Communications in Advanced Mathematical Sciences
Vol. 6, No. 4, 196-210, 2023
Research Article
e-ISSN: 2651-4001
DOI:10.33434/cams.1365151

# Quantify the Impact of Non-Response and Measurement Error of Sensitive Variable(s) under Two-Phase Sampling employing ORRT Models

Sunil Kumar[1], Sanam Preet Kour[2]*

**Abstract**
Throughout this article, a two-phase sampling (TPS) technique is employed to estimate the population mean of the sensitive variable. The current article endeavours to develop a chain ratio type estimator for the estimation of sensitive variable(s) in the presence of non-response and measurement error simultaneously by utilizing ORRT models under a two-phase sampling technique. The significant aspects associated with the suggested estimator characterized by bias and mean squared error have been evaluated. Besides this, the utterance for the minimum mean squared error for the optimal values has also been identified. The supremacy of the proposed estimator has been compared with the modified existing estimators under the TPS scheme by using two sensitive auxiliary variables. To clarify the theoretical findings, a simulation study along with a hypothetical generated population and a real population which is based on abortion rates from Statistical Abstract of the United States: 2011 are also addressed in this study.

**Keywords:** Measurement error, Non-response, Optional Randomized Response Models (ORRT), Sensitive variable(s), Two-phase Sampling (TPS)
**2010 AMS:** 62D05

[1] *Department of Statistics, University of Jammu, J&K, India, sunilbhougal06@gmail.com, ORCID: 0000-0003-0249-8415*
[2] *Department of Statistics, University of Jammu, J&K, India, sanamkour903@gmail.com, ORCID: 0000-0002-2894-2556*
***Corresponding author:** sanamkour903@gmail.com
**Received:** 23 September 2023, **Accepted:** 16 November 2023, **Available online:** 5 December 2023
**How to cite this article:** S. Kumar, S. P. Kour, *Quantify the Impact of Non-Response and Measurement Error of Sensitive Variable(s) under Two-Phase Sampling employing ORRT Models*, Commun. Adv. Math. Sci., 6(4) (2023), 196-210.

## 1. Introduction

Sympathetic or contentious issues that are raised in a brusque way may cause some respondents to feel anxious or insecure. As a consequence, they may hide the truth because they donot want their personal intentions to be revealed. Because of the perversion against negative behaviours, respondents may answer 'No' to questions like addiction of drugs, gambling, criminal conviction, domestic abuse, induced abortions, illegal income, tax evasion, even if they have. Such questionnaires encompassing sensitive characteristics necessarily entail the use of innovative techniques such as Randomized Response Technique (RRT) to evoke responses from the sampled units. Warner [1] is the first who posit an inventive RRT for estimating an unknown population prevalence of a sensitive criterion. Greenberg et al. [2] pioneered the estimation of the mean of quantitative sensitive variable by utilising RRT models. Afterwards, Pollack and Bek [3] developed the scrambling response technique for estimating the population mean of a sensitive variable. Gupta et al. [4] models are based on multiplicative scrambling whereas Gupta et al. [5] models are based on additive scrambling which works better than multiplicative scrambling as demonstrated by Gupta et al.

[6]. The notable authors include Zhang et al. [7], Kumar and Kour [8, 9], Kumar et al. [10, 11], Zaman et al. [12] and so forth developed estimation of mean of sensitive variables under non-response and measurement error using ORRT under simple random sampling and two-phase sampling.

In medical sciences, there are well documented instances where sensitive research must be surveilled over time in order to truly comprehend the problem. The evolution of these kind of varying variables may be analyzed by using two-phase sampling (TPS) technique which was first initiated by Neyman [13] and several researchers have since used it in varied incarnations. For illustration, in a survey to estimate the manufacturing of avocado crop predicated on orchards under the crop, only a sub-sampled of the orchards chosen for deciding land area is being used to ascertain the yield rate. Individual authors have been used TPS in varied incarnations including Sanullah et al. [14] who developed a generalized exponential chain ratio estimators under stratified two-phase random sampling, Zaman and Kadilar [15] introduced a new class of exponential estimators for estimating finite population mean in two-phase sampling, Khalil et al. [16] proposed an enhanced two-phase sampling ratio estimator for estimating population mean and among others.

A bulk of studies in a research presume that the data acquisition in a survey is error-free. Unfortunately this is not the reality; measurement error and non-response are very serious flaws in survey sampling. Measurement error (ME) is the difference between the observed value and the theoretical value of the target variable. Cognitive impairment, reputation bias, processing errors and erroneous respondent responses all contribute to measurement errors. Previously, Khalil et al. [17], Onyango et al. [18] deal with the problem of estimation of sensitive variable under measurement error in simple random sampling and double sampling. Withal, it is essential to tackle the issue of non-response in a sampling survey. Non-response (NR) happens when the analyst is unable to gather information from the estimated units of the population. Hansen and Hurwitz [19] is the first one who fix the problem of non-response by conducting a strategy that entails by collecting a sub-sample of non-respondents following the initial mail effort and then analyzing information through personal interview. Diana et al. [20], Gupta et al. [21], Zhang et al. [22], Mukhopadhhyay et al. [23] and so on addressed the problem of estimating the population mean to adjust non-response in varied sampling schemes.

Although we all aware that queries in a survey may have differing levels of sensitivity, and it may be important to quantify this sensitivity. Consequently, the accentuation of this article is exclusively on the chain ratio type estimator for the estimation of sensitive variable(s) in the presence of non-response and measurement error at the same time by making use of ORRT models when study and both two auxiliary variables are sensitive in nature under TPS technique. In section 2 and section 3 there are an ORR technique, an enhanced Hansen and Hurwitz [19] technique, some usual notations and some existing estimators. The Proposed estimator is described in section 4. In section 5, we have studied the efficiency comparisons of all considered estimator(s). To validate the theoretical findings an empirical study for both hypothetical and real population is performed in section 6. Finally, an ultimate conclusion is given in section 7.

## 2. The ORR Technique

Assume that $\Theta = \Theta_1, \Theta_2, ..., \Theta_N$ be a finite population of size $N$ in which $Y$ be the sensitive study variable and $X$ and $Z$ be two sensitive auxiliary variables with means $\bar{Y}$, $\bar{X}$ and $\bar{Z}$ and variances $S_y^2$, $S_x^2$ and $S_z^2$. Take $S$ and $S'$ be two scrambling variables with means $\bar{S}$ and $\bar{S}'$ and variances $S_s^2$ and $S_{s'}^2$, respectively. Let '$\pi$' signifies the probability that the respondent will find the question sensitive. If the respondent consider the question is sensitive, then he or she is prompted to provide a scrambled response for the study $Y$ as well as the auxiliary variables $(X, Z)$, otherwise a legitimate response is recorded. Presuming simple random sampling without replacement (SRSWOR) at each phase, the TPS strategy works as follows

1. During the first phase, a large sample of fixed size $n'$ is taken from $\Theta$ to examine $X$ and $Z$ in order to find estimates of $\bar{X}$ and $\bar{Z}$.

2. In the second phase sample, a fixed-size $n$ sub-sample is taken from $n'$ to observe Y only , so that $(n < n')$.

A conventional additive RRT model with $Y + S'$ as the scrambled response (as in Gupta et al. [6]) or a more comprehensive RRT model with $SY + S'$ as the scrambled response (as in Diana and Perri [24]) could be employed. The simple additive model is a particular case of the more general model if $E(S) = 1$ and with varying variances. The basic additive approach is more efficient, according to Khalil et al. [25], whereas the general model gives greater privacy. Even yet, the generalized randomized response model performs better when we utilize Gupta et al. [21] combined measure of efficiency and privacy, i.e. $\upsilon = \frac{Var(Z_1)}{\Upsilon}$, where $Z_1$ is the scrambled response and $\Upsilon = E(Z_1 - Y)^2$ is the privacy level for the same model as given by Yan et al. [26]. It is important to note that the model with the lower value is preferable since it indicates either a higher level of privacy or a lower value of Var($\hat{y}$), or both. It is worth noting that

$$\upsilon_{additiveRRT} = 1+\frac{S_y^2}{S_{s'}^2} > 1+\frac{S_y^2}{S_{s'}^2+S_s^2(\bar{y}+S_y^2)} = \upsilon_{generalRRT}$$

Under such circumstances, we will utilize the general scrambling model $Ƶ_1 = SY+S'$ as

$$Ƶ_1 = \begin{cases} Y & \text{with probability 1-}\pi \\ SY+S' & \text{with probability }\pi, \end{cases}$$

where $S$ and $S'$ follows normal distribution with mean $(1,0)$ and variances $(S_s^2, S_{s'}^2)$ i.e. $S \sim N(1,S_s^2)$ and $S' \sim N(0,S_{s'}^2)$. The mean and variance of $Ƶ_1$ are as

$E(Ƶ_1) = E(Y)(1-\pi)+E(SY+S')\pi = E(Y)$

and $Var(Ƶ_1) = E(Ƶ_1^2)-E^2(Ƶ_1) = S_y^2+S_{s'}^2\pi+S_s^2(S_y^2+\bar{Y}^2)\pi.$

We can write the randomized linear model as follows

$Ƶ_1 = (SY+S')J+Y(1-J)$, where $J \sim$ Bernoulli($\pi$) with $E(J)=\pi, Var(J)=\pi(1-\pi)$ and $E(J^2)=Var(J)+E^2(J)=\pi$. And the expectation and variance of randomized mechanism is $E_R(Ƶ_1) = (\bar{S}\pi+1-\pi)Y+\bar{S}'\pi$ and $V_R(Ƶ_1) = (Y^2S_s^2+S_{s'}^2)\pi$.

In our research, we assume $X$ and $Z$ to be a sensitive variable(s) then first the general scrambling model for the auxiliary variable $X$ is stated as follows

$$Ƶ_2 = \begin{cases} X & \text{with probability 1-}\pi \\ SX+S' & \text{with probability }\pi, \end{cases}$$

Now, The mean and variance of $Ƶ_2$ are given by

$E(Ƶ_2) = E(X)(1-\pi)+E(SX+S')\pi = E(X)$

and $Var(Ƶ_2) = E(Ƶ_2^2)-E^2(Ƶ_2) = S_x^2+S_{s'}^2\pi+S_s^2(S_x^2+\bar{X}^2)\pi.$

Likewise, we can write randomized linear model as $Ƶ_2 = (SX+S')J+X(1-J)$, where $J \sim$ Bernoulli($\pi$) with $E(J) = \pi, Var(J) = \pi(1-\pi)$ and $E(J^2) = Var(J)+E^2(J) = \pi$. And the expectation and variance of randomized mechanism is $E_R(Ƶ_2) = (\bar{S}\pi+1-\pi)X+\bar{S}'\pi$ and $V_R(Ƶ_2) = (X^2S_s^2+S_{s'}^2)\pi$.

Similarly, for auxiliary variable $Z$, the general scrambling model is given as

$$Ƶ_3 = \begin{cases} Z & \text{with probability 1-}\pi \\ SZ+S' & \text{with probability }\pi, \end{cases}$$

Now, The mean and variance of $Ƶ_3$ are given by

$E(Ƶ_3) = E(Z)(1-\pi)+E(SZ+S')\pi = E(Z)$

and $Var(Ƶ_3) = E(Ƶ_3^2)-E^2(Ƶ_3) = S_z^2+S_{s'}^2\pi+S_s^2(S_z^2+\bar{Z}^2)\pi.$

As well, we can write randomized linear model as $Ƶ_3 = (SZ+S')J+Z(1-J)$, where $J \sim$ Bernoulli($\pi$). The expectation and variance of randomized procedure is $E_R(Ƶ_3) = (\bar{S}\pi+1-\pi)Z+\bar{S}'\pi$ and $V_R(Ƶ_3) = (Z^2S_s^2+S_{s'}^2)\pi$.

The variance of $Ƶ_1$, $Ƶ_2$ and $Ƶ_3$ increases with increase in the probability $\pi$ which demonstrates that the optional RRT model is definitely more efficient than the non-optional RRT model.

## 3. Enhanced Hansen and Hurwitz Technique [19]

From the population $\Theta$, we suppose that only $n_1$ units respond on the first call and the residual $n_2 = n-n_1$ units do not respond. Out from $n_2$ non-responding units, a subsample of size $n_s = \frac{n_2}{k}$; $(k>0)$ is selected. Also, $(N_1, N_2)$ are the sizes of the respondent and non-respondent group. Let us suppose that $\bar{Y}_{(2)}$, $\bar{X}_{(2)}$ and $\bar{Z}_{(2)}$; $S_{y_{(2)}}^2$, $S_{x_{(2)}}^2$ and $S_{z_{(2)}}^2$ are the mean and variances of non-respondent group of size $N_2$, respectively. Hansen and Hurwitz [19] conducted a mail survey on the first conversation and then face-to-face interview on the second call.

The entire population mean of study variable is given by

$$\bar{Y} = W_1\bar{Y}_{(1)} + W_2\bar{Y}_{(2)},$$

where $W_1 = \frac{N_1}{N}$ and $W_2 = \frac{N_2}{N}$.

Let $\bar{y}_1 = \frac{\sum_{i=1}^{N_1} y_i}{n_1}$ be the sample mean for the response group, and $\bar{y}_2 = \frac{\sum_{i=1}^{N_2} y_i}{n_2}$ be the sample mean for non-response group. It is worth noting note that $\bar{y}_1$ and $\bar{y}_2$ are unbiased estimators of $Y_1$ and $Y_2$, respectively.

Hansen and Hurwitz [19] suggested an unbiased population mean estimator which is given by

$$\bar{y} = w_1\bar{y}_1 + w_2\bar{y}_{2s}$$

where $w_1 = \frac{n_1}{n}$ and $w_2 = \frac{n_2}{n}$.

The variance of $\bar{y}$ is given by

$$Var(\bar{y}) = \left(\frac{N-n}{Nn}\right)S_y^2 + \frac{W_2(k-1)}{n}S_{y(2)}^2$$

Within the second phase of the Hansen and Hurwitz [19] methodology, wherein face-to-face interviews of subsampled units of non-respondents are undertaken, we give respondents the opportunity to scramble their response using ORRT to incentivize them to answer a sensitive question honestly. In this scenario, we adapt Hansen and Hurwitz's technique by stating that the respondent group provides direct responses in the first phase, and then the ORRT model is being applied in the second phase to obtain answers from a sample of non-respondents.

Let $\hat{y}_i$ denote a transformation of the randomized response on the $i^{th}$ unit, the expectation of which is the true response $y_i$ under the randomization startegy is given by

$$\hat{y}_i = \frac{z_{1i} - \bar{S}'}{\bar{S}\pi + 1 - \pi}$$

with $E_R(\hat{y}_i) = y_i$ and $V_R(\hat{y}_i) = \frac{V_R(z_{1i})}{(\bar{S}\pi+1-\pi)^2} = \frac{(y_i^2 S_s^2 + S_{s'}^2)\pi}{(\bar{S}\pi+1-\pi)^2} = \delta_{1i}$

Contrastingly, assume that $\hat{x}_i$ and $\hat{z}_i$ denote a transformation of the randomized response on the $i^{th}$ block, the expectation of which is the true response $x_i$ and $z_i$, respectively under the mechanism and is given by

$$\hat{x}_i = \frac{z_{2i} - \bar{S}'}{\bar{S}\pi + 1 - \pi}$$

with $E_R(\hat{x}_i) = x_i$ and $V_R(\hat{x}_i) = \frac{V_R(z_{2i})}{(\bar{S}\pi+1-\pi)^2} = \frac{(x_i^2 S_s^2 + S_{s'}^2)\pi}{(\bar{S}\pi+1-\pi)^2} = \delta_{2i}$

Analogously

$$\hat{z}_i = \frac{z_{3i} - \bar{S}'}{\bar{S}\pi + 1 - \pi}$$

with $E_R(\hat{z}_i) = z_i$ and $V_R(\hat{z}_i) = \frac{V_R(z_{3i})}{(\bar{S}\pi+1-\pi)^2} = \frac{(z_i^2 S_s^2 + S_{s'}^2)\pi}{(\bar{S}\pi+1-\pi)^2} = \delta_{3i}$

From the previous discussions, we alter the Hansen and Hurwitz [19] estimator in the presence of non-response by utilizing ORRT.

$$\hat{\bar{y}} = w_1\bar{y}_1 + w_2\hat{\bar{y}}_2$$

$$\hat{\bar{x}} = w_1\bar{x}_1 + w_2\hat{\bar{x}}_2$$

$$\hat{\bar{z}} = w_1\bar{z}_1 + w_2\hat{\bar{z}}_2$$

where $\hat{\bar{y}}_2 = \sum_{i=1}^{n_s}(\frac{\hat{y}_i}{n_s})$, $\hat{\bar{x}}_2 = \sum_{i=1}^{n_s}(\frac{\hat{x}_i}{n_s})$ and $\hat{\bar{z}}_2 = \sum_{i=1}^{n_s}(\frac{\hat{z}_i}{n_s})$.

It is simple to illustrate that

$E(\hat{\bar{y}}) = \bar{Y}$; $E(\hat{\bar{x}}) = \bar{X}$; $E(\hat{\bar{z}}) = \bar{Z}$

and

The variance of $\hat{\bar{y}}$ is given by

$$Var(\hat{\bar{y}}) = \lambda S_y^2 + \lambda^* S_{y(2)}^2 + \frac{W_2 k}{n}\left[\frac{\{(S_{y(2)}^2 + \bar{y}_{(2)}^2)S_s^2 + S_{s'}^2\}\pi}{(\bar{S}\pi + 1 - \pi)^2}\right]$$

Similarly, the variance of $\hat{\bar{x}}$ is given by

$$Var(\hat{\bar{x}}) = \lambda S_x^2 + \lambda^* S_{x(2)}^2 + \frac{W_2 k}{n}\left[\frac{\{(S_{x(2)}^2 + \bar{x}_{(2)}^2)S_s^2 + S_{s'}^2\}\pi}{(\bar{S}\pi + 1 - \pi)^2}\right]$$

and

$$Var(\hat{\bar{z}}) = \lambda S_z^2 + \lambda^* S_{z(2)}^2 + \frac{W_2 k}{n}\left[\frac{\{(S_{z(2)}^2 + \bar{z}_{(2)}^2)S_s^2 + S_{s'}^2\}\pi}{(\bar{S}\pi + 1 - \pi)^2}\right]$$

where $\lambda = \frac{(N-n)}{Nn}$ and $\lambda^* = \frac{(k-1)W_2}{n}$.

Measurement error, additionally to non-response, is a prominent cause of non-sampling errors in a survey. Let $U_i = y_i - Y_i$, $V_i = x_i - X_i$ and $W_i = z_i - Z_i$ be the measurement error for the study variable $(Y)$ and auxiliary variables $(X, Z)$ in the population. Let $P_i = \mathfrak{z}_{1i} - \mathfrak{Z}_{1i}$, $Q_i = \mathfrak{z}_{2i} - \mathfrak{Z}_{2i}$ and $R_i = \mathfrak{z}_{3i} - \mathfrak{Z}_{3i}$ indicate the respective measurement error associated with the sensitive variables ($\mathfrak{Z}_1$, $\mathfrak{Z}_2$ and $\mathfrak{Z}_3$) in the face-to-face interview phase. These measurement errors are recognised to be random and uncorrelated, with mean zero and variances $S_u^2$, $S_v^2$, $S_w^2$, $S_p^2$, $S_q^2$ and $S_r^2$, respectively.

Numerous notations are presented here, supposing that the population mean of the sensitive auxiliary variable(s) are unknown and that non-response happens on both the study as well as on both the auxiliary variables i.e $X$, $Y$ and $Z$.

$$\hat{\Delta}_y^* = \sum_{i=1}^{n}(y_i - \bar{Y});\ \hat{\Delta}_x^* = \sum_{i=1}^{n}(x_i - \bar{X});\ \hat{\Delta}_z^* = \sum_{i=1}^{n}(z_i - \bar{Z})$$

$$\hat{\Delta}_u^* = \sum_{i=1}^{n_1} U_i + \sum_{i=1}^{n_2} P_i;\ \hat{\Delta}_v^* = \sum_{i=1}^{n_1} V_i + \sum_{i=1}^{n_2} Q_i;\ \hat{\Delta}_w^* = \sum_{i=1}^{n_1} W_i + \sum_{i=1}^{n_2} R_i$$

where $U_i$, $V_i$, $W_i$, $P_i$, $Q_i$ and $R_i$ are measurement errors on $Y$, $X$, $Z$, $\mathfrak{Z}_1$, $\mathfrak{Z}_2$ and $\mathfrak{Z}_3$ respectively.

Furthermore, in the presence of measurement error, the variance of $\hat{\bar{y}}$, $\hat{\bar{x}}$ and $\hat{\bar{z}}$ is given by

$$Var(\hat{\bar{y}}^*) = \lambda(S_y^2 + S_u^2) + \lambda^*(S_{y(2)}^2 + S_p^2) + \kappa_1;$$

$$Var(\hat{\bar{x}}^*) = \lambda(S_x^2 + S_v^2) + \lambda^*(S_{x(2)}^2 + S_q^2) + \kappa_2$$

and

$$Var(\hat{\bar{z}}^*) = \lambda(S_z^2 + S_w^2) + \lambda^*(S_{z(2)}^2 + S_r^2) + \kappa_3$$

where $\kappa_1 = \frac{W_2 k}{n}\left[\frac{\{(S_{y(2)}^2 + \bar{y}_{(2)}^2)S_s^2 + S_{s'}^2\}\pi}{(\bar{S}\pi + 1 - \pi)^2}\right]$; $\kappa_2 = \frac{W_2 k}{n}\left[\frac{\{(S_{x(2)}^2 + \bar{x}_{(2)}^2)S_s^2 + S_{s'}^2\}\pi}{(\bar{S}\pi + 1 - \pi)^2}\right]$ and $\kappa_3 = \frac{W_2 k}{n}\left[\frac{\{(S_{z(2)}^2 + \bar{z}_{(2)}^2)S_s^2 + S_{s'}^2\}\pi}{(\bar{S}\pi + 1 - \pi)^2}\right]$.

Taking $\hat{\bar{y}}^* = \bar{Y}(1 + \hat{e}_0^*)$, $\hat{\bar{x}}^* = \bar{X}(1 + \hat{e}_1^*)$, $\hat{\bar{z}}^* = \bar{Z}(1 + \hat{e}_2^*)$, $\bar{x}' = \bar{X}(1 + e_1')$ and $\bar{z}' = \bar{Z}(1 + e_2')$ such that $E(\hat{e}_0^*) = E(\hat{e}_1^*) = E(\hat{e}_2^*) = E(e_1') = E(e_2') = 0$

To acquire mean squared error, we will used the following notations

$E(\hat{e}_0^{*2}) = \frac{1}{\bar{Y}^2}\left[\lambda(S_y^2 + S_u^2) + \lambda^*(S_{y(2)}^2 + S_p^2) + \kappa_1\right] = \frac{1}{\bar{Y}^2}(A + \kappa_1) = A_1$;

$E(\hat{e}_0^2) = \frac{1}{\bar{Y}^2}\left[\lambda S_y^2 + \lambda^* S_{y(2)}^2 + \kappa_1\right] = \frac{1}{\bar{Y}^2}(\hat{A} + \kappa_1) = A_2$;

$E(e_0^{*2}) = \frac{1}{\bar{Y}^2}\left[\lambda(S_y^2 + S_u^2) + \lambda^*(S_{y(2)}^2 + S_{u(2)}^2)\right] = \frac{1}{\bar{Y}^2}A^* = A_3$;

$E(\hat{e}_1^{*2}) = \frac{1}{\bar{X}^2}\left[\lambda(S_x^2+S_v^2)+\lambda^*(S_{x(2)}^2+S_q^2)+\kappa_2\right] = \frac{1}{\bar{X}^2}(B+\kappa_2) = B_1;$

$E(\hat{e}_1^{2}) = \frac{1}{\bar{X}^2}\left[\lambda S_x^2+\lambda^* S_{x(2)}^2+\kappa_2\right] = \frac{1}{\bar{X}^2}(\hat{B}+\kappa_2) = B_2;$

$E(e_1^{*2}) = \frac{1}{\bar{X}^2}\left[\lambda(S_x^2+S_v^2)+\lambda^*(S_{x(2)}^2+S_{v(2)}^2)\right] = \frac{1}{\bar{X}^2}B^* = B_3;$

$E(\hat{e}_2^{*2}) = \frac{1}{\bar{Z}^2}\left[\lambda(S_z^2+S_w^2)+\lambda^*(S_{z(2)}^2+S_r^2)+\kappa_3\right] = \frac{1}{\bar{Z}^2}(C+\kappa_3) = C_1;$

$E(\hat{e}_2^{2}) = \frac{1}{\bar{Z}^2}\left[\lambda S_z^2+\lambda^* S_{z(2)}^2+\kappa_3\right] = \frac{1}{\bar{Z}^2}(\hat{C}+\kappa_3) = C_2;$

$E(e_2^{*2}) = \frac{1}{\bar{Z}^2}\left[\lambda(S_z^2+S_w^2)+\lambda^*(S_{z(2)}^2+S_{w(2)}^2)\right] = \frac{1}{\bar{Z}^2}A^* = C_3;$

$E(e_1') = \frac{1}{\bar{X}^2}\lambda S_x^2 = A_{11};\ E(e_2') = \frac{1}{\bar{Z}^2}\lambda S_z^2 = C_{11};$

$E(\hat{e}_0^*\hat{e}_1^*) = \frac{1}{\bar{Y}\bar{X}}\left[\lambda\rho_{yx}S_yS_x+\lambda^*\rho_{yx(2)}S_{y(2)}S_{x(2)}\right] = \frac{1}{\bar{Y}\bar{X}}D = D_1;$

$E(\hat{e}_0\hat{e}_1) = \frac{1}{\bar{Y}\bar{X}}\left(\lambda\rho_{yx}S_yS_x\right) = \frac{1}{\bar{Y}\bar{X}}\hat{D} = D_2;$

$E(\hat{e}_1^*\hat{e}_2^*) = \frac{1}{\bar{X}\bar{Z}}\left[\lambda\rho_{xz}S_xS_z+\lambda^*\rho_{xz(2)}S_{x(2)}S_{z(2)}\right] = \frac{1}{\bar{X}\bar{Z}}E = E_1;$

$E(\hat{e}_1\hat{e}_2) = \frac{1}{\bar{X}\bar{Z}}\left(\lambda\rho_{xz}S_xS_z\right) = \frac{1}{\bar{X}\bar{Z}}\hat{E} = E_2;$

$E(\hat{e}_0^*\hat{e}_2^*) = \frac{1}{\bar{Y}\bar{Z}}\left[\lambda\rho_{yz}S_yS_z+\lambda^*\rho_{yz(2)}S_{y(2)}S_{z(2)}\right] = \frac{1}{\bar{Y}\bar{Z}}F = F_1;$

$E(\hat{e}_0\hat{e}_2) = \frac{1}{\bar{Y}\bar{Z}}\left(\lambda\rho_{yz}S_yS_z\right) = \frac{1}{\bar{Y}\bar{Z}}\hat{F} = F_2;$

$E(\hat{e}_0^*e_1') = \frac{1}{\bar{Y}\bar{X}}\lambda'\rho_{yx}S_yS_x = \frac{1}{\bar{Y}\bar{X}}G = G_1;$

$E(\hat{e}_0^*e_2') = \frac{1}{\bar{Y}\bar{Z}}\lambda'\rho_{yz}S_yS_z = \frac{1}{\bar{Y}\bar{Z}}H = H_1;$

$E(\hat{e}_1^*e_1') = \frac{1}{\bar{X}^2}\lambda' S_x^2 = \frac{1}{\bar{X}^2}I = I_1;$

$E(\hat{e}_1^*e_2') = \frac{1}{\bar{X}\bar{Z}}\lambda'\rho_{xz}S_yS_z = \frac{1}{\bar{X}\bar{Z}}J = J_1;$

$E(\hat{e}_2^*e_1') = \frac{1}{\bar{Z}\bar{X}}\lambda'\rho_{zx}S_zS_x = \frac{1}{\bar{Z}\bar{X}}K = K_1;$

$E(\hat{e}_2^*e_2') = \frac{1}{\bar{Z}^2}\lambda' S_z^2 = \frac{1}{\bar{Z}^2}L = L_1;$

$E(e_1'e_2') = \frac{1}{\bar{X}\bar{Z}}\lambda'\rho_{xz}S_xS_z = \frac{1}{\bar{X}\bar{Z}}M = M_1.$

where $\lambda = \left(\frac{1}{n}-\frac{1}{N}\right)$; $\lambda' = \left(\frac{1}{n'}-\frac{1}{N}\right)$ and $\lambda^* = \frac{W_2(k-1)}{n}$.

Next we take the modified conventional estimators i.e. ratio and product estimators into six antithetic strategies depending upon the accessible sensitive auxiliary variables using ORRT models under two-phase sampling (TPS) scheme.

| *Strategies* | *Conventional Estimator(s)* | *Bias* | *Mean Squared Error (MSE)* |
|---|---|---|---|
| **Strategy 1:** When $\hat{\bar{y}}^*$, $\hat{\bar{x}}^*$ and $\bar{x}'$ are used and NR and ME occurs on both sensitive study and auxiliary variable | $\hat{T}_r^* = \hat{\bar{y}}^*\left(\frac{\hat{\bar{x}}^*}{\bar{x}'}\right)$ | $Bias(\hat{T}_r^*) = \varkappa\varphi B - \varkappa\varphi I - \varphi G + \varphi D + \varkappa\varphi\kappa_2$ | $MSE(\hat{T}_r^*) = A + \varkappa^2 B + \varkappa^2 A_{11} - 2\varkappa^2 I + 2\varphi G + 2\varphi D + (\kappa_1 + \varkappa^2\kappa_2)$ where $\varkappa = \frac{\bar{Y}}{\bar{X}}$ and $\varphi = \frac{1}{\bar{X}}$ |
| **Strategy 2:** When $\hat{\bar{y}}$, $\hat{\bar{x}}$ and $\bar{x}'$ are in use and there is absence of NR and ME | $\hat{T}_r = \hat{\bar{y}}\left(\frac{\hat{\bar{x}}}{\bar{x}'}\right)$ | $Bias(\hat{T}_r) = \varkappa\varphi\hat{B} - \varkappa\varphi I - \varphi G + \varphi\hat{D} + \varkappa\varphi\kappa_2$ | $MSE(\hat{T}_r) = \hat{A} + \varkappa^2\hat{B} + \varkappa^2 A_{11} - 2\varkappa^2 I + 2\varphi G + 2\varphi\hat{D} + (\kappa_1 + \varkappa^2\kappa_2)$ |
| **Strategy 3:** When $\bar{y}^*$, $\bar{x}^*$ and $\bar{x}'$ are utilized and NR and ME both occurs on study as well as on auxiliary variable | $T_r^* = \bar{y}^*\left(\frac{\bar{x}^*}{\bar{x}'}\right)$ | $Bias(T_r^*) = \varkappa\varphi B^* - \varkappa\varphi I - \varphi G + \varphi D + \varkappa\varphi\kappa_2$ | $MSE(T_r^*) = A^* + \varkappa^2 B^* + \varkappa^2 A_{11} - 2\varkappa^2 I + +2\varphi G + 2\varphi D + (\kappa_1 + \varkappa^2\kappa_2)$ |
| **Strategy 4:** When $\hat{\bar{y}}^*$, $\hat{\bar{x}}^*$, $\hat{\bar{z}}^*$, $\bar{x}'$ and $\bar{z}'$ are used and NR and ME happens on both the sensitive study as well as auxiliary variables | $\hat{T}_p^* = \hat{\bar{y}}^*\left(\frac{\hat{\bar{x}}^*}{\bar{x}'}\right)\left(\frac{\hat{\bar{z}}^*}{\bar{z}'}\right)$ | $Bias(\hat{T}_r^*) = \varkappa\varphi A_{11} - \varpi\rho L + \varpi\rho C_{11} - \varkappa\rho K + \varkappa\rho E - \varkappa\varphi I - \rho H - \varphi G$ | $MSE(\hat{T}_p^*) = A + \varkappa^2 B + \varpi^2 C + \rho^2 C_{11} + 4\varkappa^2 A_{11} + 4\varkappa\varpi M - 2\varpi^2 L - 2\varpi H - 4\varkappa^2 I - 4\varkappa\varpi K - 4\varkappa G + 2\varkappa\varpi E + 2\varkappa D + 2\varpi F + \kappa_1 + +\varkappa^2\kappa_2 + \varpi^2\kappa_3$ |
| **Strategy 5:** When $\hat{\bar{y}}$, $\hat{\bar{x}}$, $\hat{\bar{z}}$, $\bar{x}'$ and $\bar{z}'$ are utilized and there is no NR and ME happens | $\hat{T}_p = \hat{\bar{y}}\left(\frac{\hat{\bar{x}}}{\bar{x}'}\right)\left(\frac{\hat{\bar{z}}}{\bar{z}'}\right)$ | $Bias(\hat{T}_r) = \varkappa\varphi A_{11} - \varpi\rho L + \varpi\rho C_{11} - \varkappa\rho K + \varkappa\rho\hat{E} - \varkappa\varphi I - \rho H - \varphi G$ | $MSE(\hat{T}_p) = \hat{A} + \varkappa^2\hat{B} + \varpi^2\hat{C} + \rho^2 C_{11} + 4\varkappa^2 A_{11} + 4\varkappa\varpi M - 2\varpi^2 L - 2\varpi H - 4\varkappa^2 I - 4\varkappa\varpi K - 4\varkappa G + 2\varkappa\varpi\hat{E} + 2\varkappa\hat{D} + 2\varpi\hat{F} + \kappa_1 + \varkappa^2\kappa_2 + \varpi^2\kappa_3$ |
| **Strategy 6:** When $\bar{y}^*$, $\bar{x}^*$, $\bar{z}^*$, $\bar{x}'$ and $\bar{z}'$ are employed and NR and ME both occurs on study as well as on the auxiliary variables | $T_p^* = \bar{y}^*\left(\frac{\bar{x}^*}{\bar{x}'}\right)\left(\frac{\bar{z}^*}{\bar{z}'}\right)$ | $Bias(T_r^*) = \varkappa\varphi A_{11} - \varpi\rho L + \varpi\rho C_{11} - \varkappa\rho K + \varkappa\rho E - \varkappa\varphi I - \rho H - \varphi G$ | $MSE(T_p^*) = A^* + \varkappa^2 B^* + \varpi^2 C^* + \rho^2 C_{11} + 4\varkappa^2 A_{11} + 4\varkappa\varpi M - 2\varpi^2 L - 2\varpi H - 4\varkappa^2 I - 4\varkappa\varpi K - 4\varkappa G + 2\varkappa\varpi E + 2\varkappa D + 2\varpi F + \kappa_1 + \varkappa^2\kappa_2 + \varpi^2\kappa_3$ |

**Table 1.** Conventional estimators with their bias and mean square errors using ORRT

## 4. Proposed Chain Ratio Type Estimator

Grabbing inspiration from the existing evidences, an efforts have been made to propose an estimator to improve conventional estimators by multiplying a tuning constant term $\alpha$ whose optimum value is based on the coefficient of variation, which is relatively a stable variable. In addition, inspired by Kumar and Kour [8] and Zhang et al. [22], an information on more than one auxiliary variable can be utilized to suggest a more efficient chain ratio type estimator in the presence of non-response and measurement error simultaneously when the study as well as both the auxiliary variables are sensitive in its essence under three different strategies in two-phase sampling technique by utilizing ORRT models so that one could get a more precise estimate of the population mean.

**Methodology 1:** Assuming $\hat{\bar{y}}^*$, $\hat{\bar{x}}^*$, $\hat{\bar{z}}^*$, $\bar{x}'$ and $\bar{z}'$ are deployed and non-response and measurement error occured on both the sensitive study as well as auxiliary variables i.e. $(X, Z)$ under TPS scheme then the chain ratio type estimator is given as

$$\hat{T}_s^* = \alpha\hat{\bar{y}}^*\left[\alpha_1\left(\frac{\hat{\bar{x}}^*}{\bar{x}'}\right)\left(\frac{\hat{\bar{z}}^*}{\bar{z}'}\right) + (1-\alpha_1)\left(\frac{\bar{x}'}{\hat{\bar{x}}^*}\right)\left(\frac{\bar{z}'}{\hat{\bar{z}}^*}\right)\right]$$

where $\alpha = \frac{\bar{Y}^2}{\bar{Y}^2 + \lambda S_y^2 + \lambda^* S_{y(2)}^2}$.

To evaluate the bias and mean squared error of $\hat{T}_s^*$ by reducing and eliminating terms to first order of approximation, one could verify that

$$(\hat{T}_s^* - \bar{Y}) = \bar{Y}(\alpha - 1) + \alpha\bar{Y}(\hat{e}_2^{*2} - \hat{e}_2^* + e_2' - \hat{e}_2^* e_2' + \hat{e}_1^*\hat{e}_2^* - \hat{e}_2^* e_2' + e_1' - \hat{e}_2^* e_1' + e_1' e_2' - \hat{e}_1^* e_1' - 2\alpha_1 e_2' + 2\alpha_1 \hat{e}_2^* - 2\alpha_1 e_1' + \alpha_1 e_2'^2 + \alpha_1 e_1'^2 - \alpha_1 \hat{e}_2^{*2} - \alpha_1 \hat{e}_1^{*2} + 2\alpha_1 \hat{e}_1^* + \hat{e}_0^* - \hat{e}_0^*\hat{e}_2^* + \hat{e}_0^* e_2' - \hat{e}_0^*\hat{e}_1^* + \hat{e}_0^* e_1' - 2\alpha_1 \hat{e}_0^* e_2' + 2\alpha_1 \hat{e}_0^*\hat{e}_2^* - 2\alpha_1 \hat{e}_0^* e_1' + 2\alpha_1 \hat{e}_0^*\hat{e}_1^*).$$

The bias and mean squared error of the chain ratio type estimator $\hat{\hat{T}}_s^*$ in the inclusion of non-response and measurement error at the same time, is given by

$$Bias(\hat{T}_s^*) = \alpha^*\{(2\alpha_1 - 1)(F_1 + D_1 - H_1 - G_1) + E_1 - (\alpha_1 - 1)C_1 - \alpha_1 B_1 - 2L_1 - K_1 + M_1 + J_1 + \alpha_1\zeta\} - \beta^*$$

where $\beta^* = \left(\frac{\bar{Y}^2 + \lambda S_y^2 + \lambda^* S_{y(2)}^2}{\bar{Y}} + \bar{Y}\right)$; $\alpha^* = \frac{\bar{Y}^3}{\bar{Y}^2 + \lambda S_y^2 + \lambda^* S_{y(2)}^2}$; $\zeta = \lambda'\left(\frac{1}{\bar{Z}^2} S_z^2 + \frac{1}{\bar{X}^2} S_x^2\right)$.

and

$$MSE(\hat{T}_s^*) = \gamma^2 + \alpha^{*2}\left[A_1 + \theta C_1 + \theta A_{11} + \theta C_{11} + 4\alpha_1^2 B_1 + 2\phi F_1 - 2\phi G_1 - 2\phi H_1 + 4\alpha_1 D_1 - 2\theta K_1 + 2\theta L_1 + 4\alpha_1\phi E_1 + 2\theta M_1 + 4\alpha_1\phi I_1 - 4\alpha_1\phi J_1\right] \tag{4.1}$$

where $\theta = 1 + 4\alpha_1^2 - 4\alpha_1$; $\phi = 2\alpha_1 - 1$ and $\gamma = \left(\frac{-(\lambda S_y^2 + \lambda^* S_{y(2)}^2)\bar{Y}^2}{\bar{Y}^2 + \lambda S_y^2 + \lambda^* S_{y(2)}^2}\right)$.

To get the optimum solution of the constant '$\alpha_1$' in $\hat{\hat{T}}_s^*$, we differentiate (4.1) with respect to '$\alpha_1$' and equating it to zero, we have

$$\hat{\alpha}_{1opt.}^* = \frac{-(\gamma^2 + 4\eta)}{8\alpha^{*2}\eta^*} \tag{4.2}$$

where $\eta = F_1 - G_1 - H_1 + D_1$;

and $\eta^* = B_1 + C_1 + A_{11} + C_{11} - 2K_1 + 2L_1 + 2M_1 + 2E_1 + 2I_1 - 2J_1$.

Substituting the optimum value from (4.2), the minimum mean squared error of $\hat{T}_s^*$ is given as

$$min.MSE(\hat{T}_s^*) = \gamma^2 + \alpha^{*2}(A_1 + C_1 + A_{11} + C_{11} - 2K_1 + 2L_1 + 2M_1 - 2F_1 + 2G_1 + 2H_1) - \psi(D_1 - F_1 + G_1 + H_1 - B_1) + \psi^*$$

where $\psi = \frac{-(\gamma^2 + 4\eta)}{2}$ and $\psi^* = \frac{\psi^2}{\alpha^{*2}\eta^*}$.

**Methodology 2:** Letting $\hat{\bar{y}}$, $\hat{\bar{x}}$, $\hat{\bar{z}}$, $\bar{x}'$ and $\bar{z}'$ are being used and there is absence of non-response and measurement error on the sensitive study as well as both the sensitive auxiliary variables i.e. $(X, Z)$ under TPS technique then the chain ratio type estimator is given as

$$\hat{T}_s = \hat{\alpha}\hat{\bar{y}}\left[\alpha_1\left(\frac{\hat{\bar{x}}}{\bar{x}'}\right)\left(\frac{\hat{\bar{z}}}{\bar{z}'}\right) + (1 - \alpha_1)\left(\frac{\bar{x}'}{\hat{\bar{x}}}\right)\left(\frac{\bar{z}'}{\hat{\bar{z}}}\right)\right]$$

where $\hat{\alpha} = \frac{\bar{Y}^2}{\bar{Y}^2 + \lambda S_y^2}$.

The expressions for the bias as well as mean squared error are expressed as

$$Bias(\hat{T}_s) = \hat{\alpha}^*\{(2\hat{\alpha}_1 - 1)(F_2 + D_2 - H_1 - G_1) + E_2 - (\hat{\alpha}_1 - 1)C_2 - \hat{\alpha}_1 B_2 - 2L_1 - K_1 + M_1 + J_1 + \hat{\alpha}_1\zeta\} - \hat{\beta}^*$$

where $\hat{\beta}^* = \left(\frac{\bar{Y}^2 + \lambda S_y^2}{\bar{Y}} + \bar{Y}\right)$; $\hat{\alpha}^* = \frac{\bar{Y}^3}{\bar{Y}^2 + \lambda S_y^2}$.

$$MSE(\hat{T}_s) = \hat{\gamma}^2 + \alpha^{*2}\left[A_2 + \theta C_2 + \theta A_{11} + \theta C_{11} + 4\alpha_1^2 B_2 + 2\phi F_2 - 2\phi G_1 - 2\phi H_1 + 4\alpha_1 D_2 - 2\theta K_1 + 2\theta L_1 + 4\alpha_1\phi E_2 + 2\theta M_1 + 4\alpha_1\phi I_1 - 4\alpha_1\phi J_1\right]$$

where $\hat{\gamma} = \left(\frac{-(\lambda S_y^2)\bar{Y}^2}{\bar{Y}^2 + \lambda S_y^2}\right)$.

which is optimum when

$$\hat{\alpha}_{1opt.} = \frac{-(\hat{\gamma}^2 + 4\eta)}{8\alpha^{*2}\eta^*}$$

where $\hat{\eta} = F_2 - G_1 - H_1 + D_2$;

and $\hat{\eta}^* = B_2 + C_2 + A_{11} + C_{11} - 2K_1 + 2L_1 + 2M_1 + 2E_2 + 2I_1 - 2J_1$.

The minimum mean squared error for this methodology is given as

$$min.MSE(\hat{T}_s) = \hat{\gamma}^2 + \hat{\alpha}^{*2}(A_2 + C_2 + A_{11} + C_{11} - 2K_1 + 2L_1 + 2M_1 - 2F_2 + 2G_1 + 2H_1) - \hat{\psi}(D_2 - F_2 + G_1 + H_1 - B_2) + \hat{\psi}^*$$

where $\hat{\psi} = \frac{-(\hat{\gamma}^2+4\hat{\eta})}{2}$; $\hat{\psi}^* = \frac{\hat{\psi}^2}{\hat{\alpha}^{*2}\hat{\eta}^*}$.

**Methodology 3:** Suppose $\bar{y}^*$, $\bar{x}^*$, $\bar{z}^*$, $\bar{x}'$ and $\bar{z}'$ are employed and there is presence both non-response and measurement error on the sensitive study and auxiliary variables i.e. $(X,Z)$ using TPS technique. For this strategy the chain ratio type estimator is given as

$$T_s^* = \alpha\bar{y}^*\left[\alpha_1\left(\frac{\bar{x}^*}{\bar{x}'}\right)\left(\frac{\bar{z}^*}{\bar{z}'}\right) + (1-\alpha_1)\left(\frac{\bar{x}'}{\bar{x}^*}\right)\left(\frac{\bar{z}'}{\bar{z}^*}\right)\right]$$

where $\alpha = \frac{\bar{Y}^2}{\bar{Y}^2 + \lambda S_y^2 + \lambda^* S_{y(2)}^2}$.

The formulation of bias and MSE when there is a presence of non-response and measurement error are given as

$$Bias(T_s^*) = \alpha^*\{(2\alpha_1 - 1)(F_1 + D_1 - H_1 - G_1) + E_1 - (\alpha_1 - 1)C_3 - \alpha_1 B_3 - 2L_1 - K_1 + M_1 + J_1 + \alpha_1\zeta\} - \beta^*$$

and

$$MSE(T_s^*) = \gamma^2 + \alpha^{*2}\left[A_3 + \theta C_3 + \theta A_{11} + \theta C_{11} + 4\alpha_1^2 B_3 + 2\phi F_1 - 2\phi G_1 - 2\phi H_1 + 4\alpha_1 D_1 - 2\theta K_1 + 2\theta L_1 + 4\alpha_1\phi E_1 + 2\theta M_1 + 4\alpha_1\phi I_1 - 4\alpha_1\phi J_1\right]$$

which in itself is optimal when

$$\alpha^*_{1opt.} = \frac{-(\gamma^2 + 4\eta)}{8\alpha^{*2}\eta^{**}}$$

where $\eta^{**} = B_3 + C_3 + A_{11} + C_{11} - 2K_1 + 2L_1 + 2M_1 + 2E_1 + 2I_1 - 2J_1$.

Then, the min.MSE for this strategy is expressed as

$$min.MSE(T_s^*) = \gamma^2 + \alpha^{*2}(A_3 + C_3 + A_{11} + C_{11} - 2K_1 + 2L_1 + 2M_1 - 2F_1 + 2G_1 + 2H_1) - \psi(D_1 - F_1 + G_1 + H_1 - B_3) + \psi^{**}$$

where $\psi^{**} = \frac{\psi^2}{\alpha^{*2}\eta^{**}}$.

## 5. Efficiency Comparisons of Estimator(s)

To assess the effectiveness of the chain type proposed estimator, we relate it to the ratio and product estimator in different strategic plans as

$(i)$ $\quad min.MSE(\hat{T}_s^*) - MSE(\hat{T}_r^*) < 0$

if $\quad \gamma^2 + \alpha^{*2}\hat{a}^* + \psi^* - \psi\hat{b}^* - \hat{c}^* < 0$

$(ii)\quad min.MSE(\hat{T}_s) - MSE(\hat{T}_r) < 0$

if $\quad \hat{\gamma}^2 + \hat{\alpha}^{*2}\hat{a} + \psi^* - \psi\hat{b} - \hat{c} < 0$

$(iii)\quad min.MSE(T_s^*) - MSE(T_r^*) < 0$

if $\quad \gamma^2 + \alpha^{*2}a^* + \psi^* - \psi b^* - c^* < 0$

$(iv)\quad min.MSE(\hat{T}_s^*) - MSE(\hat{T}_p^*) < 0$

if $\quad \gamma^2 + \alpha^{*2}\hat{a}^* + \psi^* - \psi\hat{b}^* - \hat{d}^* < 0$

$(v)\quad min.MSE(\hat{T}_s) - MSE(\hat{T}_p) < 0$

if $\quad \hat{\gamma}^2 + \hat{\alpha}^{*2}\hat{a} + \psi^* - \psi\hat{b} - \hat{d} < 0$

$(vi)\quad min.MSE(T_s^*) - MSE(T_p^*) < 0$

if $\quad \gamma^2 + \alpha^{*2}a^* + \psi^* - \psi b^* - d^* < 0$

where $\hat{a}^* = A_1 + C_1 + A_{11} + C_{11} - 2K_1 + 2L_1 + 2M_1 - 2F_1 + 2G_1 + 2H_1$;

$\hat{b}^* = D_1 - F_1 + G_1 + H_1 - B_1$;

$\hat{c}^* = A + \varkappa^2 B + \varkappa^2 A_{11} - 2\varkappa^2 I + 2\varphi G + 2\varphi D + (\kappa_1 + \varkappa^2\kappa_2)$;

$\hat{a} = A_2 + C_2 + A_{11} + C_{11} - 2K_1 + 2L_1 + 2M_1 - 2F_2 + 2G_1 + 2H_1$;

$\hat{b} = D_2 - F_2 + G_1 + H_1 - B_2$;

$\hat{c} = \hat{A} + \varkappa^2\hat{B} + \varkappa^2 A_{11} - 2\varkappa^2 I + 2\varphi G + 2\varphi\hat{D} + (\kappa_1 + \varkappa^2\kappa_2)$;

$a^* = A_3 + C_3 + A_{11} + C_{11} - 2K_1 + 2L_1 + 2M_1 - 2F_1 + 2G_1 + 2H_1$;

$b^* = D_1 - F_1 + G_1 + H_1 - B_3$;

$c^* = A + \varkappa^2 B + \varkappa^2 A_{11} - 2\varkappa^2 I + 2\varphi G + 2\varphi D + (\kappa_1 + \varkappa^2\kappa_2)$;

When the above conditions from $(i) - (vi)$ are met then it is evident that the suggested estimators i.e. $\hat{T}_s^*$, $\hat{T}_s$ and $T_s^*$ are efficient than the existing one.

$min.MSE(\hat{T}_s) < min.MSE(T_s^*) < min.MSE(\hat{T}_s^*) < MSE(\hat{T}_p) < MSE(T_p^*) < MSE(\hat{T}_p^*) < MSE(\hat{T}_r) < MSE(T_r^*) < MSE(\hat{T}_p^*)$.

To verify the performance of the above relations, we execute a simulation study by using both hypothetical and real populations in R software which is relinquished in the next section.

## 6. Simulation Study

To gain a better understanding of the efficiency of the recommended estimators, we leverage R software to perform a simulation study to validate the effectiveness of our proposed estimator as compare to the ratio and the product type estimator(s). We generated a population of $N = 8000$ we take sample of size $n' = 6000$ and suppose that the response rate is 40% in the first phase. From $n'$ we take sample of size $n = 2000$ using R software for different values of $k$ and $\pi$ sequentially. A variable $X \sim N(a,b)$; $Z \sim N(a,b)$ and variable $Y$ which is related with $X$ and $Z$ is defined as $Y = X + Z + N(0,1)$ also generated from normal distribution where $a = 0.5$ and $b = 1.5$. The scrambling variables $S \sim N(1,a)$ and $S' \sim N(0,1)$, both taken from normal distribution and results are averaged over $8,000$ iterations.

The unified measure $\omega$ as described by Gupta et al. [21] are represented by

$$\hat{\omega}^* = \frac{MSE(\hat{T}_i^*)}{\Upsilon}; \tag{6.1}$$

where $\Upsilon = E(Z_1 - Y)^2$ is the privacy level of sensitive models and $T_i^* = \hat{T}_r^*$, $\hat{T}_p^*$ and $\hat{T}_s^*$.

$$\hat{\omega} = \frac{MSE(\hat{T}_i)}{\Upsilon}; \tag{6.2}$$

where $T_i^* = \hat{T}_r$, $\hat{T}_p$ and $\hat{T}_s$.

| $\pi$ | $k$ | *Estimator(s)* | | | *Unified Measure*$(\hat{\omega}^*)$ | | |
|---|---|---|---|---|---|---|---|
| | | $MSE(\hat{T}_r^*)$ | $MSE(\hat{T}_p^*)$ | $MSE(\hat{T}_s^*)$ | $\hat{\omega}^*(\hat{T}_r^*)$ | $\hat{\omega}^*(\hat{T}_p^*)$ | $\hat{\omega}^*(\hat{T}_s^*)$ |
| 0.2 | 2 | 0.0747 | 0.0601 | 0.0255 | 0.0254 | 0.0205 | 0.0087 |
| | 3 | 0.0963 | 0.0767 | 0.0326 | 0.0328 | 0.0261 | 0.0111 |
| | 4 | 0.1189 | 0.0974 | 0.0404 | 0.0406 | 0.0332 | 0.0138 |
| | 5 | 0.1403 | 0.1131 | 0.0475 | 0.0479 | 0.0385 | 0.0162 |
| 0.4 | 2 | 0.0755 | 0.0616 | 0.0262 | 0.0248 | 0.0202 | 0.0086 |
| | 3 | 0.0977 | 0.0790 | 0.0339 | 0.0321 | 0.0260 | 0.0111 |
| | 4 | 0.1210 | 0.1006 | 0.0423 | 0.0398 | 0.0331 | 0.0139 |
| | 5 | 0.1429 | 0.1159 | 0.0494 | 0.0470 | 0.0381 | 0.0162 |
| 0.6 | 2 | 0.0757 | 0.0656 | 0.0286 | 0.0246 | 0.0213 | 0.0092 |
| | 3 | 0.0983 | 0.0843 | 0.0371 | 0.0319 | 0.0274 | 0.0121 |
| | 4 | 0.1220 | 0.1050 | 0.0455 | 0.0396 | 0.0343 | 0.0148 |
| | 5 | 0.1440 | 0.1229 | 0.0537 | 0.0468 | 0.0399 | 0.0174 |
| 0.8 | 2 | 0.0764 | 0.0671 | 0.0294 | 0.0240 | 0.0210 | 0.0092 |
| | 3 | 0.0995 | 0.0867 | 0.0385 | 0.0312 | 0.0272 | 0.0121 |
| | 4 | 0.1230 | 0.1087 | 0.0474 | 0.0388 | 0.0341 | 0.0149 |
| | 5 | 0.1464 | 0.1257 | 0.0557 | 0.0459 | 0.0395 | 0.0174 |
| 1 | 2 | 0.0837 | 0.0709 | 0.0298 | 0.0256 | 0.0216 | 0.0091 |
| | 3 | 0.1109 | 0.0911 | 0.0386 | 0.0339 | 0.0278 | 0.0118 |
| | 4 | 0.1351 | 0.1106 | 0.0464 | 0.0413 | 0.0338 | 0.0141 |
| | 5 | 0.1613 | 0.1314 | 0.0557 | 0.0493 | 0.0401 | 0.0170 |

**Table 2.** Comparison of Mean squared error and privacy and efficiency $(\hat{\omega}^*)$ of $\hat{T}_s^*$, $\hat{T}_r^*$ and $\hat{T}_p^*$ at varying values of $k$ and $\pi$ with non-response and measurement error.

Table 1 delineates the comparison of mean squared error of the suggested estimator $\hat{T}_s^*$ with other conventional estimators i.e. $\hat{T}_r^*$ and $\hat{T}_p^*$ and privacy protection measure suggested by Gupta et al. [21] which is represented in (6.1) at distinct values of $k$ and $\pi$ in the presence of non-response and measurement error at the same time under TPS technique. For increase in the value of $\pi$ from 0.2 to 1 and $k$ from 2 to 5, the mean squared error of each estimator grows and same behaviour is observed for the unified measure $(\hat{\omega}^*)$.

Table 2 depicts the comparison of mean squared error of the suggested estimator $\hat{T}_s$ with other existing estimators i.e. $\hat{T}_r$ and $\hat{T}_p$ and privacy protection measure which is represented in (6.2) at distinct values of $k$ and $\pi$ in the absence of non-response and measurement error. The mean squared error of each estimator increases with increase in the value of $\pi$ from 0.2 to 1 and $k$ from 2 to 5, and same performance is detected for the privacy protection $(\hat{\omega}^*)$.

It is also visualize from Tables 1 and Table 2 that the MSEs of ratio estimators ( $\hat{T}_r^*$, $\hat{T}_r$) and product estimators ($\hat{T}_p^*$, $\hat{T}_p$) are the highest for all analyzed values of $k$, whereas our recommended estimators, ($\hat{T}_s^*$, $\hat{T}_s$) is the lowest among the ratio and the product type estimators. Also, the privacy measure is least for the proposed estimator ($\hat{T}_s^*$, $\hat{T}_s$) in the presence and absence of non-response and measurement error simultaneously. In both the scenario‘s, ($MSE(\hat{T}_s^*)$, $MSE(\hat{T}_s)$), i.e. the recommended estimator, is the most efficient amongst the alternatives. Furthermore, Table 1 and Table 2 indicates that the proposed estimator outperformed existing estimators also in terms of the unified measure ($\hat{\omega}^*$ and $\hat{\omega}$) of privacy and efficiency.

Table 3 illustrates the comparison of mean squared error of the suggested estimator $T_s^*$ with other existing estimators i.e. $T_r^*$ and $T_p^*$ at specific values of $k$ in the absence of non-response and measurement error entirely at the same time. When the value

| $\pi$ | $k$ | *Estimator(s)* | | | *Unified Measure*($\hat{\omega}$) | | |
|---|---|---|---|---|---|---|---|
| | | $MSE(\hat{T}_r)$ | $MSE(\hat{T}_p)$ | $MSE(\hat{T}_s)$ | $\hat{\omega}(\hat{T}_r)$ | $\hat{\omega}(\hat{T}_p)$ | $\hat{\omega}(\hat{T}_s)$ |
| 0.2 | 2 | 0.0209 | 0.0269 | 0.0080 | 0.0254 | 0.0092 | 0.0027 |
| | 3 | 0.0217 | 0.0274 | 0.0082 | 0.0328 | 0.0093 | 0.0028 |
| | 4 | 0.0226 | 0.0279 | 0.0086 | 0.0406 | 0.0095 | 0.0029 |
| | 5 | 0.0233 | 0.0283 | 0.0088 | 0.0479 | 0.0096 | 0.0030 |
| 0.4 | 2 | 0.0227 | 0.0285 | 0.0087 | 0.0248 | 0.0094 | 0.0028 |
| | 3 | 0.0243 | 0.0294 | 0.0093 | 0.0321 | 0.0097 | 0.0030 |
| | 4 | 0.0261 | 0.0306 | 0.0100 | 0.0398 | 0.0100 | 0.0033 |
| | 5 | 0.0275 | 0.0313 | 0.0105 | 0.0470 | 0.0103 | 0.0034 |
| 0.6 | 2 | 0.0240 | 0.0311 | 0.0099 | 0.0246 | 0.0101 | 0.0032 |
| | 3 | 0.0263 | 0.0325 | 0.0108 | 0.0319 | 0.0105 | 0.0035 |
| | 4 | 0.0289 | 0.0341 | 0.0118 | 0.0390 | 0.0110 | 0.0038 |
| | 5 | 0.0310 | 0.0354 | 0.0126 | 0.0468 | 0.0115 | 0.0040 |
| 0.8 | 2 | 0.0257 | 0.0328 | 0.0107 | 0.0240 | 0.0103 | 0.0033 |
| | 3 | 0.0287 | 0.0347 | 0.0120 | 0.0312 | 0.0109 | 0.0037 |
| | 4 | 0.0321 | 0.0368 | 0.0133 | 0.0388 | 0.0115 | 0.0041 |
| | 5 | 0.0348 | 0.0384 | 0.0143 | 0.0459 | 0.0120 | 0.0044 |
| 1 | 2 | 0.0294 | 0.0348 | 0.0116 | 0.0256 | 0.0106 | 0.0035 |
| | 3 | 0.0335 | 0.0371 | 0.0129 | 0.0339 | 0.0113 | 0.0039 |
| | 4 | 0.0372 | 0.0392 | 0.0142 | 0.0413 | 0.0119 | 0.0043 |
| | 5 | 0.0410 | 0.0416 | 0.0158 | 0.0493 | 0.0127 | 0.0048 |

**Table 3.** Comparison of Mean squared error and privacy and efficiency ($\hat{\omega}$) of $\hat{T}_r$, $\hat{T}_p$ and $\hat{T}_s$ at varying values of $k$ and $\pi$ without non-response and measurement error.

| $k$ | *Estimator(s)* | | |
|---|---|---|---|
| | $MSE(T_r^*)$ | $MSE(T_p^*)$ | $MSE(T_s^*)$ |
| 2 | 0.0556 | 0.0529 | 0.0205 |
| 3 | 0.0692 | 0.0659 | 0.0254 |
| 4 | 0.0849 | 0.0815 | 0.0301 |
| 5 | 0.0960 | 0.0926 | 0.0347 |

**Table 4.** Comparison of Mean squared error of $T_r^*$, $T_p^*$ and $T_s^*$ at varying values of $k$ with complete non-response and measurement error.

of $k$ tends to increase, the mean squared error of each estimator also increases. The MSE of the suggested estimator i.e. $T_s^*$ is minimal as the MSEs of the conventional one viz $T_r^*$ and $T_p^*$ are highest.

## 6.1 Natural population data set

The natural population dataset is based on abortion rates form Statistical Abstract of the United States: 2011 to elucidate the efficacious performance of our proposed estimator. The data is of $N = 51$ states and union territories of United States then a random sample is drawn from the population i.e., $n' = 20$. From $n'$ we take sample of size $n = 12$. Let $y$, $x$, $z$ be the number of abortions reported in the state of US during the years 2000, 2004, and 2005 respectively. The results are shown in Table 5 for different probability levels of sensitive variables, i.e. $\pi = 0.2, 0.4, 0.6, 0.8, 1$ when $k = 2$.

| *Estimator(s)* | $\pi$ | | | | |
|---|---|---|---|---|---|
| (*Unified Measure*) | 0.2 | 0.4 | 0.6 | 0.8 | 1 |
| $MSE(\hat{T}_r^*)$ $(\hat{\omega}^*(\hat{T}_r^*))$ | 0.1500161 (0.0002191) | 0.1500946 (0.0002192) | 0.1501633 (0.0002193) | 0.1502234 (0.0002194) | 0.1502760 (0.0002195) |
| $MSE(\hat{T}_p^*)$ $(\hat{\omega}^*(\hat{T}_p^*))$ | 0.1550136 (0.0002264) | 0.1551421 (0.0002266) | 0.1552544 (0.0002267) | 0.1553527 (0.0002269) | 0.1554388 (0.0002270) |
| $MSE(\hat{T}_s^*)$ $(\hat{\omega}^*(\hat{T}_p^*))$ | 0.0346015 (0.0000505) | 0.0345727 (0.0000505) | 0.0345476 (0.0000504) | 0.0345256 (0.0000505) | 0.03450639 (0.0000504) |
| $MSE(\hat{T}_r)$ $(\hat{\omega}(\hat{T}_r))$ | 0.1492835 (0.0002180) | 0.1493620 (0.0002181) | 0.1494306 (0.0002182) | 0.1494907 (0.0002183) | 0.1495434 (0.0002184) |
| $MSE(\hat{T}_p)$ $(\hat{\omega}(\hat{T}_p))$ | 0.1540969 (0.0002251) | 0.1542253 (0.0002252) | 0.1543376 (0.0002254) | 0.1544359 (0.0002256) | 0.1545220 (0.0002257) |
| $MSE(\hat{T}_s)$ $(\hat{\omega}(\hat{T}_s))$ | 0.0348155 (0.0005085) | 0.0347863 (0.0005081) | 0.0347609 (0.0005077) | 0.0347386 (0.0005074) | 0.0347191 (0.0005071) |

**Table 5.** Comparison of Mean squared error and unified measure at varying values of $\pi$ when $k = 2$ and $(MSE(T_r^*) = 0.081953$, $MSE(T_p^*) = 0.088362$ & $MSE(T_s^*) = 0.057562)$

Table 5 represents the comparison of mean squared error and unified measure of the proposed estimator i.e., ($\hat{T}_s^*$, $\hat{T}_s$ and $T_s^*$) with other existing estimators i.e. ($\hat{T}_r^*$, $\hat{T}_r$ and $T_r^*$) and ($\hat{T}_p^*$, $\hat{T}_p$ and $T_p^*$) at specific values of $\pi$ in the presence and absence of non-response and measurement error simultaneously. When the value of $\pi$ increases, the mean squared error and unified measure of existing estimators also increases but the mean squared error and unified measure of proposed estimator decreases. The MSE of proposed estimator is lowest and unified measure is highest which finds that the proposed estimator is better and each respondent privacy is protected as compared to the other existing estimators.

## 7. Conclusion

This study demonstrates a new chain ratio type estimator for estimating the population mean of the sensitive study as well as auxiliary variables in the presence of non-response and measurement error under two-phase sampling technique by utilizing ORRT models. The bias and mean squared errors of the proposed estimator are assessed up to the first order approximation. The efficiency of the proposed chain ratio type estimator has been compared with that of the existing one under TPS using two auxiliary variables. The condition by which the proposed estimator $\hat{T}_s^*$ proven to be more efficient than other existing estimators, notably $\hat{T}_r^*$ and $\hat{T}_p^*$ are also formed. The theoretical facts have been supported by conducting an empirical study. We executed a model-based simulation and a real dataset in R software to verify the theoretical results, and from the simulation results i.e., both hypothetical and real population shows that the suggested estimator outperform the other conventional estimators. As a result, if the requirements in Section 5 are satisfied, then the suggested estimators are encouraged for use in practice.

## Article Information

**Acknowledgements:** The authors would like to express their sincere thanks to the editor and the anonymous reviewers for their helpful comments and suggestions.

**Author's contributions:** All authors contributed equally to the writing of this paper. All authors read and approved the final manuscript.

**Conflict of Interest Disclosure:** No potential conflict of interest was declared by the author.



**Supporting/Supporting Organizations:** No grants were received from any public, private or non-profit organizations for this research.

**Ethical Approval and Participant Consent:** It is declared that during the preparation process of this study, scientific and ethical principles were followed and all the studies benefited from are stated in the bibliography.

**Plagiarism Statement:** This article was scanned by the plagiarism program. No plagiarism detected.

**Availability of data and materials:** Not applicable.

## References


[1] S. L. Warner, *Randomized response: a survey technique for eliminating evasive answer bias*, J. Am. Stat. Assoc., **60**(309) (1965), 63–69.

[2] B. G. Greenberg, R. R. Jr. Kuebler, J. R. Abernathy, D. G. Hovertiz, *Application of the randomized response techniques in obtaining quantitative data*, J. Am. Stat. Assoc., **66**(334) (1971), 243-250.

[3] K. Pollock, Y. Bek, *A comparison of three randomized response models for quantitative data*, J. Am. Stat. Assoc., **71**(356) (1976), 884-886.

[4] S. Gupta, B. Gupta, S. Singh, *Estimation of sensitivity level of personal interview survey questions*, J. Stat. Plan. Inference, **100**(2) (2002), 239-247.

[5] S. Gupta, J. Shabbir, S. J. Sehra, *Mean and sensitivity estimation in optional randomized response models*, J. Stat. Plan. Inference, **140**(10) (2010), 2870-2874.

[6] S. Gupta, J. Shabbir, R. Sousa, P. Corte-Real, *Estimation of the mean of a sensitive variable in the presence of auxiliary information*, Commun. Stat. Theory Methods, **41**(13-14) (2012), 13-14.

[7] Q. Zhang, G. Kalucha, S. Gupta, S. Khalil, *Ratio estimation of the mean under RRT models*, Int. J. Stat. Manag. Syst., **22**(1) (2018), 97-113.

[8] S. Kumar, S. P. Kour, *Estimation of Sensitive Variable in Two-Phase Sampling under Measurement Error And Non-Response Using ORRT Models*, Sri Lankan J. Appl. Stat., **22**(3) (2021), 95-122.

[9] S. Kumar, S. P. Kour, *The joint influence of estimation of sensitive variable under measurement error and non-response using ORRT models*, J. Stat. Comput. Simul., **92**(17) (2022), 3583-3604.

[10] S. Kumar, S. P. Kour, Q. Zhang *An enhanced ratio-cum-product estimator with non-response and observational error by utilizing ORRT models: a sensitive estimation approach*, J. Stat. Comput. Simul., **93**(5) (2023), 818-836.

[11] S. Kumar, S. P. Kour, R. Gupta, J. P. S. Joorel *A Class of Logarithmic Type Estimator Under Non-Response and Measurement Error Using ORRT Models*, J. Indian Soc. Probab. Stat., (2023), doi:10.1007/s41096-023-00156-7.

[12] Q. Zaman, M. Ijaz, T. Zaman, *A randomization tool for obtaining efficient estimators through focus group discussion in sensitive surveys*, Commun. Stat. - Theory Methods, **52**(10) (2023), 3414-3428.

[13] J. Neyman, *Contribution to the theory of sampling human populations*, J Am Stat Assoc., **33**(201) (1938), 101-116.

[14] A. Sanaullah, H. Ali, M. Noor-ul-Amin, M. Hanif, *Generalized exponential chain ratio estimators under stratified two-phase random sampling*, Appl. Math. Comput., **226** (2014), 541-547.

[15] T. Zaman, C. Kadilar, *New class of exponential estimators for finite population mean in two-phase sampling*, Commun. Stat. Theory Methods, **50**(4) (2021), 874-889.

[16] S. Khalil, Q. Zhang, S. Gupta, *An enhanced two phase sampling ratio estimator for estimating population mean*, J. Sci. Res. **65**(3) (2021), 1-16.

[17] S. Khalil, Q. Zhang, S. Gupta, *Mean estimation of sensitive variables under measurement errors using optional RRT models*, Commun. Stat. Simul. Comput., **50**(5) (2021), 1417-1426.

[18] R. Onyango, B. Oduor, F. Odundol, *Joint influence of measurement errors and randomized response technique on mean estimation under stratified double sampling*, Open J. Math. Sci., **5**(1) (2021), 192-199.

[19] M. H. Hansen, W. N. Hurwitz, *The problem of non-response in sample surveys*, J. Am. Stat. Assoc., **41**(236) (1946), 517-529.

[20] G. Diana, S. Riaz, J. Shabbir, *Hansen and Hurwitz estimator with scrambled response on the second call*, J. Appl. Stat., **41**(3) (2014), 596-611.

[21] S. Gupta, S. Mehta, J. Shabbir, S. Khalil, *A unified measure of respondent privacy and model efficiency in quantitative RRT models*, J. Stat. Theory Pract., **12**(3) (2018), 506-511.

[22] Q. Zhang, S. Khalil, S. Gupta, *Ratio estimation of the mean under RRT models*, J. Stat. Theory Pract., **15**(3) (2021), 97-113.

[23] P. Mukhopadhyay, G. N. Singh, A. Bandyopadhyay, *A general estimation technique of population mean under stratified successive sampling in presence of random scrambled response and non-response*, Commun. Stat. Simul. Comput, **50**(5) (2021), 1417-1426.

[24] G. Diana, P. F. Perri, *A class of estimators for quantitative sensitive data*, Stat. Pap., **52**(3) (2011), 633-650.

[25] S. Khalil, M. Amin, M. Hanif, *Estimation of population mean for a sensitive variable in the presence of measurement error*, Int. J. Stat. Manag. Syst., **21**(1) (2018), 81-91.

[26] Z. Yan, J. Wang, J. Lai, *An efficiency and protection degree-based comparison among the quantitative randomized response strategies*, Commun. Stat. - Theory Methods, **38**(3) (2008), 400-408.