Brazilian Journal of Biometrics, **43**, 1–16 (2025)
doi:10.28951/bjb.v43i2.730

BRAZILIAN JOURNAL OF BIOMETRICS
ISSN:2764-5290

ARTICLE

Advanced Optional Randomized Response Models For Mean Estimation Under Non-Response And Measurement Error

Sunil Kumar[1] and Sanam Preet Kour[⋆,2]

[1]Department of Statistics, University of Jammu, Jammu & Kashmir, India.
[2]Department of Clinical Research, Sharda School of Allied Health Sciences, Sharda University, Greater Noida, Uttar Pradesh, India.
[⋆]Corresponding author. Email: sanamkour903@gmail.com

(Received: March 22, 2024; Revised: August 7, 2024; Accepted: September 17, 2024; Published: April 3, 2025)

Abstract

A regression estimator for estimating the population mean of sensitive variable(s) in the presence of non-response and measurement error simultaneously using two scrambling variables are introduced. Comparisons are made with the mean squared error of the proposed estimator with some of the commonly used estimators i.e. Hansen and Hurwitz estimator, linear regression estimator and Diana et al. estimator. Under large sample approximation, their biases and mean square errors are estimated. In addition, an extensive simulation study with real and hypothetical population are also conducted to evaluate the performance of proposed estimator which show that these estimators perform better than the other considered estimators. A graphically representations are also used to represent the simulation results.

Keywords: Sensitive variable(s); Measurement error; Non-response; Optional Randomized Response Model; Bias; Mean square error.

1. Introduction

Equivocating direct questioning about intimate and confidential matters including gambling, alcoholism, abortion, drug usage, tax evasion, illegal income and so on can cause respondents humiliation or fear of societal condemnation. Even if interviewers make every effort to maintain confidentially, interviewees may remain distrustful or reluctant to provide accurate responses. Survey statisticians have established a lot of approaches to secure interviewee anonymity or, at the very least, a high degree of survey, in order to reduce non-response and minimize under-reporting of embarrassing, threatening or even convicting behaviors.

© Brazilian Journal of Biometrics. This is an open access article distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/)

To prevail trustworthy data on stigmatizing characteristics, Warner (1965) proposed an interviewing strategy, known as Randomized response technique in which the respondents choose one of the two complementary questions at random on a probability basis to reduce Social desirability response bias without exposing respondents privacy. Greenberg *et al.* (1971) relied on Warner (1965) study by gathering information on quantitative stigmatizing factors. Since then, numerous authors, notably Eichhorn & Hayre (1983), Gupta *et al.* (2002), Gupta & Shabbir (2004), Gupta *et al.* (2006), Hussain & Shabbir (2009), Zhang *et al.* (2018), Mushtaq & Amin (2020), Kumar & Kour (2022), Kumar *et al.* (2023) and Kumar *et al.* (2024) and others have worked on quantitative randomized response models. In addition to Social desirability response bias, there are several additional non-sampling flaws that can impact population mean estimate, including measurement error, which is the discrepancy between the real value of the variable being monitored and its recorded value. In the presence of measurement error, Khalil *et al.* (2018) and Khalil *et al.* (2021) investigated mean estimate of sensitive variable using auxiliary information under optional randomized response models.

One typical non-sampling problem that researchers must cope with is non-response. After the first call Hansen & Hurwitz (1946) were among the first to advise collecting a sub-sample of non-respondents, and then conducting a personal interview with this group to acquire information. Kumar *et al.* (2011), Yaqub *et al.* (2017), Guha & Chandra (2019), Unal & Kadilar (2021), Kumar & Kour (2023), Kumar *et al.* (2024) and Kumar *et al.* (2025) all have analyzed mean estimate in the context of non-response by utilizing information on auxiliary variable. While Hansen & Hurwitz (1946) methodology might provide additional information from face-to-face interview in the second stage, it could also yield in social desirability response bias if the variable of interest is hypersensitive. In face-to-face interview, respondents are unwilling to give honest answers to such questions. When we target the group of non-respondents in the second stage, we might utilize randomized response technique to mitigate the social desirability response bias caused by sensitive questions. Diana *et al.* (2014), Gupta *et al.* (2018) and others used scrambles response to answer the sensitive questions directly.

Taking motivation from the earlier research on optional randomized response models, the present study addresses a hybrid regression estimator for the estimation of sensitive variable by utilizing optional randomized response models in the presence of non-response and measurement error simultaneously. The paper is organized as follows: section 2 discusses a optional randomized response technique, while ections 3 examine revised Hansen & Hurwitz (1946) technique and existing estimators for the estimation of sensitive variables using the ORRT model. In section 4, there are existing scrambled models. The proposed scrambled estimator and its properties are explored in section 5. Simulation study with hypothetical and real population are presented in section 6 and 7, where conditions are identified under which the proposed estimator outperforms existing ones. Finally, Section 8 offers concluding remarks.

2. Optional Randomized Response Technique

Consider Y and X be positively correlated sensitive variable(s) with unknown mean $\bar{Y}$ and $\bar{X}$, and unknown variance S_y^2 and S_x^2. Let S_1 and S_2 be two scrambling variables with mean $\bar{S}_1$ and $\bar{S}_2$, and known variances $S_{S_1}^2$ and $S_{S_2}^2$, respectively. Let W represents the probability that the respondent consider the question sensitive. If the respondent thinks the question is sensitive, he or she is requested to report a scrambled response for study and auxiliary variable (X, Y) otherwise a valid response reported.

A basic additive randomized response model with $Y + S_2$ as the scrambled response (as in Gupta *et al.* (2012)) or a more general randomized response model with $S_1 Y + S_2$ as the scrambled response (as in Diana & Perri (2011)). If $Var(S_1) = 0$ and $E(S_1) = 1$ then the simple additive model is a special instance of the second model. The basic additive approach is more efficient, but the general model provides better privacy, according to Khalil *et al.* (2018).

On the other hand, the generalized randomized response model performs better when we use Gupta *et al.* (2018) combined measure of efficiency and privacy i.e $\gamma = \frac{Var(Z_1)}{\Gamma}$, where Z_1 is the scrambled response and $\Gamma = E(Z_1 - Y)^2$ is the privacy level for the same model, as provided by Yan *et al.* (2008). It should be noted that the model with a smaller value is preferable because it indicates either a higher level of privacy or a lower value of $Var(\hat{y})$, or both. It should be noted that

$$\gamma_{additiveRRT} = 1 + \frac{S_y^2}{S_{S_2}^2} > 1 + \frac{S_y^2}{S_{S_2}^2 + S_{S_1}^2(\bar{y} + S_y^2)} = \gamma_{generalRRT} \tag{1}$$

As a result, while dealing with the general randomized response model, the scrambling variable S_1 will reduce model efficiency while increasing privacy. Overall, the general model outperforms the specific approach in terms of efficiency and privacy. Therefore, in this investigation, one will utilize the general scrambling model. The optional version of model $Z_1 = S_1Y + S_2$ is given by

$$Z_1 = \begin{cases} Y & \text{with probability } 1 - W \\ S_1Y + S_2 & \text{with probability } W, \end{cases} \tag{2}$$

where S_1 and S_2 follows Normal distribution with mean $(1, 0)$ and variances $(S_{S_1}^2, S_{S_2}^2)$ i.e. $S_1 \sim N(1, S_{S_1}^2)$ and $S_2 \sim N(0, S_{S_2}^2)$. The mean and variance of Z_1 are given by

$E(Z_1) = E(Y)(1 - W) + E(S_1Y + S_2)W = E(Y)$

and $\text{Var}(Z_1) = E(Z_1^2) - E^2(Z_1) = S_y^2 + S_{S_2}^2 W + S_{S_1}^2(S_y^2 + \bar{Y}^2)W$.

From here, the randomized linear model is given as $Z_1 = (S_1Y + S_2)J + Y(1 - J)$, where $J \sim \text{Bernoulli}(W)$ with $E(J) = W, \text{Var}(J) = W(1 - W)$ and $\text{E}(J^2) = \text{Var}(J) + E^2(J) = W$. And the expectation and variance of randomized mechanism is $E_R(Z_1) = (\bar{S}_1W + 1 - W)Y + \bar{S}_2W$ and $\text{V}_R(Z_1) = (Y^2S_{S_1}^2 + S_{S_2}^2)W$.

Since the variance of Z_1 increases as W increases, the optional randomized response model is obviously more efficient than the non-optional randomized response model. The randomized response model becomes non-optional when $W = 1$.

In present research, the auxiliary variable X to be a sensitive variable as well. For illustration, if income (Y) is a sensitive variable, but investments and expenditure (X) must also be sensitive variables. As a result, the general scrambling model for auxiliary variable is given as

$$Z_2 = \begin{cases} X & \text{with probability } 1 - W \\ S_1X + S_2 & \text{with probability } W, \end{cases} \tag{3}$$

where $S_1 \sim N(1, S_{S_1}^2)$, $S_2 \sim N(0, S_{S_2}^2)$ and W denotes the probability that a respondent find the question sensitive. The mean and variance of Z_2 are given by

$E(Z_2) = E(X)(1 - W) + E(S_1X + S_2)W = E(X)$

and $\text{Var}(Z_2) = E(Z_2^2) - E^2(Z_2) = S_x^2 + S_{S_2}^2 W + S_{S_1}^2(S_x^2 + \bar{X}^2)W$.

Similarly, the randomized linear model can be written as $Z_2 = (S_1X + S_2)J + X(1 - J)$, where $J \sim \text{Bernoulli}(W)$ with $E(J) = W, \text{Var}(J) = W(1 - W)$ and $E(J^2) = \text{Var}(J) + E^2(J) = W$. And the expectation and variance of randomized mechanism is $E_R(Z_2) = (\bar{S}_1W + 1 - W)X + \bar{S}_2W$ and $\text{V}_R(Z_2) = (X^2S_{S_1}^2 + S_{S_2}^2)W$.

The optional randomized response model is clearly more efficient than the non-optional randomized response model, as the variance of Z_2 increases with W increases. When $W = 1$, the randomized response model becomes non-optional.

3. Revised Hansen & Hurwitz (1946) Technique

Consider $\xi = \xi_1, \xi_2, ..., \xi_N$ be a finite population of size N and a random sample without replacement of size n is drawn from this population. Then only n_1 units provide a response on the first call and remaining $n_2 = n - n_1$ units do not respond. A sub sample of size $n_s = \frac{n_2}{k}$; $(k > 1)$ is taken from the n_2 non-responding units. Hansen & Hurwitz (1946) used mail survey at the first attempt and used face-to-face interview at the second call.

Let $\bar{Y} = \frac{\sum_{i=1}^{N} y_i}{N}$ and $S_y^2 = \frac{\sum_{i=1}^{N}(y_i - \bar{Y})^2}{N-1}$ be the population mean and variance of the study variable Y. Let $\bar{Y}_{(1)} = \frac{\sum_{i=1}^{N_1} y_i}{N_1}$ and $S^2_{y_{(1)}} = \frac{\sum_{i=1}^{N_1}(y_i - \bar{Y}_{(1)})^2}{N_1-1}$ respectively, be the population mean and variance of respondent group of size N_1, $\bar{Y}_{(2)} = \frac{\sum_{i=1}^{N_2} y_i}{N_2}$ and $S^2_{y_{(2)}} = \frac{\sum_{i=1}^{N_2}(y_i - \bar{Y}_{(2)})^2}{N_2-1}$ be the population mean and variance of non-respondent group of size N_2. Let $\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$ and $S_x^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})^2}{N-1}$ be the population mean and variance of the study variable X. Let $\bar{X} = \frac{\sum_{i=1}^{N_1} x_i}{N_1}$ and $S^2_{X_{(1)}} = \frac{\sum_{i=1}^{N_1}(x_i - \bar{X}_{(1)})^2}{N_1-1}$ be the population mean and variance of respondent group of size N_1, $\bar{X}_{(2)} = \frac{\sum_{i=1}^{N_2} y_i}{N_2}$ and $S^2_{X_{(2)}} = \frac{\sum_{i=1}^{N_2}(x_i - \bar{X}_{(2)})^2}{N_2-1}$ be the population mean and variance of non-respondent group of size N_2.

The overall population mean of study variable is given by

$$\bar{Y} = W_1 \bar{Y}_{(1)} + W_2 \bar{Y}_{(2)}, \tag{4}$$

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

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

Hansen & Hurwitz (1946) suggested an unbiased population mean estimator given by

$$\bar{y} = w_1 \bar{y}_1 + w_2 \bar{y}_{2s} \tag{5}$$

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

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

$$Var(\bar{y}) = \left(\frac{N-n}{Nn}\right) S_y^2 + \frac{W_2(k-1)}{n} S^2_{y(2)} \tag{6}$$

In the second phase of the Hansen & Hurwitz (1946) process, when there is face-to-face interview of sub-sampled units of non-respondents in first phase, one provide the respondents the chance to scramble their response using optional randomized response models to motivate respondents to answer a sensitive question honestly. In this instance, a modified Hansen and Hurwitz procedure by assuming that in the first phase, the respondent group provides direct responses, and then in the second phase, the optional randomized response model is utilized to collect responses 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 mechanism and is given by

$$\hat{y}_i = \frac{z_{1i} - \bar{S}_2 W}{\bar{S}_1 W + 1 - W} \tag{7}$$

with

$$E_R(\hat{y}_i) = y_i$$

and

$$\mathrm{V}_R(\hat{y}_i) = \frac{V_R(z_{1i})}{(\bar{S}_1 W+1-W)^2} = \frac{(y_i^2 S_{S_1}^2 + S_{S_2}^2)W}{(\bar{S}_1 W+1-W)^2} = \tau_{1i}$$

On the other hand, assume $\hat{x}_i$ denote a transformation of the randomized response on the i^{th} unit, the expectation of which is the true response x_i under the mechanism and is given by

$$\hat{x}_i = \frac{z_{2i} - \bar{S}_2 W}{\bar{S}_1 W + 1 - W} \tag{8}$$

with

$$E_R(\hat{x}_i) = x_i$$

and

$$V_R(\hat{x}_i) = \frac{V_R(z_{2i})}{(\bar{S}_1 W+1-W)^2} = \frac{(x_i^2 S_{S_1}^2 + S_{S_2}^2)W}{(\bar{S}_1 W+1-W)^2} = \tau_{2i}$$

On the basis of above discussion, a modified Hansen & Hurwitz (1946) estimator in the presence of non-response by using optional randomized response technique as

$$\hat{\bar{y}} = w_1\bar{y}_1 + w_2\hat{\bar{y}}_2 \tag{9}$$

$$\hat{\bar{x}} = w_1\bar{x}_1 + w_2\hat{\bar{x}}_2 \tag{10}$$

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

It is easy to verify that

$$E(\hat{\bar{y}}) = \bar{Y}; E(\hat{\bar{x}}) = \bar{X} \tag{11}$$

i.e. $\hat{\bar{y}}$ and $\hat{\bar{x}}$ are usual unbiased estimators.

The variance of $\hat{\bar{y}}$ and $\hat{\bar{x}}$ can be written as

$$\mathrm{V}(\hat{\bar{y}}) = \mathrm{V}(\bar{y}) + \frac{W_2 k}{n}\frac{\sum_{i=1}^{N_2}\tau_{1i}}{N_2}$$

and

$$\mathrm{V}(\hat{\bar{x}}) = \mathrm{V}(\bar{x}) + \frac{W_2 k}{n}\frac{\sum_{i=1}^{N_2}\tau_{2i}}{N_2}$$

Since $\bar{y}$ is the original Hansen & Hurwitz (1946) mean estimator, the variance of $\hat{\bar{y}}$ is given by

$$\mathrm{Var}(\hat{\bar{y}}) = \lambda S_y^2 + \lambda^* S_{y(2)}^2 + \frac{W_2 k}{n}\left(\frac{(S_{y(2)}^2 + \bar{y}_{(2)}^2)S_{S_1}^2 W + S_{S_2}^2 W}{(\bar{S}_1 W + 1 - W)^2}\right) \tag{12}$$

As well $\bar{x}$ is the original Hansen & Hurwitz (1946) mean estimator, the variance of $\hat{\bar{x}}$ is given by

$$\mathrm{Var}(\hat{\bar{x}}) = \lambda S_x^2 + \lambda^* S_{x(2)}^2 + \frac{W_2 k}{n}\left(\frac{(S_{x(2)}^2 + \bar{x}_{(2)}^2)S_{S_1}^2 W + S_{S_2}^2 W}{(\bar{S}_1 W + 1 - W)^2}\right) \tag{13}$$

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

In addition to non-response, measurement error is an important source of non-sampling errors in a survey. Let the measurement error for the study variable (Y) and auxiliary variable (X) in the population be given by $U_i = y_i - Y_i$ and $V_i = x_i - X_i$. Let the respective measurement error associated with the sensitive variables (Z_1 and Z_2) in face-to-face interview phase be given by $P_i = z_{1i} - Z_{1i}$ and $Q_i = z_{2i} - Z_{2i}$. These measurement errors are assumed to be random and uncorrelated with mean zero and variances S_u^2, S_v^2, S_p^2 and S_q^2, respectively.

Let us assume that the population mean of the sensitive auxiliary variable is known and non-response happens on the both X and Y then some notations are given below

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

$\hat{\Omega}_u^* = \sum_{i=1}^{n_1} U_i + \sum_{i=1}^{n_2} P_i; \hat{\Omega}_v^* = \sum_{i=1}^{n} V_i + \sum_{i-1}^{n_2} Q_i$

where U_i, V_i, P_i and Q_i are measurement errors on Y, X, Z_1 and Z_2 respectively.

Now, the variance of $\hat{\bar{y}}$ in the presence of measurement error is given by

$$\mathrm{Var}(\hat{\bar{y}}^*) = \lambda(S_y^2 + S_u^2) + \lambda^*(S_{y(2)}^2 + S_p^2) + \kappa_1 \tag{14}$$

Likewise, the variance of $\hat{\bar{x}}$ in the presence of measurement error is given by

$$\mathrm{Var}(\hat{\bar{x}}^*) = \lambda(S_x^2 + S_v^2) + \lambda^*(S_{x(2)}^2 + S_q^2) + \kappa_2 \tag{15}$$

where $\kappa_1 = \frac{W_2 k}{n}\left(\frac{(S_{y(2)}^2 + \bar{y}_{(2)}^2)S_{S_1}^2 W + S_{S_2}^2 W}{(\bar{S}_1 W + 1 - W)^2}\right)$; $\kappa_2 = \frac{W_2 k}{n}\left(\frac{(S_{x(2)}^2 + \bar{x}_{(2)}^2)S_{S_1}^2 W + S_{S_2}^2 W}{(\bar{S}_1 W + 1 - W)^2}\right)$.

Next, consider a linear regression estimator of y on x when there is presence of non-response and measurement error on y with full information on x as

$$\hat{T}_{reg}^* = \bar{y}^* + \hat{\beta}_{yx}(\bar{X} - \bar{x}) \tag{16}$$

where $\hat{\beta}_{yx} = s_{yx}/s_x^2$ is a sample estimate of the population regression coefficient $\beta_{yx} = S_{yx}/S_x^2$.

The mean squared error of $\hat{T}_{reg}^*$ is

$$\mathrm{MSE}(\hat{T}_{reg}^*) \cong \lambda((S_y^2 + S_u^2)(1 - \rho_{yx}^2)) + \lambda^*(S_{y(2)}^2 + S_p^2) \tag{17}$$

where $\rho_{yx}^2 = S_{yx}^2/S_y^2 S_x^2$.

Further, Diana *et al.* (2014) proposed a regression estimator, assuming X be a non-sensitive auxiliary variable and Y be a sensitive study variable in the presence of non-response and measurement error is given as

$$\hat{T}_D^* = \hat{\bar{y}}^* + \hat{\beta}_{yx}^{***}(\bar{X} - \bar{x}^*) \tag{18}$$

where $\hat{\beta}_{yx}^{***} = \hat{s_{yx}}^*/s_x^{*2}$ is the estimate of the population regression coefficient $\beta_{yx} = S_{yx}/S_x^2$.

The mean squared error of $\hat{T}_D^*$ is

$$\mathrm{MSE}(\hat{T}_D^*) \cong \lambda((S_y^2 + S_u^2)(1 - \rho_{yx}^2)) + \lambda^*((S_{y(2)}^2 + S_p^2) + \beta_{yx}^2(S_{x(2)}^2 + S_v^2) - 2\beta_{yx}S_{yx(2)}) + \frac{k}{nN}\sum_{i=1}^{N_2}\phi_i \tag{19}$$

In the next section, one can define four different scrambled models to check the efficiency of the proposed models.

4. Used Scrambled Models

In addition, a study focus on four well-known additive, multiplicative, and mixed nature models to get an idea of how they perform in terms of efficiency and privacy. To demonstrate this, suppose L_1 and L_2 are two mutually independent scrambling variables that are equally independent of Y. Then, consider the following linear models as

- M_1: Pollock & Bek (1976) additive model is $Z = Y + L_2$.
- M_2: Eichhorn & Hayre (1983) multiplicative model is $Z = L_1 Y$.
- M_3: Saha (2008) mixed model is $Z = L_1(Y + L_2)$.
- M_4: Diana *et al.* (2013) model is $Z = W(Y + L_2) + (1 - W)L_1 Y$.

Perhaps one may represent the variance using four different models in the presence of non-response and measurement error simultaneously,

$$\mathrm{Var}(\hat{T}^*_{D(M_1)}) = \lambda(S_y^2 + S_u^2) + \lambda^*(S^2_{y(2)} + S_p^2) + \frac{kN_2}{nN} S_{S_2} \tag{20}$$

$$\mathrm{Var}(\hat{T}^*_{D(M_2)}) = \lambda(S_y^2 + S_u^2) + \lambda^*(S^2_{y(2)} + S_p^2) + \frac{kN_2}{nN}\left(\frac{S^2_{S_1}(S^2_{S_1} + \bar{Y}_2^2)}{\bar{S}_1{}^2}\right) \tag{21}$$

$$\mathrm{Var}(\hat{T}^*_{D(M_3)}) = \lambda(S_y^2 + S_u^2) + \lambda^*(S^2_{y(2)} + S_p^2) + \frac{kN_2}{nN}\left(\frac{S^2_{S_1}(S^2_{S_1} + \bar{Y}_2^2) + 2\bar{S}_2 S^2_{S_1}\bar{Y}_2 + S^2_{S_2}\bar{S}_1{}^2 + S^2_{S_1}(\bar{S}_2{}^2 + S^2_{S_2})}{\bar{S}_1{}^2}\right) \tag{22}$$

$$\mathrm{Var}(\hat{T}^*_{D(M_4)}) = \lambda(S_y^2 + S_u^2) + \lambda^*(S^2_{y(2)} + S_p^2) + \frac{kN_2}{nN}\left(\frac{(1-W)^2 S^2_{S_1}(S^2_{S_1} + \bar{Y}_2^2) + W^2 S^2_{S_1}}{(W + (1-W)\bar{S}_1)^2}\right) \tag{23}$$

5. Proposed Scrambled Estimator

In human surveys, collection of information on sensitive variables is very serious. Many times this tends to increase the non-sampling errors i.e. non-response and measurement error. To overcome the problem of increase in non-response and measurement error, Diana *et al.* (2014) modified the Hansen & Hurwitz (1946) technique by collecting direct response from a respondent in the first phase and by collecting scrambled responses at second phase. Further, they used non-sensitive auxiliary variable to reduce the loss of efficiency due to scrambled responses. But, on the same side, auxiliary information can be sensitive to the respondents which may increase the possibility of non-response and measurement errors. So, an estimator for the estimation of population mean of a sensitive variable in the presence of non-response and measurement error by using auxiliary variable which may be sensitive in nature under optional randomized response models are proposed and is given as

$$\hat{T}^*_p = \hat{\bar{y}}^* + \hat{b}^*_{yx}(\bar{X} - \hat{\bar{x}}^*) + \hat{b}^{**}_{yx}(\bar{X} - \bar{x}) \tag{24}$$

where $\hat{b}^*_{yx} = \hat{s}^*_{yx}/\hat{s}^{*2}_x$ is the estimate of the population regression coefficient $\hat{\beta}^*_{yx} = \hat{S}^*_{yx}/\hat{S}^{*2}_x$.
and $\hat{b}^{**}_{yx} = \hat{s}_{yx}/s^2_x$ is the estimate of the population regression coefficient $\hat{\beta}^{**}_{yx} = \hat{S}_{yx}/S^2_x$.

To obtain the mean squared error of this estimator, we use models which are given in (2) and (3), define $\hat{\bar{y}}^* = \bar{Y}(1+\hat{e}^*_0)$, $\hat{\bar{x}}^* = \bar{X}(1+\hat{e}^*_1)$, $\bar{x} = \bar{X}(1+e_2)$, $\hat{s}^{*2}_x = \hat{S}^2_x(1+\hat{e}^*_3)$, $\hat{s}^*_{yx} = \hat{S}_{yx}(1+\hat{e}^*_4)$, $s^2_x = S^2_x(1+e_5)$, $\hat{s}_{yx} = \hat{S}_{yx}(1+\hat{e}_6)$ such that $E(\hat{e}^*_0) = E(\hat{e}^*_1) = E(e_2) = E(\hat{e}^*_3) = E(\hat{e}^*_4) = E(e_5) = E(\hat{e}_6) = 0$;

$E(\hat{e}_0^{*2}) = \frac{1}{\bar{Y}^2}\left(\lambda(S_y^2 + S_u^2) + \lambda^*(S_{y(2)}^2 + S_p^2) + \frac{W_2 k}{n}\left(\frac{\{(S_{y(2)}^2+\bar{y}_{(2)}^2)S_{S_1}^2 + S_{S_2}^2\}W}{(\bar{S}_1 W+1-W)^2}\right)\right);$

$E(\hat{e}_1^{*2}) = \frac{1}{\bar{X}^2}\left(\lambda(S_x^2 + S_v^2) + \lambda^*(S_{x(2)}^2 + S_q^2) + \frac{W_2 k}{n}\left(\frac{\{(S_{x(2)}^2+\bar{x}_{(2)}^2)S_{S_1}^2 + S_{S_2}^2\}W}{(\bar{S}_1 W+1-W)^2}\right)\right);\ E(e_2^2) = \frac{1}{\bar{X}^2}\lambda S_x^2;$

$E(\hat{e}_0^*\hat{e}_1^*) = \lambda\rho_{yx}\frac{S_y S_x}{\bar{Y}\bar{X}} + \lambda^*\rho_{yx(2)}\frac{S_y S_x}{\bar{Y}\bar{X}};\ E(\hat{e}_1^* e_2) = \lambda\rho_{z_2 x}\frac{S_{z_2} S_x}{\bar{Z}_2\bar{X}};\ E(\hat{e}_0^* e_2) = \lambda\rho_{yx}\frac{S_y S_x}{\bar{Y}\bar{X}}$

$E(\hat{e}_1^*\hat{e}_3^*) = \frac{1}{\bar{X}}\left(\lambda\frac{\hat{\alpha}_{03}^*}{\hat{\alpha}_{02}^*} + \lambda^*\frac{\hat{\alpha}_{03(2)}^*}{\hat{\alpha}_{02(2)}^*}\right);\ E(\hat{e}_1^*\hat{e}_4^*) = \frac{1}{\bar{X}}\left(\lambda\frac{\hat{\alpha}_{12}^*}{\hat{\alpha}_{11}^*} + \lambda^*\frac{\hat{\alpha}_{12(2)}^*}{\hat{\alpha}_{11(2)}^*}\right)$

$E(e_2 e_5) = \frac{1}{\bar{X}}\left(\lambda\frac{\mu_{03}^*}{\mu_{02}^*}\right)$ and $E(e_2\hat{e}_6) = \frac{1}{\bar{X}}\left(\lambda\frac{\mu_{12}^*}{\mu_{11}^*}\right)$

$\hat{\alpha}_{rs}^* = \sum_{i=1}^{N}\frac{(y_i-\bar{Y})^r(x_i-\bar{X})^s}{N-1};\ \hat{\alpha}_{rs(2)}^* = \sum_{i=1}^{N}\frac{(y_i-\bar{Y}_2)^r(x_i-\bar{X}_2)^s}{N_2-1}$

$\mu_{pq}^* = \sum_{i=1}^{N}\frac{(y_i-\bar{Y})^p(x_i-\bar{X})^q}{N-1};\ \mu_{pq(2)}^* = \sum_{i=1}^{N}\frac{(y_i-\bar{Y}_2)^p(x_i-\bar{X}_2)^q}{N_2-1}$

Using Taylor's approximation up to the first order, one can have

$$(\hat{T}_p^* - \bar{Y}) = \bar{Y}\hat{e}_0^* - \beta_{yx}(\hat{e}_1^* - \hat{e}_1^*\hat{e}_3^* + \hat{e}_1^*\hat{e}_4^*) - \beta_{yx}^*(e_2 - e_2\hat{e}_5^* + e_2\hat{e}_6^*) \tag{25}$$

The bias and mean squared error of the proposed estimator to second order of approximation in the presence of non-response and measurement error simultaneously, is given by

$$\text{Bias}(\hat{T}_p^*) = \hat{\beta}_{yx}^*\left(\left(\lambda\frac{\hat{\alpha}_{03}^*}{\hat{\alpha}_{02}^*} + \lambda^*\frac{\hat{\alpha}_{03(2)}^*}{\hat{\alpha}_{02(2)}^*}\right) - \left(\lambda\frac{\hat{\alpha}_{12}^*}{\hat{\alpha}_{11}^*} + \lambda^*\frac{\hat{\alpha}_{12(2)}^*}{\hat{\alpha}_{11(2)}^*}\right)\right) + \hat{\beta}_{yx}^{**}\left(\lambda\left(\frac{\mu_{03}^*}{\mu_{02}^*} - \frac{\mu_{12}^*}{\mu_{11}^*}\right)\right) \tag{26}$$

and

$$\begin{aligned}\text{MSE}(\hat{T}_p^*) = \lambda\big((S_y^2 + S_u^2) + \hat{\beta}_{yx}^{*2}(S_x^2 + S_v^2) - \hat{\beta}_{yx}^{**2}S_x^2 - 2\hat{\beta}_{yx}^*\rho_{yx}S_y S_x - 2\hat{\beta}_{yx}^*\hat{\beta}_{yx}^{**}\rho_{z_2x}S_{z_2}S_x - \\ 2\hat{\beta}_{yx}^{**}\rho_{z_1x}S_{z_1}S_x\big) + \lambda^*\big((S_{y(2)}^2 + S_p^2) + \hat{\beta}_{yx}^{*2}(S_{x(2)}^2 + S_q^2) - 2\hat{\beta}_{yx}^*\rho_{yx(2)}S_{y(2)}S_{x(2)}\big) \\ +\kappa_1 + \hat{\beta}_{yx}^{*2}\kappa_2\end{aligned} \tag{27}$$

By putting $S_u^2 = S_v^2 = S_p^2 = S_q^2 = 0$ in the above expression, we get the mean squared error of the proposed scrambled estimator without measurement error which is given by

$$\begin{aligned}\text{MSE}(\hat{T}_p^*) = \lambda\big(S_y^2 + \hat{\beta}_{yx}^{*2}S_x^2 - \hat{\beta}_{yx}^{**2}S_x^2 - 2\hat{\beta}_{yx}^*\rho_{yx}S_y S_x - 2\hat{\beta}_{yx}^*\hat{\beta}_{yx}^{**}\rho_{z_2x}S_{z_2}S_x - 2\hat{\beta}_{yx}^{**}\rho_{z_1x}S_{z_1}S_x\big) + \\ \lambda^*\left(S_{y(2)}^2 + \hat{\beta}_{yx}^{*2}S_{x(2)}^2 - 2\hat{\beta}_{yx}^*\rho_{yx(2)}S_{y(2)}S_{x(2)}\right) + \kappa_1 + \hat{\beta}_{yx}^{*2}\kappa_2\end{aligned} \tag{28}$$

Consequently, as mentioned in section 4, the four models of additive, multiplicative and mixed nature to have some idea that how they work in terms of efficiency and privacy are used and the variance expressions of $\hat{T}_p^*$ are given as

$$\text{Var}(\hat{T}_{p(M_1)}^*) = \sigma + \frac{kN_2}{nN}S_{S_2} \tag{29}$$

$$\text{Var}(\hat{T}_{p(M_2)}^*) = \sigma + \frac{kN_2}{nN}\left(\frac{S_{S_1}^2\mu}{\bar{S}_1^{\,2}}\right) \tag{30}$$

$$\text{Var}(\hat{T}_{p(M_3)}^*) = \sigma + \frac{kN_2}{nN}\left(\frac{S_{S_1}^2\mu + 2\bar{S}_2 S_{S_1}^2\bar{Y}_2 + S_{S_2}^2\bar{S}_1^{\,2} + S_{S_1}^2(\bar{S}_2^{\,2} + S_{S_2}^2)}{\bar{S}_1^{\,2}}\right) \tag{31}$$

$$\mathrm{Var}(\hat{T}^*_{p(M_4)}) = \sigma + \frac{kN_2}{nN}\left(\frac{(1-W)^2 S^2_{S_1}\mu + W^2 S^2_{S_1}}{(W + (1-W)\bar{S}_1)^2}\right) \tag{32}$$

where $\mu = (S^2_{S_1} + \bar{Y}^2_2)$;

and $\sigma = \lambda(S^2_y + S^2_u) + \lambda^*(S^2_{y(2)} + S^2_p)$.

From (29) and (30), it is noted that $\mathrm{Var}(\hat{T}^*_{p(M_2)}) > \mathrm{Var}(\hat{T}^*_{p(M_1)})$

$$\textit{iff} \qquad \mu = \frac{S^2_{S_2}}{S^2_{S_1}/\bar{S}_1{}^2} \tag{33}$$

holds true.

The multiplicative model-based estimator $\hat{T}^*_{p(M_2)}$ is a better choice in terms of privacy protection than $\hat{T}^*_{p(M_1)}$, although $\hat{T}^*_{p(M_1)}$ definitely performs better in terms of efficiency. The relationship between the model $\hat{T}^*_{p(M_4)}$ with the models $\hat{T}^*_{p(M_1)}$ and $\hat{T}^*_{p(M_2)}$ is fascinating. If $W = 0$ in (32), then

$$\mathrm{Var}(\hat{T}^*_{p(M_4)}) = \mathrm{Var}(\hat{T}^*_{p(M_2)})$$

and for $W = 1$,

$$\mathrm{Var}(\hat{T}^*_{p(M_4)}) = \mathrm{Var}(\hat{T}^*_{p(M_1)})$$

Now, it is easy to verify that $\mathrm{Var}(\hat{T}^*_{p(M_4)})$ achieves its minimum value for $W_0 = S^2_{S_2}\mu/(\bar{S}_{S_1}S^2_{S_2} + S^2_{S_1}\mu)$ from (29)-(32), it is easy to relate the performance of additive and multiplicative models with the mixed models and it can be seen that variance of $\hat{T}^*_{p(M_4)}$ in the optimal case (W_0) is lower than others if condition (33) is satisfied.

$$\mathrm{Var}(\hat{T}^*_{p(M_3)}) > \mathrm{Var}(\hat{T}^*_{p(M_2)}) > \mathrm{Var}(\hat{T}^*_{p(M_1)}) \geq \mathrm{Var}(\hat{T}^*_{p(M_4)})$$

It is interesting to note that this relation is expressing both aspects, i.e. efficiency and privacy. For biased estimators, we use mean squared errors in place of variances, therefore the following relation of mean squared error's holds for regression estimators as

$$\mathrm{MSE}(\hat{T}^*_{p(M_3)}) > \mathrm{MSE}(\hat{T}^*_{p(M_2)}) > \mathrm{MSE}(\hat{T}^*_{p(M_1)}) \geq \mathrm{MSE}(\hat{T}^*_{p(M_4)}) \tag{34}$$

Also, $\mathrm{MSE}(\hat{T}^*_p) < \mathrm{MSE}(\hat{T}^*_{p(M_j)}) < \mathrm{MSE}(\hat{T}^*_{reg}) < \mathrm{MSE}(\hat{T}^*_{D(M_j)})$; $j = 1, 2, 3, 4$.

To confirm the behavior of the above relations, we perform a numerical comparison by using *R* software.

6. Simulation Study

In this section, a simulation with particular focus on the performance of the proposed regression estimator $(\hat{T}^*_p)$ as compared to the regression estimator $(\hat{T}^*_{reg})$ and Diana *et al.* (2014) estimator $(\hat{T}^*_D)$ in the presence of non-response and measurement error.

6.1 Numerical illustration using Hypothetical Population

A finite population of size $N = 5,000$ is considered. A variable $X \sim N(0,1)$ is generated from normal distribution and variables Y which are related with X is defined as $Y = N(0,1) + X$. The scrambling variable S_1 is taken from normal distribution with mean equal to 1 and variance 0.5 and scrambling variable S_2 is also taken from normal distribution with mean equal to 0 and variance 0.5. Coding for simulation is done in R software, and the results are averaged over $5,000$ iterations.

Next, let us consider samples of size $n = 700$ using simple random sampling without replacement and assume a response rate of 40% in the first phase. This means in the first phase, only $240(n_1)$ provide a response to the survey question and $460(n_2)$ of them don't respond. In the second phase, we take another sample $(n_s = \frac{n_2}{k})$ from the non-respondent group by using $k = 1, 2, 3, 4, 5$, respectively. For various values of k, we analyze the behavior of the following estimators

$$\hat{T}^*_{reg}, \hat{T}^*_{D(M_j)}, \hat{T}^*_{p}, \hat{T}^*_{p(M_j)}; j = 1, 2, 3, 4.$$

The efficiency and privacy of unified measure ω as defined in Gupta *et al.* (2018) is given by

$$\omega = \frac{\text{MSE}(\hat{T}^*_i)}{\Gamma}; \tag{35}$$

where $\Gamma = E(Z_1 - Y)^2$ is the privacy level of sensitive models and $T^*_i = \hat{T}^*_p, \hat{T}^*_{p(M_j)}$; j=1, 2, 3, 4.

Table 1. Mean Squared Error of the suggested estimators for different values of k without measurement error.

k	Estimator(s)							
	$\hat{T}^*_{reg}$	$\hat{T}^*_{D(M_1)}$	$\hat{T}^*_{D(M_2)}$	$\hat{T}^*_{D(M_3)}$	$\hat{T}^*_{p}$	$\hat{T}^*_{p(M_1)}$	$\hat{T}^*_{p(M_2)}$	$\hat{T}^*_{p(M_3)}$
2	0.0055	3.5968	3.5963	3.5971	0.0025	0.0038	0.0037	0.0033
3	0.0093	5.3957	5.3952	5.3962	0.0058	0.0072	0.0070	0.0066
4	0.0133	7.1947	7.1943	7.1953	0.0094	0.0108	0.0107	0.0103
5	0.0160	8.9935	8.9921	8.9943	0.0125	0.0141	0.0139	0.0135

Table 2. Mean Squared Error of the suggested estimators for different values of k with measurement error.

k	Estimator(s)							
	$\hat{T}^*_{reg}$	$\hat{T}^*_{D(M_1)}$	$\hat{T}^*_{D(M_2)}$	$\hat{T}^*_{D(M_3)}$	$\hat{T}^*_{p}$	$\hat{T}^*_{p(M_1)}$	$\hat{T}^*_{p(M_2)}$	$\hat{T}^*_{p(M_3)}$
2	0.0037	3.5949	3.5943	3.5951	0.0020	0.0033	0.0031	0.0028
3	0.0063	5.3923	5.3917	5.3927	0.0054	0.0068	0.0066	0.0062
4	0.0093	7.1898	7.1893	7.1904	0.0092	0.0106	0.0105	0.0101
5	0.0104	8.9866	8.9852	8.9873	0.0100	0.0116	0.0114	0.0110

Table 1 and Table 2 represents the mean squared errors for $(\hat{T}^*_{reg})$, $(\hat{T}^*_{D(M_1)}, \hat{T}^*_{D(M_2)}, \hat{T}^*_{D(M_3)})$ estimators and our suggested estimators $(\hat{T}^*_{p(M_1)}, \hat{T}^*_{p(M_1)}, \hat{T}^*_{p(M_2)}$ and $\hat{T}^*_{p(M_3)})$ and from Table 1 and Table 2, we envisaged that as the value of k increases from 2 to 5, the MSE of each estimator increases. Moreover, Table 1 and Table 2 shows that the mean squared errors of Diana *et al.* (2014) estimator i.e. $\hat{T}^*_{D(M_3)}$ are the highest for all considered values of k and our suggested estimators i.e. $\hat{T}^*_{p}$ is lowest among all considered estimators. Also, the proposed estimators $\hat{T}^*_{p}$, $\hat{T}^*_{p(M_1)}$, $\hat{T}^*_{p(M_2)}$ and $\hat{T}^*_{p(M_3)}$ are having minimum mean squared error in compare to $\hat{T}^*_{reg}$, $\hat{T}^*_{D(M_1)}$, $\hat{T}^*_{D(M_2)}$ and $\hat{T}^*_{D(M_3)}$, respectively in the presence and absence of measurement error.

At last, the proposed estimator $(\hat{T}^*_p)$ under non-response and measurement error using optional randomized response models performing efficiently in comparison to the other estimators i.e. Regression estimator $(\hat{T}^*_{reg})$, Diana *et al.* (2014) estimators $\hat{T}^*_{D(M_1)}$, $\hat{T}^*_{D(M_2)}$ and $\hat{T}^*_{D(M_3)}$, $\hat{T}^*_{p(M_1)}$, $\hat{T}^*_{p(M_2)}$ and $\hat{T}^*_{p(M_3)}$. Above results are shown graphically in Figure 1.

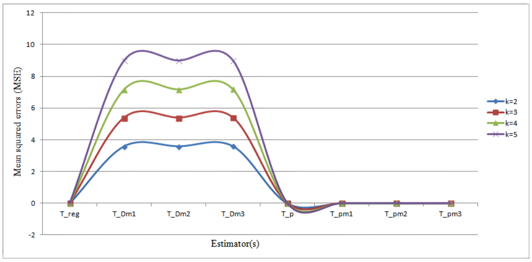

Figure 1. Mean Squared Error of the existing estimators and suggested estimators for different values of k in the presence of non-response and measurement error.

Table 3. Mean Squared Error of the suggested estimators for different values of (W, k) with measurement error.

Estimator(s)						W					
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
						for k = 2					
$\hat{T}^*_{D(M_4)}$	3.5962	3.5963	3.5964	3.5964	3.5965	3.5966	3.5967	3.5967	3.5969	3.5970	3.5970
$\hat{T}^*_{p(M_4)}$	0.0033	0.0032	0.0033	0.0033	0.0034	0.0034	0.0034	0.0035	0.0035	0.0035	0.0038
$\hat{T}^*_p$	0.0019	0.0020	0.0021	0.0022	0.0023	0.0024	0.0024	0.0025	0.0026	0.0027	0.0028
						for k = 3					
$\hat{T}^*_{D(M_4)}$	5.3955	5.3952	5.3953	5.3954	5.3954	5.3955	5.3956	5.3957	5.3959	5.3961	5.3968
$\hat{T}^*_{p(M_4)}$	0.0062	0.0063	0.0064	0.0065	0.0066	0.0066	0.0067	0.0068	0.0068	0.0069	0.0070
T^*_p	0.0049	0.0050	0.0051	0.0052	0.0053	0.0055	0.0056	0.0057	0.0058	0.0059	0.0061
						for k = 4					
$\hat{T}^*_{D(M_4)}$	7.1945	7.1943	7.1943	7.1944	7.1945	7.1946	7.1947	7.1948	7.1949	7.1951	7.1957
$\hat{T}^*_{p(M_4)}$	0.0093	0.0096	0.0098	0.0099	0.0100	0.0101	0.0102	0.0103	0.0104	0.0105	0.0105
$\hat{T}^*_p$	0.0080	0.0084	0.0085	0.0087	0.0086	0.0088	0.0090	0.0091	0.0093	0.0096	0.0096
						for k = 5					
$\hat{T}^*_{D(M_4)}$	8.9931	8.9922	8.9923	8.9925	8.9926	8.9928	8.9931	8.9933	8.9936	8.9939	8.9952
$\hat{T}^*_{p(M_4)}$	0.0130	0.0125	0.0127	0.0129	0.0130	0.0132	0.0133	0.0135	0.0136	0.0137	0.0146
$\hat{T}^*_p$	0.0117	0.0111	0.0113	0.0115	0.0117	0.0119	0.0121	0.0123	0.0125	0.0127	0.0137

Table 3 and 4 represents the mean squared error's of the estimators for different values of (W, k) with and without measurement error, respectively. From Table 3, the mean square error of Diana *et al.* (2014) estimator $(\hat{T}^*_{D(M_4)})$ is highest among all other estimators and on the same side the mean

Table 4. Mean Squared Error of the suggested estimators for different values of (W, k) without measurement error.

Estimator(s)	W										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
						for k = 2					
$\hat{T}^*_{D(M_4)}$	3.5941	3.5944	3.5944	3.5945	3.5945	3.5946	3.5947	3.5948	3.5949	3.5950	3.5948
$\hat{T}^*_{p(M_4)}$	0.0019	0.0026	0.0027	0.0027	0.0028	0.0028	0.0028	0.0029	0.0029	0.0029	0.0023
$\hat{T}^*_p$	0.0005	0.0015	0.0015	0.0016	0.0017	0.0018	0.0018	0.0019	0.0020	0.0020	0.0013
						for k = 3					
$\hat{T}^*_{D(M_4)}$	5.3918	5.3917	5.3918	5.3918	5.3919	5.3920	5.3921	5.3922	5.3924	5.3926	5.3930
$\hat{T}^*_{p(M_4)}$	0.0066	0.0058	0.0059	0.0060	0.0061	0.0062	0.0063	0.0063	0.0064	0.0064	0.0073
$\hat{T}^*_p$	0.0053	0.0046	0.0047	0.0048	0.0049	0.0051	0.0052	0.0053	0.0054	0.0055	0.0064
						for k = 4					
$\hat{T}^*_{D(M_4)}$	7.1892	7.1893	7.1894	7.1894	7.1895	7.1896	7.1897	7.1898	7.1900	7.1902	7.1903
$\hat{T}^*_{p(M_4)}$	0.0097	0.0094	0.0096	0.0097	0.0098	0.0099	0.0100	0.0101	0.0102	0.0103	0.0108
$\hat{T}^*_p$	0.0084	0.0082	0.0083	0.0085	0.0086	0.0088	0.0089	0.0091	0.0092	0.0094	0.0099
						for k = 5					
$\hat{T}^*_{D(M_4)}$	8.9865	8.9853	8.9854	8.9855	8.9857	8.9859	8.9861	8.9864	8.9866	8.9869	8.9884
$\hat{T}^*_{p(M_4)}$	0.0122	0.0100	0.0102	0.0104	0.0105	0.0107	0.0109	0.0110	0.0111	0.0113	0.0137
$\hat{T}^*_p$	0.0109	0.0086	0.0088	0.0090	0.0092	0.0094	0.0096	0.0098	0.0100	0.0102	0.0128

Squared Error of proposed estimator $(\hat{T}^*_p)$ is lowest among other existing estimators.

Table 3:

- For $k = 2, 3, 4$, $W = 0$; MSE of $\hat{T}^*_p$ is minimum and for $k = 5$, $W = 0.1$, MSE of $\hat{T}^*_p$ is minimum.

Table 4:

- For $k = 2$, $W = 0$, the MSE of $\hat{T}^*_p$ is minimum and for $k = 3, 4, 5$, $W = 0.1$, the MSE of $\hat{T}^*_p$ is minimum.

For $k = 2, 3, 4, 5$, $W = 0$ to 1, the proposed estimator $\hat{T}^*_p$ outperforms the other estimators i.e. $\hat{T}^*_{D(M_4)}$ and $\hat{T}^*_{p(M_4)}$ in terms of having minimum mean squared error. Also, with increase in k, the mean squared error of all estimators increases.

Table 5. Privacy (ω) of the suggested models for different values of W.

Estimator(s)	W										
	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\omega^*_{p(M_1)}$	0.0141	0.0134	0.0137	0.0140	0.0143	0.0146	0.0149	0.0151	0.0154	0.0157	0.0173
$\omega^*_{p(M_2)}$	0.0134	0.0127	0.0130	0.0133	0.0136	0.0139	0.0142	0.0145	0.0147	0.0150	0.0166
$\omega^*_{p(M_3)}$	0.0117	0.0113	0.0116	0.0119	0.0122	0.0125	0.0127	0.0130	0.0133	0.0136	0.0149
$\omega^*_{p(M_4)}$	0.0134	0.0127	0.0130	0.0132	0.0134	0.0135	0.0137	0.0138	0.0138	0.0139	0.0149
ω^*_p	0.0079	0.0082	0.0085	0.0087	0.0090	0.0093	0.0096	0.0099	0.0102	0.0105	0.0110

Table 5 indicates the privacy protection measure suggested by Gupta *et al.* (2018) which is given in (35). To see how this measure works in this study, we considered ω as ($\omega^*_{p(M_1)}$, $\omega^*_{p(M_2)}$, $\omega^*_{p(M_3)}$, $\omega^*_{p(M_4)}$ and ω^*_p) for five models in Table 5. W increases from 0 to 1, ω^*_p increases from 0.0079 to 0.0110 but first decrease for W = 0 to 1 and then increase for W = 0.2 to 1.0 other four models i.e. ($\omega^*_{p(M_1)}$, $\omega^*_{p(M_2)}$, $\omega^*_{p(M_3)}$, $\omega^*_{p(M_4)}$).

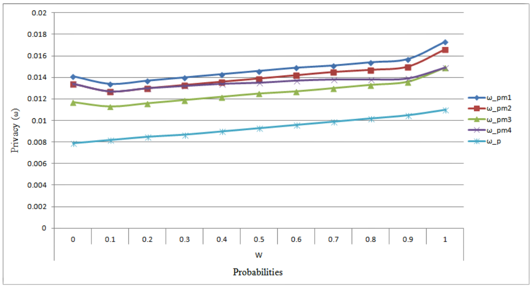

Figure 2. ω (privacy) of the suggested estimators.

Moreover, Table 5 and figure 2 depicts that the proposed estimator is better if we examine the performance of other estimators with respect to the unified measure (ω) of privacy and efficiency.

6.2 Numerical illustration using real population

The data set is based on Census 2011 literacy rates in India. The data is of N = 35 Indian states and union territories and then a random sample is drawn from population of size n = 10. The literacy rate is spread across the major parameters–Overall, Rural and Urban. Let y and x denotes the number of literates (people) in 2001 and the total literacy rate (2001), respectively. The scrambling variables S_1 and S_2 once again taken from normal distribution i.e., $S_1 \sim N(0, 0.5)$ and $S_2 \sim N(0, 0.5)$.

The results are shown in Table 6 and 7 for k = 2 and the probability level of sensitive variables, i.e. W = 0.8 are used.

Table 6. Mean Squared Error of the suggested estimators for k = 2 and W = 0.8 with measurement error.

Estimator(s)	Mean Squared Error	
	With M. E.	Without M. E.
$\hat{T}^*_{reg}$	3.2682	3.2160
$\hat{T}^*_{D(M_1)}$	2.5267	2.5135
$\hat{T}^*_{D(M_2)}$	2.3361	2.3348
$\hat{T}^*_{D(M_3)}$	4.2169	4.2156
$\hat{T}^*_{D(M_4)}$	4.3242	4.3111
$\hat{T}^*_{p}$	0.0483	0.0419
$\hat{T}^*_{p(M_1)}$	1.3239	1.1176
$\hat{T}^*_{p(M_2)}$	1.1613	1.1549
$\hat{T}^*_{p(M_3)}$	1.1763	1.1699
$\hat{T}^*_{p(M_4)}$	1.7489	1.7425

Table 6 and 7 indicates the mean squared error and privacy protection measure using real data set is based on Census 2011 literacy rates in India. From Table 6, the mean square error of Diana

Table 7. Privacy (ω) of the suggested models for $k = 2$ and $W = 0.8$.

Estimator(s)	Privacy (ω)
$\omega^*_{p(M_1)}$	0.0033
$\omega^*_{p(M_2)}$	0.0011
$\omega^*_{p(M_3)}$	0.0012
$\omega^*_{p(M_4)}$	0.0038
ω^*_{p}	0.0001

et al. (2014) estimator ($\hat{T}^*_{D(M_4)}$) is highest among all other estimators and on the same side the mean Squared Error of proposed estimator ($\hat{T}^*_p$) is lowest among other existing estimators. Moreover, from Table 7, the privacy protection measure of our proposed estimator (ω^*_p) is lowest which indicates that lesser the unified measure (i.e. Privacy and Efficiency) more efficient is our ORRT model.

7. Conclusions

An advanced optional randomized response technique is used in this study by making use of correlated quantitative scrambling variables. We have formulated a regression estimator for estimating the population mean of sensitive variable(s) in the presence of non-response and measurement error simultaneously using optional randomized response models. The optional randomized response model leads to better results than the other considered estimators in the presence of non-response and measurement errors simultaneously. The properties of the proposed estimator have also been obtained. The mean squared errors of the proposed estimator with other existing estimators i.e linear regression estimator ($\hat{T}^*_{reg}$) and Diana *et al.* (2014) estimator ($\hat{T}^*_{D(M_j)}$) has also been discussed and the conditions have been obtained. To verify the theoretical results, we have performed simulation using *R*. Tables and Figures shows that the proposed estimator is more efficient than the other conventional estimators. Thus, we recommend that our proposed estimator performs efficiently in a situation when the respondent finds the question more sensitive, always opts for more privacy.

Acknowledgments

The authors would like to express their gratitude to the esteemed referees for their valuable comments and suggestions, which have significantly enhanced the quality of this paper.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization: KUMAR, S.; KOUR, S. P. **Data curation**: KUMAR, S.; KOUR, S. P. **Formal analysis**: KUMAR, S.; KOUR, S. P. **Funding acquisition**: KUMAR, S.; KOUR, S. P. **Investigation**: KUMAR, S.; KOUR, S. P. **Methodology**: KUMAR, S.; KOUR, S. P. **Project administration**: KUMAR, S.; KOUR, S. P. **Software**: KUMAR, S.; KOUR, S. P. **Resources**: KUMAR, S.; KOUR, S. P. **Supervision**: KUMAR, S.; KOUR, S. P. **Validation**: KUMAR, S.; KOUR, S. P. **Visualization**: KUMAR, S.; KOUR, S. P. **Writing – original draft**: KUMAR, S.; KOUR, S. P. **Writing – review and editing**: KUMAR, S.; KOUR, S. P.

References

1. Diana, G., Giordan, M. & Perri, P. F. Randomized response surveys: A note on some privacy protection measures. *Model Assisted Statistics and Applications* **8,** 19–28 (2013).
2. Diana, G. & Perri, P. F. A class of estimators for quantitative sensitive data, Statistical Papers. *Statistical Papers* **52,** 633–650 (2011).
3. Diana, G., Riaz, S. & Shabbir, J. Hansen and Hurwitz estimator with scrambled response on the second call. *Journal of Applied Statistics* **41,** 596–611 (2014).
4. Eichhorn, B. H. & Hayre, L. S. Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference* **7,** 307–316 (1983).
5. Greenberg, B. G., Kuebler, R. R. J., Abernathy, J. R. & Hovertiz, D. G. Application of the randomized response techniques in obtaining quantitative data. *Journal of the American Statistical Association* **66,** 243–250 (1971).
6. Guha, S. & Chandra, H. Improved estimation of finite population mean in two-phase sampling with sub-sampling of the non-respondents. *Mathematical Population Studies* **28,** 24–44 (2019).
7. Gupta, S., Gupta, B. & Singh, S. Estimation of sensitivity level of personal interview survey questions. *Journal of Statistical Planning and inference* **100,** 239–247 (2002).
8. Gupta, S., Mehta, S., Shabbir, J. & Khalil, S. A unified measure of respondent privacy and model efficiency in quantitative RRT models. *Journal of Statistical Theory and Practice* **12,** 506–511 (2018).
9. Gupta, S. & Shabbir, J. Sensitivity estimation for personal interview survey questions. *Statistica* **64,** 643–653 (2004).
10. Gupta, S., Shabbir, J., Sousa, R. & Corte-Real, P. Estimation of the mean of a sensitive variable in the presence of auxiliary information. *Communications in Statistics-Theory and Methods* **41,** 13–14 (2012).
11. Gupta, S. N., Thornton, B., Shabbir, J. & Singhal, S. A comparison of multiplicative and additive optional RRT models. *Journal of Statistical Theory and Applications* **5,** 226–239 (2006).
12. Hansen, M. H. & Hurwitz, W. N. The Problem of Non-Response in Sample Surveys. *Journal of the American Statistical Association* **41,** 517–529 (1946).
13. Hussain, Z. & Shabbir, J. Improved estimation procedures for the mean of sensitive variable using randomized response model. *Pakistan Journal of Statistics* **25,** 205–220 (2009).
14. Khalil, S., Amin, M. & Hanif, M. Estimation of population mean for a sensitive variable in the presence of measurement error. *Journal of Statistics and Management Systems* **21,** 81–91 (2018).
15. Khalil, S., Zhang, Q. & Gupta, S. Mean estimation of sensitive variables under measurement errors using optional RRT models. *Communications in Statistics-Simulation and Computation* **50,** 1417–1426 (2021).
16. Kumar, M., Singh, R., Singh, A. K. & Smarandache, F. Some ratio type estimators under measurement errors. *World Applied Sciences Journal* **14,** 272–276 (2011).
17. Kumar, S. & Kour, S. P. Quantify the Impact of Non-Response and Measurement Error of Sensitive Variable(s) under Two-Phase Sampling employing ORRT Models. *Communications in Advanced Mathematical Sciences* **6,** 196–210 (2023).
18. Kumar, S. & Kour, S. P. The joint influence of estimation of sensitive variable under measurement error and non-response using ORRT models. *Journal of Statistical Computation and Simulation* **92,** 3583–3604 (2022).

19. Kumar, S., Kour, S. P. & Zhang, Q. An enhanced ratio-cum-product estimator with non-response and observational error by utilizing ORRT models: a sensitive estimation approach. *Journal of Statistical Computation and Simulation* **93,** 818–836 (2023).

20. Kumar, S., Kour, S. & Singh, H. P. A Class of Ratio Cum Product Estimators with Non-Response and Measurement Error Using ORRT Models: A Double Sampling Scheme. *Revista Colombiana de Estadística* **48,** 215–237 (2025).

21. Kumar, S., Kour, S., Singh, H. P. & Verma, T. An efficient class of ratio cum product estimators with non-response and measurement error using ORRT models: a sensitive estimation procedure. *Revista Investigation Operacional* **45,** 740–765 (2024).

22. Mushtaq, N. & Amin, M. Joint Influence of Double Sampling and Randomized Response Technique on Estimation Method of Mean. *Applied Mathematics* **10,** 12–19 (2020).

23. Pollock, K. & Bek, Y. A comparison of three randomized response models for quantitative data. *Journal of the American Statistical Association* **71,** 884–886 (1976).

24. Saha, A. A randomized response technique for quantitative data under unequal probability sampling. *Journal of Statistical Theory and Practice* **2,** 589–596 (2008).

25. Unal, C. & Kadilar, C. Improved Family of Estimators Using Exponential Function for the Population Mean in the Presence of Non-response. *Communications in Statistics: Theory and Methods* **50,** 237–248 (2021).

26. Warner, S. L. Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* **60,** 63–69 (1965).

27. Yan, Z., Wang, J. & Lai, J. An efficiency and protection degree-based comparison among the quantitative randomized response strategies. *Communications in Statistics-Theory and Methods* **38,** 400–408 (2008).

28. Yaqub, M., Shabbir, J. & Gupta, S. N. Estimation of population mean based on dual use of auxiliary information in non response. *Communications in Statistics-Theory and Methods* **46,** 12130–12151 (2017).

29. Zhang, Q., Kalucha, G., Gupta, S. & Khalil, S. Ratio Estimation of the Mean under RRT Models. *Journal of Statistics and Managements Systems* **22,** 97–113 (2018).