hansen and Hurwitz (1946). They proposed a sampling scheme that involves taking a subsample of non-respondents after the first mail attempt and then obtain the information by personal interview. The usual approach to overcome non-response problem is to contact the non-respondent and obtain the information as much as possible.

The information on auxiliary information can be utilized to suggest improved estimators of the population mean in typical surveys and plays a vital role in addressing non-response problems as well. Cochran (1977), Rao (1986), Khare and Srivastava (1993, 1995, 1997), Okafor and Lee (2000), Singh and Kumar (2008, 2009) are among those who have utilized auxiliary information to develop improved estimators of the population mean in the presence of non-response.

Considering that non-response and measurement error are commonly encountered in surveys, it is important to study both of these factors simultaneously. However, there have been limited studies in this area, Singh and Sharma (2015), Azeem and Hanif (2016), Singh and Singh (2018) are among those who have made contributions in this field.

Sharma and Singh (2015) introduced a class of estimator for population mean in the presence of non-response and measurement error both. However, the minimum MSE up to the first order of approximation same as that of regression estimators so both are equally efficient. The present paper is an attempt to suggest an improved class of estimator of population mean which is more efficient than Sharma and Singh (2015) and hence with regression and other estimators in presence of non-response and measurement error simultaneously. The expression for bias and mean square error are derived up to the first order of approximation. An empirical study is carried out to show the efficiency of the proposed estimators over other estimators.

The remainder of the paper is organized as follows. Section 2 presents the notations and Sampling Procedure. Section 3 reviews some existing estimators of the finite population mean in the presence of non-response and measurement error under two-phase sampling. In Section 4, a new class of estimators that simultaneously incorporates measurement error and non-response is proposed. Section 5 Efficiency comparisons using Simulation Section 6 presents simulation and numerical study, Section 7 discusses the results, and concluding remarks are provided in Section 8.

2. Notations and Sampling Procedure

Consider a finite population of size N denoted as $\mathfrak{P}=\{P_1, P_2, P_3, \cdots, P_N\}$. A sample of size n with simple random sampling without replacement is selected and the observations are noted on two characteristics the study variable Y and the auxiliary variable X. Suppose that (y_i, x_i) be the values for $i^{th}(i=1,2,3,\ldots,n)$ observed unit with measurement error for study variable Y and auxiliary variable X respectively and Y_i, X_i be the true value of Y and X for $i^{th}(i=1,2,3,\ldots,n)$ unit respectively in the sample.

Define the variables $U_i=y_i-Y_i$, $V_i=x_i-X_i$ for the measurement error on the study and the auxiliary variable with mean zero and the population variances $S_U^2=\frac{1}{N-1}\sum_{i=1}^{N}(U_i-\bar{U})^2$ and $S_V^2=\frac{1}{N-1}\sum_{i=1}^{N}(V_i-\bar{V})^2$ respectively. Also, let the $U_i's$ and $V_i's$ are uncorrelated but $X_i's$ and $Y_i's$ are correlated and ρ be the population correlation coefficient between X and Y. The population mean and variance for the study variable Y and X are $\mu_y=\frac{1}{N}\sum_{i=1}^{N}y_i$, $S_Y^2=\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-\bar{Y})^2$and $\mu_x=\frac{1}{N}\sum_{i=1}^{N}x_i$, $S_X^2=\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})$ respectively.

The usual unbiased estimator for the population mean of the study variable in the presence of measurement error is given as

$$\bar{y}^*=\frac{1}{n}\sum_{i=1}^{n}y_i$$

The variance of $\bar{y}^*$is given as

$$Var(\bar{y}^*)=\left(\frac{1}{n}-\frac{1}{N}\right)(S_Y^2+S_U^2)$$

The problem of non-response is common and is more widespread in mail surveys than in personal interviews. Hansen and Hurwitz (1946) were the first to deal with the problem of non-response. They proposed a double sampling scheme for estimating population mean, where a sample of size n is drawn randomly from the population, among which n_1 are respondent and n_2 are nonrespondent units. accordingly, the given population be divided into two strata, respondent and nonrespondents of size N_1 and N_2 respectably, Further, a sub sample of size $r\left(=\frac{n_2}{k}\right), k>0$ is taken from n_2 non respondent units. Let population variance for $Y$and X for the group of non-respondents is $S_{Y_{(2)}}^2=\frac{1}{N_2-1}\sum_{i=1}^{N_2}(Y_i-\mu_y)^2$and $S_{X_{(2)}}^2=\frac{1}{N_2-1}\sum_{i=1}^{N_2}(X_i-\mu_x)^2$ respectively.
Then the Hansen and Hurwitz (1946) estimator of the population mean in the presence of non-response and measurement error is given as:

$$\bar{y}^*=w_1\bar{y}_{n_1}+w_2\bar{y}_r$$

where $w_1=\left(\frac{n_1}{n}\right)$, $w_2=\left(\frac{n_2}{n}\right)$, *and* $\bar{y}_{n_1}=\frac{1}{n_1}\sum_{i=1}^{n_1}y_i$, $\bar{y}_r=\frac{1}{r}\sum_{i=1}^{r}y_i$.

The variance $\bar{y}^*$ is given as

$$V(\bar{y}^*)=\lambda S_Y^2\left(1+\frac{S_U^2}{S_Y^2}\right)+\theta S_{y_2}^2\left(1+\frac{S_{U(2)}^2}{S_{Y(2)}^2}\right)$$

where $\lambda=\left(\frac{1-f}{n}\right)$, $\theta=\frac{W_2\,(K-1)}{n}$, $C_Y^2=\frac{S_Y^2}{\mu_y^2}$ and $C_{Y(2)}^2=\frac{S_{Y(2)}^2}{\mu_y^2}$, and the population variance of U and V for the non respondent group are given as $S_{U_{(2)}}^2=\frac{1}{N_2-1}\sum_{i=1}^{N_2}(U_i-\bar{U})^2$and $S_{V_{(2)}}^2=\frac{1}{N_2-1}\sum_{i=1}^{N_2}(V_i-\bar{V})^2$respectively.

The similar expression will be used for the auxiliary variable X.

$$\bar{x}^*=w_1\bar{x}_{n_1}+w_2\bar{x}_r$$

where, $\bar{x}_{n_1}=\frac{1}{n_1}\sum_{i=1}^{n_1}x_i$, $\bar{x}_r=\frac{1}{r}\sum_{i=1}^{r}x_i$.

The variance $\bar{x}^*$ is given as

$$V(\bar{x}^*)=\lambda S_X^2\left(1+\frac{S_V^2}{S_X^2}\right)+\theta S_{X_2}^2\left(1+\frac{S_{V(2)}^2}{S_{X(2)}^2}\right)$$

where $C_X^2=\frac{S_X^2}{\mu_x^2}$ *and* $C_{X(2)}^2=\frac{S_{X(2)}^2}{\mu_x^2}$.

For finding the properties of existing and proposed estimators let us define, $\bar{y}^*=\mu_y(1+e_0)$, $\bar{x}^*=\mu_x(1+e_1)$, *then* $E(e_0)=0=E(e_1)$, *and*

$$E(e_o^2)=\lambda C_Y^2\left(1+\frac{S_U^2}{S_Y^2}\right)+\theta C_{Y_2}^2\left(1+\frac{S_{U_2}^2}{S_{Y_2}^2}\right)=M(say)$$

$$E(e_1^2)=\lambda C_X^2\left(1+\frac{S_V^2}{S_X^2}\right)+\theta C_{X_2}^2\left(1+\frac{S_{V_2}^2}{S_{X_2}^2}\right)=N(say)$$

$$E(e_oe_1)=\lambda\,\rho_{YX}C_YC_X+\theta\rho_{YX_2}C_{Y_2}C_{X_2}=O(say)$$

Where, $\rho_{YX}=\frac{s_{YX}}{s_Ys_X}$, $R=\frac{\mu_y}{\mu_x}$,

3. Existing Estimators of Population Mean in Presence of Both the Errors

1) A traditional unbiased estimator for estimating population mean in the simultaneous presence of and non-response and measurement error is given by

$$T_1=\frac{1}{n}\sum_{i=1}^{n}y_i=\bar{y}^* \tag{3.1}$$

The mean square error of t_1 is given as

$$MSE(T_1)=\mu_y{}^2\left[\lambda C_Y^2\left(1+\frac{S_U^2}{S_Y^2}\right)+\theta\,C_{Y_2}^2\left(1+\frac{S_{U_2}^2}{S_{Y_2}^2}\right)\right]=\mu_y^2\,M \tag{3.2}$$

2) A ratio type estimator for estimating population mean in presence of non-response and measurement error,

$$T_2=\mu_x\left(\frac{\bar{y}^*}{\bar{x}^*}\right) \tag{3.3}$$

$$MSE(T_2)=\mu_y{}^2\left[\lambda C_Y^2\left(1+\frac{S_U^2}{S_Y^2}\right)+\theta\,C_{Y_2}^2\left(1+\frac{S_{U_2}^2}{S_{Y_2}^2}\right)+\lambda C_X^2\left(1+\frac{S_V^2}{S_X^2}\right)+\theta C_{X_2}^2\left(1+\frac{S_{V_2}^2}{S_{X_2}^2}\right)-2(\lambda\,\rho_{YX}C_YC_X+\theta\rho_{YX_2}C_{Y_2}C_{X_2})\right]$$

$$=\mu_y{}^2(M+N-2O) \tag{3.4}$$

3) A regression estimator in presence of non-response and measurement error

$$T_3 = \bar{y}^* + b_{yx}(\mu_x - \bar{x}^*) \quad (3.5)$$

The optimum value of b_{yx} is obtained by minimizing MSE, which is given as

$$b_{yx} = R\frac{O}{N}$$

$$MSE_{min}(T_3) = \mu_y^2\left[M - \frac{O^2}{N}\right] \quad (3.6)$$

Sharma and Singh (2015) estimator:

$$T_4 = m_1\bar{y}^{**} + m_2\frac{\bar{y}^{**}}{\bar{x}^{**}}\cdot\mu_x \quad (3.7)$$

where $m_1 + m_2 = 1$, The optimum value of m_2 is obtained by minimizing MSE, which is given as $m_2 = \frac{O}{N}$, and the minimum MSE is then given by

$$MSE_{min}(T_4) = \mu_y^2\left[M - \frac{O^2}{N}\right] \quad (3.8)$$

4. Proposed Estimator

We propose a new estimator based on Singh and Sharma (2015) estimator to estimate the population mean of variable Y using an auxiliary variable X, accounting for both measurement error and non-response simultaneously. The estimator is given as

$$T_a = m_1\bar{y}^{**} + m_2\frac{\bar{y}^{**}}{\bar{x}^{**}}\cdot\mu_x \quad (4.1)$$

where m_1 and m_2 are the unknown constants, whose values are determined by minimizing the mean square error of the proposed estimator.

Some of the particular case of the proposed estimator are
1) $m_1 = 1, m_2 = 0$, $T_a = \bar{y}^{**}$: The sample mean
2) $m_1 = 0, m_2 = 1$, $T_a = \bar{y}_R{}^{**}$: The Ratio Estimator
3) $m_1 + m_2 = 1$, $T_a = T_4$: The Sharma and Singh (2015) estimator.

4.1 Properties of the Proposed Estimator:

To derive the expression of the bias and mean square error of the proposed estimator, let us rewrite equation (9) in terms of $e's$ which are defined in section 2. Then we have

$$T_a = m_1\mu_y(1+e_0) + m_2\mu_y(1+e_0)(1+e_1)^{-1}$$

$$T_a - \mu_y = \mu_y[(m_1 - 1) + m_1e_0 + m_2(1 + e_0 - e_1 + e_1^2 - e_0e_1)] \quad (4.2)$$

Bias of the Proposed Estimator T_a

Taking expectation on the both sides of equation (10), we get

$$Bias(T_a) = E(T_a - \mu_y) = \mu_y[(m_1 - 1) + m_2(1 + N - O))]$$

Mean Square Error of the Proposed Estimator T_a

Squaring both sides of equation (4.2), we get

$$(T_a - \mu_y)^2 = [\mu_y\{(m_1 - 1) + m_1e_0 + m_2(1 + e_0 - e_1 + e_1^2 - e_0e_1)\}]^2$$

$$(T_a - \mu_y)^2 = \mu_y^2(m_1^2(1 + e_0^2 + 2e_0) + m_2^2(1 + 2e_0 - 2e_1 + e_0^2 + 3e_1^2 - 4e_0e_1) + 2m_1m_2(1 + 2e_0 - e_1 + e_0^2 + e_1^2 - 2e_0e_1) - 2m_1 + 2m_2(1 + e_0 - e_1 + e_1^2 - e_0e_1) + 1) \quad (4.3)$$

Taking expectation on both sides of equation (4.3), we get

$$MSE(T_a) = \mu_y^2(m_1^2(1+M) + m_2^2(1 + M + 3N - 4O) + 2m_1m_2(1 + M + N - 2O) - 2m_1 - 2m_2(1 + N - O) + 1)$$

$$MSE(T_a) = \mu_y^2(m_1^2A + m_2^2B + 2m_1m_2C - 2m_1 - 2m_2D + 1) \quad (4.4)$$

where, $A = 1 + M$; $B = 1 + M + 3N - 4O$; $C = 1 + M + N - 2O$; $D = 1 + N - O$

4.2 Optimum Values of m_1 and m_2 and minimum mean square error:

Differentiating the mean square error (4.4) with respect to m_1 and m_2, and setting the derivatives equal to zero, yields the optimal values of m_1 and m_2, gives

$$m_1 = \frac{B - CD}{AB - C^2}; \quad m_2 = \frac{AD - C}{AB - C^2} \quad (4.5)$$

Placing these values in the equation (4.4) we get

$$MSE_{min}(T_a) = \mu_y^2\left(\left\{1 - \frac{(B - 2CD + AD^2)}{(AB - C^2)}\right\}\right) \quad (4.6)$$

5. Theoretical Efficiency Comparisons

The minimum mean square error $MSE_{min}(T_a)$ of the proposed estimator T_a yields in feasible values when the term $L = \frac{(B - 2CD + AD^2)}{(AB - C^2)} < 1$.The efficiency comparisons with other estimators such as mean, ratio and regression estimators in simultaneous presence of measurement error and non-response, establishes specific conditions under which our proposed estimator outperforms these methods.

(i) The proposed estimator T_a is better than the estimator t_1, if

$MSE(T_1) - MSE_{min}(T_a) > 0$

or, if $\mu_y^2 M - \mu_y^2(\{1 - L\}) > 0$

or, if $L > M - 1$ (5.1)

(ii) The proposed estimator T_a is better than ratio estimator, if

$MSE(T_2) - MSE_{min}(T_a) > 0$

or, if $\mu_y{}^2(M + N - 2O) - \mu_y^2(\{1 - L\}) > 0$

or, if $L > 1 - (M + N - 2O)$ (5.2)

(iii) The proposed estimator T_a is better than Sharma and Singh (2015) estimator if

$MSE_{min}(T_4) - MSE_{min}(T_a) > 0$

or, if $\mu_y^2\left[M - \frac{O^2}{N}\right] - \mu_y^2(\{1 - L\}) > 0$

or, if $L > 1 - M + \frac{O^2}{N}$ (5.3)

Paper ID: SR26714220312 DOI: https://dx.doi.org/10.21275/SR26714220312

6. Empirical Study

To demonstrate the efficiency of our proposed estimator in comparison of other existing methods, this section presents both theoretical comparisons based on mean square errors (MSE's) and a simulation study using one artificially generated population and one real population:

Population 1: An artificial population is generated using MVNORM function of MASS package of R, with the following characteristics:

$N = 250, n = 80$, mean vector is $\begin{bmatrix}\mu_x\\ \mu_y\end{bmatrix} = \begin{bmatrix}40\\ 20\end{bmatrix}$ and covariance matrix is $\sum = \begin{bmatrix}75 & 60\\ 60 & 55\end{bmatrix}$.

The measurement error is introduced using $U = rnorm\ (N, 0,0.7)\ and\ V = rnorm\ (N, 0,0.9)$.

To incorporate non-response for Hansen Hurvitz framework, the population is divided into $N_1 = 150$ respondent and $N_2 = 100$ nonrespondent. Erom a sample of size $n = 80$, we take $n_1 = 48$ respondent units and $n_2 = 32$ non-respondent units.

Population 2: The second population is a real dataset taken from Kish (1967), A list of 270 blocks in Ward 1 of Fall River, Massachusetts from its volume of Block Statistics of the 1950 Census. These blocks are arranged in 27 groups of 10 blocks each. After removing missing values and reindexing the units, the dataset contains $N = 214$ units. The auxiliary variable X represents the number of dwellings and the study variable Y denote the dwellings occupied by the renters.

Since the original data do not explicitly contain both measurement error and non-response, these features are artificially introduced. Non-response is generated by randomly deleting a 0.25 proportion of observations from the study variable, while measurement error is incorporated by contaminating the observed values with 10% of SD of the study and auxiliary variables. Subsampling of the non-respondent group is then carried out following the Hansen–Hurwitz procedure for different values of the subsampling rate (k=2, 3,4,5).

The mean square errors (MSEs) and percent relative efficiencies (PREs) are using two approaches:

(i) **Theoretical Comparison**: MSEs are calculated using parametric expressions and the proposed estimator is compared with mean ratio regression and Sharma and Singh (2015) $(T_1, T_2, T_3, T_4, T_a)$ estimators.

The Percent Relative Efficiency (PRE) of an estimator with respect to the sample mean estimator T_1, in the presence of both measurement error and non-response, is defined as:

$$PRE(T_1, T_j) = \frac{MSE(T_1)}{MSE(T_j)} \times 100, \quad j = 1, 2, 3, 4, a.$$

(ii) **Simulation Study**
A Simulation study is conducted by drawing 50000 samples of size n from respective populations and each sample is divided randomly into n_1 respondent units and n_2 non-respondent units such that $n_1 + n_2 = n$ and further from non-respondent group, a sub-sample of size $r = \frac{n_2}{k}, k = 2,3,4,5,$ is selected. The estimators- mean (t_1) ratio (t_2), regression (t_3) Sharma and Singh (t_4), and our proposed estimator(t_a) are computed for each sample.

The MSE and PRE are given by:

$$MSE(t_j) = \frac{1}{50000} \sum_{s_i=1}^{50000} \left(T_i(s_i) - \mu_y\right)^2 \quad j = 1,2,3,4, a$$

$$PRE(t_j) = \frac{MSE(t_1)}{MSE(t_j)} \times 100 \quad j = 1,2,3,4, a$$

The results for population 1 are reported in Table 1 (theoretical comparison) and Table 2 (simulation study). For Population 2, the results are reported in Table 3 (theoretical comparison) and Table 4 (simulation study). The comparisons have also been made among the situations when there is no error, or either of the one error i.e. non-response or measurement error present in the data. These results are reported for different values of k.

Table 1: Theoretical compression of the Estimators presence of different errors (Population 1)

Estimators		Both error		Non response only		Measurement only		No error	
		MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
T_1	k=2	0.9940	100.00	0.9842	100.00	0.6704	100.00	0.6630	100.00
T_2		0.8003	124.21	0.7525	130.78	0.5175	129.54	0.4924	134.64
$T_3 = T_4$		0.1547	642.52	0.1340	734.47	0.1028	652.03	0.0902	735.08
T_a		0.1542	644.58	0.1335	737.00	0.1026	653.40	0.0900	736.73
T_1	k=3	1.3175	100.00	1.3053	100.00	0.6704	100.00	0.6630	100.00
T_2		1.0830	121.65	1.0126	128.91	0.5175	129.54	0.4924	134.64
$T_3 = T_4$		0.2065	638.15	0.1777	734.39	0.1028	652.03	0.0902	735.08
T_a		0.2056	640.90	0.1769	737.80	0.1026	653.40	0.0900	736.73
T_1	k=4	1.6410	100.00	1.6265	100.00	0.6704	100.00	0.6630	100.00
T_2		1.3657	120.16	1.2727	127.80	0.5175	129.54	0.4924	134.64
$T_3 = T_4$		0.2582	635.65	0.2215	734.42	0.1028	652.03	0.0902	735.08
T_a		0.2568	639.10	0.2202	738.71	0.1026	653.40	0.0900	736.73

Table 2: Theoretical compression of the estimators presence of different errors real data(kish)

Estimators		Both error		Non response only		Measurement only		No error	
		MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
T_1	k=2	5.9453	100.00	5.8903	100.00	3.5840	100.00	3.5472	100.00
T_2		0.6044	983.66	0.5195	1133.79	0.3924	913.47	0.3331	1065.01
$T_3 = T_4$		0.4954	1200.07	0.3996	1474.08	0.3066	1168.84	0.2375	1493.57
T_a		0.4952	1200.49	0.3992	1475.48	0.3065	1169.27	0.2373	1495.06
T_1	k=3	8.3067	100.00	8.2333	100.00	3.5840	100.00	3.5472	100.00
T_2		0.8165	1017.39	0.7060	1166.24	0.3924	913.47	0.3331	1065.01
$T_3 = T_4$		0.6816	1218.65	0.5586	1474.04	0.3066	1168.84	0.2375	1493.57
T_a		0.6814	1219.09	0.5580	1475.55	0.3065	1169.27	0.2373	1495.06
T_1	k=4	10.6680	100.00	10.5763	100.00	3.5840	100.00	3.5472	100.00
T_2		1.0285	1037.22	0.8924	1185.13	0.3924	913.47	0.3331	1065.013
$T_3 = T_4$		0.8670	1230.38	0.7166	1476.00	0.3066	1168.84	0.2375	1493.573
T_a		0.8667	1230.87	0.7157	1477.67	0.3065	1169.27	0.2373	1495.06

Table 3: Simulation results for mean square error and PRE of the Estimators in presence of different errors (Population 1)

Estimators		Both error		Non response only		Measurement only		No error	
		MSE	Pre	MSE	Pre	MSE	Pre	MSE	Pre
T_1	k=2	1.0094	100.00	1.0193	100.00	0.6868	100.00	0.6516	100.00
T_2		0.9405	107.33	0.7769	131.19	0.5842	117.57	0.4868	133.86
T_3		0.1820	554.67	0.1413	721.13	0.1228	559.23	0.0916	711.69
T_4		0.1923	525.04	0.1424	715.99	0.1277	537.94	0.0903	721.48
T_a		0.1877	537.77	0.1419	718.27	0.1231	558.10	0.0902	722.03
T_1	k=3	1.3443	100.00	1.3422	100.00	0.6868	100.00	0.6516	100.00
T_2		1.2545	107.16	1.0789	124.41	0.5842	117.57	0.4868	133.86
T_3		0.2343	573.75	0.1938	692.71	0.1228	559.23	0.0916	711.69
T_4		0.2493	539.29	0.1967	682.46	0.1277	537.94	0.0903	721.48
T_a		0.2442	550.39	0.1959	685.33	0.1231	558.10	0.0902	722.03
T_1	k=4	1.7642	100.00	1.6876	100.00	0.6868	100.00	0.6516	100.00
T_2		1.5930	110.75	1.3776	122.50	0.5842	117.57	0.4868	133.86
T_3		0.2964	595.12	0.2442	691.05	0.1228	559.23	0.0916	711.69
T_4		0.3143	561.36	0.2482	679.81	0.1277	537.94	0.0903	721.48
T_a		0.3095	570.00	0.2468	683.76	0.1231	558.10	0.0902	722.03

Table 4: Simulation results of the Estimators in presence of different errors real data(kish)

Estimators		Both error		Non response only		Measurement only		No error	
		MSE	PRE	MSE	PRE	MSE	PRE	MSE	PRE
T_1	k=2	6.0582	100.00	5.8902	100.00	3.5471	100.00	3.5725	100.00
T_2		0.6279	964.90	0.6007	980.56	0.4190	846.66	0.3329	1073.07
T_4		0.4902	1235.77	0.4828	1219.94	0.3117	1137.95	0.2446	1460.40
T_5		0.5103	1187.18	0.5037	1169.40	0.3240	1094.85	0.2370	1507.68
T_a		0.5064	1196.33	0.4998	1178.58	0.3214	1103.53	0.2368	1508.77
T_1	k=3	8.5741	100.00	8.5743	100.00	3.5471	100.00	3.5725	100.00
T_2		0.8046	1065.68	29.8867	28.69	0.4190	846.66	0.3329	1073.07
T_3		0.6627	1293.89	0.6759	1268.55	0.3117	1137.95	0.2446	1460.40
T_4		0.6854	1250.90	0.6982	1228.08	0.3240	1094.85	0.2370	1507.68
T_a		0.6818	1257.64	0.6942	1235.13	0.3214	1103.53	0.2368	1508.77
T_1	k=4	10.978	100.00	11.0677	100.00	3.5471	100.00	3.5725	100.00
T_2		1.0161	1080.46	0.9956	1111.66	0.4190	846.66	0.3329	1073.07
T_3		0.8725	1258.26	0.8616	1284.60	0.3117	1137.95	0.2446	1460.40
T_4		0.8955	1225.87	0.8816	1255.47	0.3240	1094.85	0.2370	1507.68
T_a		0.8915	1231.45	0.8775	1261.34	0.3214	1103.53	0.2368	1508.77

7. Results and Interpretation

The results of the numerical study and simulation study are presented in Tables 1–4, which reports mean square error (MSE) and percentage relative efficiency (PRE) of the proposed estimator. The performance of the estimators is examined under four distinct scenarios: (i) the simultaneous presence of measurement error and non-response, (ii) the presence of non-response only, (iii) the presence of measurement error only, and (iv) the absence of both errors. All comparisons are carried out for different subsampling rates of the non-respondent group, namely $K = 2,3,$ and 4.

From the tabulated results, it is evident that the proposed class of estimators consistently yields smaller MSEs and substantially higher PRE values than the traditional estimators $T_1, T_2, T_3,$and T_4(mean, ratio, regression and Sharma and Singh (2015) estimators) across all error configurations. T_a slightly improves upon $T_3 = T_4$in all scenarios. Though the numerical improvement is small, it is consistent across all error conditions, indicating stability and robustness. As k increases 2 to 4, MSE of all estimators increases efficiency slightly decreases. However, relative ranking remains unchanged:

$$T_a > T_3 = T_4 > T_2 > T_1$$

In summary MSE is highest when both errors are present. MSE is lowest in the no-error case. $T_a$maintains superiority in all situations which shows robustness to errors.

Table 2. reports the theoretical comparisons among estimators for real population which shows huge improvement from $T_1$to T_2, and further to T_3, T_4, and T_a.PRE values exceed 1000 in many cases i.e. auxiliary information is being used very effectively.

$T_a$again, shows the minimum MSE in all cases. Improvement over $T_3, T_4$is consistent but marginal. Even under both errors, $T_a$maintains very high efficiency. Which indicates strong resistance to combined effects of measurement error and non-response.

Performance after conducting simulation study for both the population is reported in Table 3 and Table 4. Simulation results closely match theoretical findings, validating the derivations.T_a is consistently among the best performers, often slightly better or comparable to T_3.Unlike theoretical equality $T_3 = T_4$, simulations show slight differences due to sampling variability. $T_a$remains competitive and stable, confirming practical applicability.PRE values of $T_a$are extremely high (≈ 550–720), indicating that the proposed estimator is several times more efficient than the conventional estimator. For real population also $T_4$and T_5 perform closely.

Across all tables both errors show highest MSE and No error shows lowest MSE but Efficiency ranking remains unchanged T_a (best) $> T_3/T_4 > T_2 > T_1$.Which shows errors affect magnitude of MSE but do not affect the superiority order.

Overall, the numerical evidence clearly demonstrates that the suggested class of estimators outperforms the existing estimators in terms of both MSE and PRE under all examined scenarios, with the largest gains observed when one or both sources of error are reduced or eliminated.

8. Conclusion

In this present article we proposed an improved estimators of the population mean of the study variable utilizing auxiliary information to enhance efficiency under the presence of non-response and measurement errors. The comparisons are made with the traditional mean ratio, regression and Sharma and Singh (2015) estimators. The bias and mean square errors (MSEs) for the proposed estimators were derived up to the first-order approximation. The theoretical comparison is made with the existing estimators. A numerical study and simulation study clearly established that our estimator outperforms the other estimators. The results clearly demonstrate that the proposed estimator $T_a$outperforms all competing estimators in terms of mean square error and percent relative efficiency under all considered situations. It provides consistent and significant efficiency gains, even in the presence of measurement errors and non-response. The close agreement between theoretical and simulation findings further validates its superiority and practical utility. Therefore, $T_a$is recommended for use in real-life survey situations where such errors are inevitable.

References

[1] Allen, J., Singh, H. P., & Smarandache, F. (2003). A family of estimators of population mean using multiauxiliary information in presence of measurement errors. *International Journal of Social Economics*, *30*(7), 837–849. https://doi.org/10.1108/03068290310478775

[2] Azeem, M., & Hanif, M. (2017). Joint influence of measurement error and non-response on estimation of population mean. *Communications in Statistics - Theory and Methods*, *46*(5), 1679–1693. https://doi.org/10.1080/03610926.2015.1026992

[3] Cochran, W. G. (1968). Errors of measurement in statistics. *Technometrics*, *10*(4), 637–666. https://doi.org/10.2307/1267450

[4] Cochran, W. G. (1977). *Sampling techniques*. Wiley.

[5] Hansen, M. H., & Hurwitz, W. N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, *41*(235), 517–529. https://doi.org/10.1080/01621459.1946.10501894

[6] Khare, B. B., & Srivastava, S. (1993). Estimation of population mean using auxiliary character in presence of non-response. *National Academy of Science and Letters, India*, *16*(3), 111–114.

[7] Khare, B. B., & Srivastava, S. (1995). Study of conventional and alternative two-phase sampling ratio, product and regression estimators in presence of non-response. *Proceedings of the Indian National Science Academy*, *65*(3), 195–203.

[8] Khare, B. B., & Srivastava, S. (1997). Transformed ratio type estimators for the population mean in the presence of nonresponse. *Communications in Statistics - Theory and Methods*, *26*(7), 1779–1791.

[9] Kish Leslie(1965).Survey Sampling. John Wiley & Sons,Inc,

[10] Kumar, M., Singh, R., Sawan, N., & Chauhan, P. (2011). Exponential ratio method of estimators in the presence of measurement errors. *International Journal of Agricultural and Statistical Sciences*, *7*(2), 457–461.

[11] Kumar, M., Singh, R., Singh, A. K., & Smarandache, F. (2011). Some ratio type estimators under measurement errors. *World Applied Sciences Journal*, *14*(2), 272–276.

[12] Manisha, & Singh, R. K. (2002). Role of regression estimator involving measurement errors. *Brazilian Journal of Probability and Statistics*, *16*(1), 39–46.

[13] Okafor, F. C., & Lee, H. (2000). Double sampling for ratio and regression estimation with sub-sampling the non-respondents. *Survey Methodology*, *26*(2), 183–188.

[14] Rao, P. S. R. S. (1986). Ratio estimation with sub sampling the non-respondents. *Survey Methodology*, *12*(2), 217–230.

[15] Shalabh. (1997). Ratio method of estimation in the presence of measurement errors. *Indian Journal of Agricultural Statistics*, *50*(2), 150–155.

[16] Sharma, P., & Singh, R. (2013). A generalized class of estimator for population variance in presence of measurement errors. *Journal of Modern Applied Statistical Methods*, *12*(2), 231–241. https://doi.org/10.22237/jmasm/1383279120

[17] Singh, H. P., & Karpe, N. (2008). Ratio-product estimator for population means in presence of measurement errors. *Journal of Applied Statistical Science*, *16*(4), 437–452.

[18] Singh, H. P., & 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*(5), 734–741. https://doi.org/10.1080/03610920802290713

[19] Singh, H. P., & Karpe, N. (2010). Effect of measurement errors on the separate and combined ratio and product estimators in stratified random sampling. *Journal of Modern Applied Statistical Methods*, *9*(2), 338–402. https://doi.org/10.22237/jmasm/1288584420

[20] Singh, H. P., & Kumar, S. (2008). Estimation of mean in presence of non-response using two phase sampling schemes. *Statistical Papers*, *50*(3), 559–582. https://doi.org/10.1007/s00362-008-0140-5

[21] Singh, H. P., & Kumar, S. (2009). A general class of estimators of the population mean in survey sampling using auxiliary information with sub sampling the non-respondents. *The Korean Journal of Applied Statistics*, *22*(2), 387–402. https://doi.org/10.5351/KJAS.2009.22.2.387

[22] Singh, P., Singh, R., & Bouza, C. N. (2018). Effect of measurement error and non-response on estimation of population mean. *Revita Investigation Operational*, *39*(1), 108–120.

[23] Singh, R. S., & Sharma, P. (2015). Method of estimation in the presence of non-response and measurement errors simultaneously. *Journal of Modern Applied Statistical Methods*, *14*(1), 107–121. https://doi.org/10.22237/jmasm/1430453460

[24] Tiwari, K., Bhogal, S., Kumar, S., & Rather, K. (2022). Using randomized response to estimate the population mean of a sensitive variable under the influence of measurement error. *Journal of Statistical Theory and Practice*, *16*(1), 1–17. https://doi.org/10.1007/s42519-022-00251-1