1. Introduction

In surveys we often gather information related to issues that people like to hide from their fellow human beings. Finding HIV tests positive, frequent drunken driving, abortions discreetly induced, misconduct with the spouse, false claim for social benefits, under reporting of income tax, etc., are some of the properties that may involve unethical stigmas to many delinquents. People generally dislike their revelations.

But in practice, the collection of truthful and reliably accurate data relating to sensitive issues seems crucial and highly challenging as respondents often provide untrue responses or even refuse to respond due to social stigma and or fear.

Hence, in such circumstances the methods that protect anonymity are a solution. Two widely practised ways has been noticed that protect the anonymity of respondents. One is randomised response technique and other is scrambled response technique. Warner (1965) initiated a technique to deal with sensitive issues which is to obtain responses through a randomized response (RR) survey where every sampled unit is asked to give a response through a RR device as per instruction from the investigator. One can refer to Greenberg et al. (1971), Bar-lev et al. (2004), Diana and Perri (2011) and Arcos et al. (2015) etc. for a comprehensive review of such RR procedure. However, there is another approach to deal with sensitive issue called scrambled response technique introduced by Pollock and Bek (1976). Many researchers such as Eichhorn and Hayre (1983), Saha (2007) and Diana and Perri (2010), etc. considered the scrambled response models to deal with sensitive issues.

There are many sensitive issues which need to be monitored over time as they may change over time. To address the character changing over time Jessen (1942) initiated the sampling procedure. His marvelous ideas was carried forward by many others such as...
Patterson (1950), Sen (1973), Feng and Zou (1997), Singh and Priyanka (2008), Priyanka and Mittal (2014, 2015a,b) and Priyanka et al. (2015, 2018a) etc. However, if the variable which opt to change over time is also sensitive in nature, then their arises a need to apply randomized/scrambled response techniques on successive occasions. Arnab and Singh (2013), Yu et al. (2015), Naeem and Shabbir (2018), Singh et al. (2018) and Priyanka et al. (2018b) have put their efforts to deal with sensitive issues on successive occasions.

In the present work an improved class of estimators have been proposed for estimating sensitive population mean at current occasion in two occasion successive sampling using an innocuous auxiliary variable. The behaviour of proposed improved class of estimator are discussed for scrambled response models. Many existing estimators in successive sampling literature have been modified to work under the considered scrambled response model for dealing with sensitive issues. The proposed improved class of estimators have been compared with recent estimators such as modified Singh and Pal (2015, 2017). Theoretical considerations are integrated with empirical and simulation studies to ascertain the efficiency gain derived from the proposed improved class of estimators.

2. Survey strategies and analysis

2.1. Notations and preliminaries

A finite population $U$ of $N$ units has been considered for sampling over two successive occasions. The sensitive study variable be referred as $x$ at the first occasion and $y$ at second occasion. Whereas $z$ is assumed to be innocuous auxiliary variable which is available at both the successive occasions. At first occasion a simple random sample without replacement of size $n$ is drawn and at the second occasion two independent samples are drawn by considering the partial overlapping case, one is matched sample of size $m = n \lambda$ drawn as sub sample from the sample of size $n$ from first occasion and another is unmatched simple random sample of size $u = (n - m) = n \mu$ drawn afresh at the second (current) occasion so that the sample size at both the occasion is $n$. On first (second) occasion the sensitive variables $x(y)$ are coded to $g(h)$ with the aid of scrambling variables $W_1$ and $W_2$. The scrambling variable are so considered that they may follow any distribution. The following notations to be considered here after.

$\bar{X}, \bar{Y}, \bar{Z}, \bar{G}, \bar{H}, \bar{W}_1, \bar{W}_2$: Population means of the variables $x, y, z, g, h, W_1$ and $W_2$, respectively.

$\bar{h}_u, \bar{h}_m, \bar{g}_m, \bar{h}_n, \bar{g}_n$: Sample mean of the variables based on sample sizes shown in suffices.

$\bar{z}_u, \bar{z}_m, \bar{z}_n$: Sample mean of the innocuous auxiliary variate based on sample sizes shown in suffice.

$\rho_{xy}, \rho_{xz}, \rho_{yz}, \rho_{gh}, \rho_{hz}, \rho_{gz}$: Correlation coefficient between the variables shown in suffices.
2.2. Scrambled response techniques on successive occasions

Scrambling the true response increases the participation of respondents. A recent paper by Diana and Perri (2010) developed efficient scrambled response model for one time survey. In this paper we intend to modify their scrambled response model for two occasion successive sampling with the proposed improved class of estimators. The coded response at first and second occasion under modified Diana and Perri (2010) model (say \( M_{DP} \)) are given as

\[
G = \phi_x (X + W_1) + (1 - \phi_x) W_2 X \\
H = \phi_y (Y + W_1) + (1 - \phi_y) W_2 Y,
\]

where \( 0 \leq \phi_x, \phi_y \leq 1 \).

For estimating sensitive population mean at current occasion, Equation (2) can be solved for \( \tilde{Y} \) as

\[
\tilde{Y} = \frac{\hat{H} - \phi_y \hat{W}_1}{\phi_y + (1 - \phi_y) \hat{W}_2},
\]

with

\[
\rho_{bg} = \frac{\phi_x \phi_y [A_1] + [A_2] (\phi_x + \phi_y) + A_3}{\sqrt{A_4 \sqrt{A_5}}}, \quad \rho_{bz} = \frac{\phi_y \sigma_y [\phi_y (1 - \tilde{W}_2) + \tilde{W}_2]}{\sqrt{A_4}}
\]

\[
\rho_{gz} = \frac{\rho_{xg} \sigma_x [\phi_x (1 - \tilde{W}_2) + \tilde{W}_2]}{\sqrt{A_5}}, \quad C_b = \frac{A_4}{H^2} \text{ and } C_g = \frac{A_5}{G^2},
\]

where

\[
A_1 = \rho_{yx} \sigma_y \sigma_x + \sigma^2_{W_1} - \rho_{yx} 2 \tilde{W}_2 \sigma_y \sigma_x + \sigma^2_{W_2} [\rho_{yx} \sigma_y \sigma_x + \tilde{X} \tilde{Y}] + \rho_{yx} \tilde{W}_2 \sigma_y \sigma_x,
\]

\[
A_2 = \tilde{W}_2 \rho_{yx} \sigma_y \sigma_x - \sigma^2_{W_2} [\rho_{yx} \sigma_y \sigma_x + \tilde{X} \tilde{Y}] - \tilde{W}_2^2 \rho_{yx} \sigma_y \sigma_x,
\]

\[
A_3 = \sigma^2_{W_2} [\rho_{yx} \sigma_y \sigma_x + \tilde{X} \tilde{Y}] + \tilde{W}_2^2 \rho_{yx} \sigma_y \sigma_x,
\]

\[
A_4 = (\phi_y)^2 [\sigma^2_y + \sigma^2_{W_1}] + (1 - \phi_y)^2 [\sigma^2_{W_2} (\tilde{Y}^2 + \sigma^2_{\tilde{Y}}) + \sigma^2_{\tilde{W}_2}] + 2 \phi_y (1 - \phi_y) \tilde{W}_2 \sigma^2_y,
\]

\[
A_5 = (\phi_x)^2 [\sigma^2_x + \sigma^2_{\tilde{W}_1}] + (1 - \phi_x)^2 [\sigma^2_{\tilde{W}_2} (\tilde{X}^2 + \sigma^2_{\tilde{X}}) + \sigma^2_{\tilde{W}_2}] + 2 \phi_x (1 - \phi_x) \tilde{W}_2 \sigma^2_x.
\]
Remark 1. The choice of $\phi_x(\phi_y) = 1$ in scrambled response model $M_{DP}$, generates modified additive scrambled response model by Pollock and Bek (1976). However, the choice of $\phi_x(\phi_y) = 0$ in $M_{DP}$, yields the modified multiplicative model by Pollock and Bek (1976) which is elaborated in details by Eichhorn and Hayre (1983).

Remark 2. In the above model $E(W_1) = \tilde{W}_1$, $E(W_2) = \tilde{W}_2$, $V(W_1) = \sigma^2_{W_1}$, $V(W_2) = \sigma^2_{W_2}$.

Remark 3. Suitable estimator of population mean of coded response variable $\tilde{H}$ need to be investigated and replaced in Equation (3) in order to obtain appropriate estimator of sensitive population mean at current occasion under two occasion successive sampling.

2.2.1. Design of the proposed improved class of estimators

To get maximum utilization of innocuous auxiliary variable, the availability of variance of auxiliary variable may be utilized to estimate the coded response variable which lead to improve the estimator for sensitive variable. In this section we intend to propose an improved general class of estimator for the population mean of coded response variable, $\tilde{H}$ at current occasion as

$$T = \Omega T_u + (1 - \Omega) T_m,$$

where

$$T_u = F_u(\tilde{b}_u, a_u, b_u) \text{ with } a_u = \frac{\tilde{z}_u}{Z}, \quad b_u = \frac{\tilde{s}^2_u}{s^2_Z},$$

and

$$T_m = F_m(\tilde{b}_m, t_1, t_2),$$

where

$$t_1 = K_1(\tilde{g}_m, a_m, b_m) \text{ and } t_2 = K_2(\tilde{g}_n, a_n, b_n), \quad a_m = \frac{\tilde{z}_m}{Z}, \quad b_m = \frac{s^2_m}{s^2_Z}, \quad a_n = \frac{\tilde{z}_n}{Z}, \quad b_n = \frac{s^2_n}{s^2_Z},$$

and $\Omega \in [0, 1]$ is a scalar quantity to be chosen suitably.

Therefore, substituting the improved class of estimator $T$ of the coded response variable $\tilde{H}$ in Equation (3), the estimator of sensitive population mean at current occasion is obtained as

$$\hat{Y} = \frac{T - \phi_y \tilde{W}_1}{\phi_y + (1 - \phi_y) \tilde{W}_2}.$$
2.2.2. Regularity conditions

Following Srivastava (1980) and Tracy et al. (1996) the following regularity conditions has been assumed.

(i) The points \((\bar{h}_u, a_u, b_u)\) and \((\bar{h}_m, t_1, t_2)\) assumes the values in a closed convex subset \(\mathbb{R}^3\) of three dimensional real spaces containing the point \((\bar{H}, 1, 1)\) and \((\bar{H}, \bar{G}, \bar{G})\) respectively.

(ii) The functions \(F_u\) and \(F_m\) are continuous and bounded in \(\mathbb{R}^3\).

(iii) The first and second order partial derivatives of \(F_u(\bar{h}_u, a_u, b_u)\) and \(F_m(\bar{h}_m, t_1, t_2)\) exists and are continuous and bounded in \(\mathbb{R}^3\).

(iv) \(t_1\) and \(t_2\) are two different classes of estimators of \(\bar{G}\) through samples of sizes \(m\) and \(n\) respectively such that \(K_1(\bar{G}, 1, 1) = K_2(\bar{G}, 1, 1) = \bar{G}\).

(v) \(F_u(\bar{H}, 1, 1) = \bar{H}\) and \(F_{1u}(Q) = \frac{\partial F_u(\cdot)}{\partial \bar{h}_u}\big|_Q = 1\) with \(Q = (\bar{H}, 1, 1)\).

(vi) \(F_m(\bar{H}, \bar{G}, \bar{G}) = \bar{H}\) and \(F_{1m}(R) = \frac{\partial F_m(\cdot)}{\partial \bar{h}_m}\big|_R = 1\) with \(R = (\bar{H}, \bar{G}, \bar{G})\).

3. Features of the proposed improved class of estimators

3.1. Bias and mean squared error of \(T\)

It may be observed from Section 2 that the proposed improved class of estimator \(T\) depends on \(T_u\) and \(T_m\) which are biased for \(\bar{H}\). This indicates that the combined improved class of estimators \(T\) is also biased for \(\bar{H}\). Hence, following transformations has been assumed in order to derive bias and mean squared error of the estimator \(T\):

\[
\begin{align*}
\bar{h}_u &= \bar{H}(1 + e_1), \quad \bar{b}_m = \bar{H}(1 + e_2), \quad \bar{g}_n = \bar{G}(1 + e_3), \quad \bar{g}_m = \bar{G}(1 + e_4), \quad \bar{z}_u = \bar{Z}(1 + e_5), \\
\bar{z}_m &= \bar{Z}(1 + e_6), \quad \bar{z}_u = \bar{Z}(1 + e_7), \quad \bar{s}_{zn} = S_2(1 + e_8), \quad \bar{s}_{zm} = S_2(1 + e_9), \\
\bar{s}_{zn} &= S_2(1 + e_{10}) \text{ such that, } E(e_j) = 0, |e_j| < 1, \text{ where } j = 1, 2, 3, \ldots, 10.
\end{align*}
\]

3.1.1. Bias and mean squared error of \(T_u\) and \(T_m\)

The bias and mean squared error of class of estimators \(T_u\) are derived up to first order approximations using the above transformations as:

\[
T_u = F_u(\bar{b}_u, a_u, b_u).
\]
Expanding $F_u(b_u, a_u, b_u)$ about the point $Q = (\bar{H}, 1, 1)$ in a first order Taylor series, we have

$$T_u = [F_u(Q) + (\bar{b}_u - \bar{H})F_1 + (a_u - 1)F_2 + (b_u - 1)F_3$$

$$+ \frac{1}{2}[(\bar{b}_u - \bar{H})^2F_{11} + (a_u - 1)^2F_{22} + (b_u - 1)^2F_{33} + (\bar{b}_u - \bar{H})(a_u - 1)F_{12}$$

$$+ (\bar{b}_u - \bar{H})(b_u - 1)F_{13} + (a_u - 1)(b_u - 1)F_{23} + \ldots)]],$$

(6)

where

$$F_1 = \frac{\partial F_u}{\partial b_u}|_Q = 1, \quad F_2 = \frac{\partial F_u}{\partial a_u}|_Q, \quad F_3 = \frac{\partial F_u}{\partial b_u}|_Q, \quad F_{11} = \frac{\partial^2 F_u}{\partial b_u^2}|_Q = 0,$$

$$F_{22} = \frac{\partial^2 F_u}{\partial a_u^2}|_Q, \quad F_{33} = \frac{\partial^2 F_u}{\partial b_u^2}|_Q, \quad F_{12} = \frac{\partial^2 F_u}{\partial b_u \partial a_u}|_Q, \quad F_{13} = \frac{\partial^2 F_u}{\partial b_u \partial b_u}|_Q,$$

$$F_{23} = \frac{\partial^2 F_u}{\partial a_u \partial b_u}|_Q, \quad Q = (\bar{H}, 1, 1).$$

Now, assuming $F_2 = -F_3$ as $F_2$ and $F_3$ are based on sample size $u$ and $(a_u - 1)$ and $(b_u - 1)$ assumes same value $\bar{H}$ at $(\bar{H}, 1, 1)$. Hence, the expression of $T_u$ up to first order approximation becomes

$$[T_u - \bar{H}] = [(\bar{b}_u - \bar{H}) + [(a_u - 1) - (b_u - 1)]F_2 + \frac{1}{2}[(a_u - 1)^2F_{22}$$

$$+ (b_u - 1)^2F_{33} + (\bar{b}_u - \bar{H})(a_u - 1)F_{12} + (\bar{b}_u - \bar{H})(b_u - 1)F_{13}$$

$$+ (a_u - 1)(b_u - 1)F_{23} + \ldots]].$$

(7)

Taking expectations on both sides in the above equation, we get bias of $T_u$ up to first order approximation as

$$B(T_u) = \frac{1}{2u}[F_{22}C_z^2 + H_{33}(\eta_{04} - 1) + \bar{H}F_{12}\rho_{hz} C_z - \bar{H}F_{13}C_z + \eta_{12} + F_{23}C_z\eta_{03}].$$

(8)

Now, squaring both sides of Equation (7) and retaining terms up to first order of approximations, we have

$$(T_u - \bar{H})^2 = [\bar{H}^2e_1^2 + F_2^2(e_5^2 + e_8^2 - 2e_5e_8) + 2\bar{H}F_2(e_1e_5 - e_1e_8)].$$

(9)

Taking expectations on both sides of Equation (9), the mean squared error of $T_u$ is obtained as

$$M(T_u) = \frac{1}{u}[S^2 + F_2^2C_z^2 + F_2^2(\eta_{04} - 1) - 2F_2^2C_z\eta_{03} + 2F_2^2S_h\rho_{hz} C_z - 2S_hF_2\eta_{12}],$$

(10)
which is optimized for $F_2 = \left( \frac{S_h(\eta_{12} - \rho_{bg}C_z)}{C_g + (\eta_{04} - 1) - 2C_g\eta_{03}} \right) = F^*_{2}$ (say).

Further, substituting optimum value of $F_2$ in the above equation we obtain the required optimum mean squared error of $T_u$ as

$$M(T_u)_{opt.} = \frac{1}{\mu} \left[ S_h^2 + (F^*_{2})^2 C_z + (F^*_{2})^2 (\eta_{04} - 1) - 2(F^*_{2})^2 C_z\eta_{03} + 2F^*_{2}S_h\rho_{bg}C_z - 2S_hF^*_{2}\eta_{12} \right],$$

where $\nu_{rs} = \frac{1}{N-1}\Sigma(G_i - \tilde{G})(Z_i - \tilde{Z})', \eta_{rs} = \frac{\nu_{rs}}{\nu_{ss}}$, $\nu_{\rho} = \frac{1}{N-1}\Sigma(X_i - \tilde{X})(Z_i - \tilde{Z})'$,

$$\eta_{\rho}^o = \frac{(\nu_{\rho})^o}{(\nu_{\rho}^o)^2}.$$  

Similarly, the bias and mean squared error of class of estimators $T_m$ are derived up to first order approximations using the above transformations and are obtained as:

$$B(T_m) = \frac{1}{2} \left[ \frac{1}{m} \left[ S_g^2 M_{22} + S_hS_g\rho_{bg}H_2 M_{12} \right] + \frac{1}{n} \left[ S_g^2 M_{33} + J_2^2 M_{33} + S_g^2 M_{23} + S_g J_2 M_{23} Q_3 \right] + \left( \frac{1}{m} - \frac{1}{n} \right) \left[ Q_1 H_2^2 M_{22} + 2H_2 M_{22} S_g Q_3 + H_2 M_{12} S_g Q_2 + M_{23} H_2 \rho_{bg} C_z S_g \right] \right],$$

where

$$M_1 = \frac{\partial F_m}{\partial h_m} \bigr|_R = 1, \quad M_2 = \frac{\partial F_m}{\partial t_1} \bigr|_R, \quad M_3 = \frac{\partial F_m}{\partial t_2} \bigr|_R, \quad M_4 = \frac{\partial^2 F_m}{\partial h_m^2} \bigr|_R = 0,$$

$$M_{22} = \frac{\partial^2 F_m}{\partial t_1^2} \bigr|_R, \quad M_{33} = \frac{\partial^2 F_m}{\partial t_2^2} \bigr|_R, \quad M_{12} = \frac{\partial^2 F_m}{\partial h_m \partial t_1} \bigr|_R, \quad M_{13} = \frac{\partial^2 F_m}{\partial h_m \partial t_2} \bigr|_R,$$

and

$$M(T_m)_{opt.} = \frac{1}{m} S_h^2 + \left( \frac{1}{m} - \frac{1}{n} \right) \left[ (M_2^*)^2 S_g^2 + (M_2^*)^2 (H_2^*)^2 Q_1 + 2S_hS_g M_{22}^* \rho_{bg} \right. \right.$$  

$$\left. + 2S_h M_{22}^* H_2^* Q_2 - 2(M_2^*)^2 S_g H_2^* \eta_{12}^o \right] + \frac{1}{n} \left[ (M_2^*)^2 (J_2^*)^2 Q_1 - 2M_2^* J_2^* S_h Q_2 \right],$$

which is optimized for $M_2 = \frac{(\rho_{bg} Q_1 + \eta_{12}^o Q_2^o)}{Q_1 - (\eta_{12}^o)^2} = M^*_2$ (say), $J_2 = \frac{S_h Q_2}{M_{12}^* Q_1} = J^*_2$ (say) and $H_2 = \frac{M_2 S_h \eta_{12} - S_h Q_2}{M_{12}^* Q_1} = H^*_2$ (say), where $Q_1 = C_g^2 + (\eta_{04} - 1) - 2C_g \eta_{03}, Q_2 = \rho_{bg} C_z - \eta_{12}^o, Q_3 = \rho_{bg} C_z - \eta_{12}^o$ and $Q_3 = \rho_{bg} C_z - \eta_{12}^o$. 


THEOREM 4. Bias of the improved class of estimators $T$ to the first order of approximations are obtained as

$$B(T) = \Omega B(T_u) + (1 - \Omega)B(T_m), \quad (13)$$

where $B(T_u)$ and $B(T_m)$ are given in Equations (8) and (11) respectively.

PROOF. The bias of the improved class of estimators $T$ is given by

$$B(T) = E(T - \bar{H}) = E[\Omega(T_u - \bar{H}) + (1 - \Omega)(T_m - \bar{H})] = \Omega B(T_u) + (1 - \Omega)B(T_m)$$

Substituting the values of $B(T_u)$ and $B(T_m)$ from the Equations (8) and (11) in the above equation, we have the expression for the bias of the general class of estimators $T$ given in Equation (13).

THEOREM 5. Mean squared error of the improved class of estimators $T$ to first order of approximations are obtained as

$$M(T) = \Omega^2 M(T_u)_{opt.} + (1 - \Omega)^2 M(T_m)_{opt.}, \quad (14)$$

where $M(T_u)_{opt.}, M(T_m)_{opt.}$ are given in Equations (10) and (12) respectively.

PROOF. The mean squared error of the improved class of estimators $T$ is given by

$$M(T) = E(T - \bar{H})^2 = E[\Omega(T_u - \bar{H}) + (1 - \Omega)(T_m - \bar{H})]^2 = \Omega^2 M(T_u) + (1 - \Omega)^2 M(T_m) + 2\Omega(1 - \Omega) \text{cov}(T_u, T_m) \quad (15)$$

The optimum values of $M(T_u)$ and $M(T_m)$ are computed in Equations (10) and (12) respectively as the estimators $T_u$ and $T_m$ are based on two independent samples of sizes $u$ and $m$ respectively. So, $\text{cov}(T_u, T_m) = 0$. Hence, substituting the optimum values of $M(T_u), M(T_m)$ and $\text{cov}(T_u, T_m)$ in the above Equation (15) we have the expression for the mean squared error of the general class of estimators $T$ as in Equation (14).

3.2. Minimum mean squared error of the proposed improved class of estimator $T$

The mean squared error of improved class of estimators $T$ in Equation (14) is a function of unknown constant $\Omega$ therefore, it is optimized with respect to $\Omega$ and subsequently the optimum value of $\Omega$ is obtained as

$$\Omega_{opt.} = \frac{M[T_m]_{opt.}}{M[T_u]_{opt.} + M[T_m]_{opt.}}. \quad (16)$$
Substituting the value of \( \Omega_{o,p} \) from Equation (16) in Equation (14), we get the optimum mean squared error of the class of estimator \( T \) as

\[
M[T]_{o,p} = \frac{M[T_u]_{o,p} \times M[T_m]_{o,p}}{M[T_u]_{o,p} + M[T_m]_{o,p}}.
\]

Further, substituting the values of \( M[T_u]_{o,p} \) and \( M[T_m]_{o,p} \) from Equations (10) and (12) respectively in Equation (17), the simplified values of \( M[T]_{o,p} \) is derived as

\[
M[T]_{o,p} = \frac{L_1 \mu - L_2}{(\mu)^2 K_3 - \mu L_3 - K_1} \left( \frac{S_h^2}{n} \right),
\]

where

\[
K_1 = 1 + K^2 Q_1 + 2KQ_2,
K_2 = 1 + M_2^2 + M_2^2 \hat{H}_2^2 Q_1 + 2M_2 \rho_b g + 2M_2 \hat{H}_2 Q_2 - 2M_2^2 \hat{H}_2 \tilde{\eta}_{12}^o,
K_3 = M_2^2 \hat{J}_2^2 Q_1 - 2M_2 \hat{J}_2 Q_2 - M_2^2 - M_2^2 \hat{H}_2^2 Q_1 - 2M_2 \rho_b g - 2M_2 \hat{H}_2 Q_2 + 2M_2^2 \hat{H}_2 \tilde{\eta}_{12}^o,
\hat{J}_2 = \frac{Q_2}{M_2 Q_1}, \quad \hat{H}_2 = \frac{M_2 \tilde{\eta}_{12}^o - Q_2}{M_2 Q_1},
L_1 = K_1 K_3, \quad L_2 = K_1 K_3 + K_1 K_3, \quad L_3 = K_2 + K_3 - K_1, \quad K = \frac{-Q_2}{Q_1}.
\]

3.3. Optimum replacement strategy

The mean squared error of improved class of estimator \( T \) from Equation (18) is a function of \( \mu \) which are the rotation rates or the fractions of sample to be drawn afresh at current occasion. It is also an important factor in reducing the cost of the survey, hence to estimate population mean with maximum precision and minimum cost, mean squared error of the class of estimators \( T \) derived in Equation (18) have been optimized with respect to \( \mu \). The optimum value of \( \mu \) say \( \hat{\mu}_f \) is given by

\[
\hat{\mu}_f = \min \left\{ \frac{I_2 \pm \sqrt{(I_2)^2 - I_1 I_3}}{I_1} \right\} \in [0, 1],
\]

where

\[
I_1 = L_1 K_3, \quad I_2 = L_2 K_3, \quad I_3 = L_1 K_1 + L_3 L_2.
\]

Replacing the optimum value of \( \mu \) from Equation (19) in Equation (18), the minimum mean squared error of \( T \) is obtained as

\[
M[T]_{\min} = \frac{L_1 \hat{\mu}_f - L_2}{(\hat{\mu}_f)^2 K_3 - \hat{\mu}_f L_3 - K_1} \left( \frac{S_h^2}{n} \right).
\]
3.4. Mean squared error of sensitive population mean estimator

The mean squared error of the estimator for sensitive population mean $\hat{Y}$ is obtained as

$$M[\hat{Y}] = \frac{M[T]_{\text{min}}}{[\phi_y + (1 - \phi_y)\bar{W}]^2}$$  \hspace{1cm} (21)

4. Comparison of the estimators

To judge the performance of the proposed improved class of estimators for sensitive population mean $\hat{Y}$, it has been compared with modified (Singh and Pal, 2015, 2017) which are described below.

(i) The recent estimator by Singh and Pal (2015) have been modified for estimating sensitive population mean at current occasion which is given as

$$\hat{Y}_{s1} = \frac{T_{s1} - \phi_y \bar{W}_1}{\phi_y + (1 - \phi_y)\bar{W}_2},$$  \hspace{1cm} (22)

where

$$T_{s1} = \Omega_{s1} \tilde{h} \exp\left(\frac{\hat{Z} + C_z}{\hat{z}_u + C_z}\right) + (1 - \Omega_{s1}) \tilde{h}_m \left(\frac{\hat{g}_m}{\hat{g}_n}\right) \left(\frac{\hat{Z} + C_z}{\hat{z}_u + C_z}\right) \text{ with } \Omega_{s1} \in (0, 1).$$

(ii) The estimator by Singh and Pal (2017) have also been modified for estimating sensitive population mean at current occasion and is given as

$$\hat{Y}_{s2} = \frac{T_{s2} - \phi_y \bar{W}_1}{\phi_y + (1 - \phi_y)\bar{W}_2},$$  \hspace{1cm} (23)

where

$$T_{s2} = \Omega_{s2} \tilde{h} \exp\left(\frac{\hat{Z} - \hat{z}_u}{\hat{Z} + \hat{z}_u}\right) + (1 - \Omega_{s2}) \tilde{h}_m \exp\left(\frac{\hat{g}_n - \hat{g}_m}{\hat{g}_n + \hat{g}_m}\right) \exp\left(\frac{\hat{Z} - \hat{z}_u}{\hat{Z} + \hat{z}_u}\right),$$

with $\Omega_{s2} \in (0, 1)$. Further the mean squared error of $\hat{Y}_{s1}$ and $\hat{Y}_{s2}$ are computed and presented in Table 1, where

$$K_{11} = 1 + \theta(\theta - 2\rho_{bz}), \quad K_{12} = 2 - 2\rho_{bg}, \quad K_{13} = \theta(1 - 2\rho_{bz}) + (2\rho_{bg} - 1),$$
$$K_{21} = \frac{5}{4} - \rho_{bz}, \quad K_{22} = \frac{5}{4} - \rho_{bg}, \quad K_{23} = \rho_{bg} - \rho_{bz}, \quad L_{j1} = K_{j1}K_{j3},$$
$$L_{j2} = K_{j1}K_{j2} + K_{j1}K_{j3}, \quad L_{j3} = K_{j2} + K_{j3} - K_{j1}, \quad I_{j1} = L_{j1}K_{j3},$$
$$I_{j2} = L_{j2}K_{j3}, \quad I_{j3} = L_{j1}K_{j1} + L_{j3}L_{j2}, \quad \theta = \left(\frac{\hat{Z}}{\hat{Z} + C_z}\right); \quad j = 1 \& 2.$$
to the estimators due to modified Singh and Pal (2015, 2017) under three models. The percent relative efficiency of the proposed improved class of estimators, with respect to the estimators due to modified Singh and Pal (2015, 2017) under three models $M_{PB1}$, $M_{PB2}$, and $M_{PB3}$.

### 5. Special Cases

Following Remark 1, if we consider $\phi_x (\phi_y) = 1$ in mean squared errors of $\hat{Y}, \hat{Y}_{s1}$ and $\hat{Y}_{s2}$ given in Equation (21) and Table 1 respectively, we get the mean squared errors of these estimators under modified additive model (say $M_{PB1}$). Similarly by substituting $\phi_x (\phi_y) = 0$ in mean squared errors of $\hat{Y}, \hat{Y}_{s1}$ and $\hat{Y}_{s2}$ given in Equation (21) and Table 1 respectively, the mean squared errors are obtained under modified multiplicative model (say $M_{PB2}$) which are given in Table 2.

### 6. Efficiency Comparison

The percent relative efficiency of the proposed improved class of estimators, with respect to the estimators due to modified Singh and Pal (2015, 2017) under three models $M_{PB1}$,
$M_{PB_2}$ and $M_{DP}$ have been computed as

\[ E_{11} = \frac{M[\hat{Y}_{s11}]}{M[\hat{Y}_1]} \times 100, \]
\[ E_{21} = \frac{M[\hat{Y}_{s21}]}{M[\hat{Y}_1]} \times 100, \]
\[ E_{12} = \frac{M[\hat{Y}_{s12}]}{M[\hat{Y}_2]} \times 100, \]
\[ E_{22} = \frac{M[\hat{Y}_{s22}]}{M[\hat{Y}_2]} \times 100, \]
\[ E_{13} = \frac{M[\hat{Y}_{s1}]}{M[\hat{Y}]} \times 100, \]
\[ E_{23} = \frac{M[\hat{Y}_{s2}]}{M[\hat{Y}]} \times 100. \]

Remark 6. The two scrambling variables $W_1$ and $W_2$ used to code the true response through scrambled response models may follow any distribution. But following Pollock and Bek (1976) and Eichborn and Hayre (1983), we consider scrambling variable $W_1$ to follow normal distribution with mean 0 and variance 1. However, the scrambling variable $W_2$ has been assumed to follow normal distribution with mean 1 and variance 1.

7. Numerical illustration

In order to validate the theoretical results, numerical illustrations has been supplemented. A sensitive population with a non-sensitive auxiliary variable have been considered from Statistical Abstracts of United States as:
X: Rate of abortions in 2004
Y: Rate of abortions in 2007
Z: Number of residents in 2004.

The artificial data for $W_1$ and $W_2$ have also been generated as per assumption in Remark 6. The optimum value of $\hat{\beta}$'s and percent relative efficiencies $E_{ij}; i, j = 1$ and 2 have been computed for the above data and are presented in Table 3.
## TABLE 3

Numerical results.

<table>
<thead>
<tr>
<th>i</th>
<th>Scrambled response model</th>
<th>$\hat{\mu}<em>{s</em>{1i}}$</th>
<th>$\hat{\mu}<em>{s</em>{2i}}$</th>
<th>$\hat{\mu}_{f_i}$</th>
<th>$E_{s_{1i}}$</th>
<th>$E_{s_{2i}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$M_{PB1}$</td>
<td>0.7811</td>
<td>0.6677</td>
<td>0.7230</td>
<td>205.8170</td>
<td>147.8605</td>
</tr>
<tr>
<td>2</td>
<td>$M_{PB2}$</td>
<td>0.8461</td>
<td>0.6812</td>
<td>0.7934</td>
<td>197.0943</td>
<td>152.0356</td>
</tr>
<tr>
<td>3</td>
<td>$M_{DP}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>p</th>
<th>$\hat{\mu}_{f_i}$</th>
<th>$E_{s_{1i}}$</th>
<th>$E_{s_{2i}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.8428</td>
<td>0.6807</td>
<td>0.7894</td>
</tr>
<tr>
<td>0.2</td>
<td>0.8161</td>
<td>0.6754</td>
<td>0.7620</td>
</tr>
<tr>
<td>0.3</td>
<td>0.8030</td>
<td>0.6725</td>
<td>0.7455</td>
</tr>
<tr>
<td>0.4</td>
<td>0.7951</td>
<td>0.6706</td>
<td>0.7354</td>
</tr>
<tr>
<td>0.5</td>
<td>0.7898</td>
<td>0.6694</td>
<td>0.7287</td>
</tr>
<tr>
<td>0.6</td>
<td>0.7859</td>
<td>0.6685</td>
<td>0.7239</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7832</td>
<td>0.6679</td>
<td>0.7205</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7813</td>
<td>0.6676</td>
<td>0.7182</td>
</tr>
<tr>
<td>0.9</td>
<td>0.7802</td>
<td>0.6674</td>
<td>0.7168</td>
</tr>
</tbody>
</table>

## TABLE 4

Simulation results.

<table>
<thead>
<tr>
<th>j</th>
<th>Scrambled response model</th>
<th>Sets$^*$</th>
<th>$E_{s_{1j}}$</th>
<th>$E_{s_{2j}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$M_{PB1}$</td>
<td>I</td>
<td>211.8754</td>
<td>145.1343</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td>207.7061</td>
<td>147.3438</td>
</tr>
<tr>
<td>2</td>
<td>$M_{PB2}$</td>
<td>I</td>
<td>204.3984</td>
<td>144.5748</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td>201.4420</td>
<td>149.1114</td>
</tr>
<tr>
<td>3</td>
<td>$M_{DP}$</td>
<td>p</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>I</td>
<td>204.7290</td>
<td>144.8387</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td>201.5453</td>
<td>149.5322</td>
</tr>
<tr>
<td></td>
<td>0.2</td>
<td>I</td>
<td>205.3721</td>
<td>143.7815</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td>201.9164</td>
<td>147.2712</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>I</td>
<td>206.2934</td>
<td>144.1131</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td>202.6177</td>
<td>146.7898</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>I</td>
<td>207.1571</td>
<td>144.4481</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td>203.4838</td>
<td>146.5415</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>I</td>
<td>208.1681</td>
<td>144.7478</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td>204.1736</td>
<td>146.8231</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>I</td>
<td>209.1362</td>
<td>145.1926</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td>204.9386</td>
<td>147.2050</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>I</td>
<td>209.8869</td>
<td>145.8146</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td>205.9717</td>
<td>147.4142</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>I</td>
<td>210.4815</td>
<td>146.3900</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td>206.7884</td>
<td>147.7016</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>I</td>
<td>210.8246</td>
<td>146.8191</td>
</tr>
<tr>
<td></td>
<td></td>
<td>II</td>
<td>207.3104</td>
<td>147.9268</td>
</tr>
</tbody>
</table>

$^*$ Set I: $n = 20$, $u = 12$, $m = 8$ and Set II: $n = 20$, $u = 15$, $m = 5$. 

---

Advances in Estimation of Sensitive Issues on Successive Occasions
8. ILLUSTRATIVE SIMULATION BASED FINDINGS

Simulation studies has been carried out to show the applicability of the proposed improved class of estimators using Monte Carlo simulation for data mentioned in Section 8. Under the simulation study 5,000 different samples have been examined and the process have been repeated for different combination of constants termed as different sets. The simulated percent relative efficiency of proposed improved class of estimator under considered three scrambled response models $M_{PB1}$, $M_{PB2}$ and $M_{DP}$ with respect to recent estimator by Singh and Pal (2015, 2017) modified to work for sensitive issues respectively have been computed and are denoted as $E_{s1j}$ and $E_{s2j}$, respectively, where $j = 1, 2, 3$ denote the three considered models. The simulation results are presented in Table 4.

9. A VIVID ILLUSTRATION OF THE STRATEGY INCLUDING TRUE AND MASKED RESPONSES

It is well known that the estimators under scrambled response techniques are less efficient than the estimators obtained using direct questioning method. Hence, a comparison has been done for the scrambled response technique with respect to direct questioning method. In absence of scrambling mechanism at any occasion, the similar estimator under direct method is proposed as

$$T_D = \chi T_{ud} + (1 - \chi) T_{md}; \quad \chi \in [0, 1], \quad (24)$$

where

$$T_{ud} = F_{ud}(\bar{y}_u, a_u, b_u),$$

$$T_{md} = F_{md}(\bar{y}_m, t_1^*, t_2^*)$$

with $t_1^* = K_{1d}(\bar{x}_m, a_m, b_m)$, $t_2^* = K_{2d}(\bar{x}_n, a_m, b_m)$, \( (25) \)

where $F_{ud}, F_{md}, K_{1d}$ and $K_{2d}$ follow similar regularity conditions as stated in Section 2.2.2. The minimum mean squared error of $T_D$ is obtained as

$$M[T_D]_{\text{min}} = \frac{L_{d1} \hat{\mu}_d - L_{d2}}{(\hat{\mu}_d)^2 K_{d3} - \hat{\mu}_d L_{d3} - K_{d1}} \left( \frac{S^2}{n} \right), \quad (27)$$

with

$$\hat{\mu}_d = \frac{I_{d2} \pm \sqrt{(I_{d2})^2 - I_{d1} I_{d3}}}{I_{d1}}, \quad (28)$$
where

\[ K_{d1} = 1 + K_d^2 Q_1 + 2K_d Q_{d2}, \]

\[ K_{d2} = 1 + M_{d2}^2 + M_{d2}^2 \tilde{H}_{d2}^2 Q_{d1} + 2M_{d2} \rho_{yx} + 2M_{d2} \tilde{H}_{d2} Q_{d2} - 2M_{d2} \tilde{H}_{d2} \eta_{12}^o, \]

\[ K_{d3} = M_{d2}^2 \tilde{j}_{d2}^2 Q_{d1} - 2M_{d2} \tilde{j}_{d2} Q_{d2} - M_{d2}^2 \tilde{H}_{d2}^2 Q_{d1} - 2M_{d2} \rho_{yx} \]

\[ - 2M_{d2} \tilde{H}_{d2} Q_{d2} + 2M_{d2}^2 \tilde{H}_{d2} \eta_{12}^o, \]

\[ \tilde{j}_{d2} = \frac{Q_{d2}}{M_{d2}^2 Q_{d1}}, \quad \tilde{H}_{d2} = \frac{M_{d2}^2 \eta_{12}^o - Q_{d2}}{M_{d2} Q_{d1}}, \]

\[ L_{d1} = K_{d1} K_{d3}, \quad L_{d2} = K_{d1} K_{d2} + K_{d1} K_{d3}, \quad L_{d3} = K_{d2} + K_{d3} - K_{d1}, \]

\[ K_d = \frac{-Q_{d2}}{Q_{d1}}, \quad I_{d1} = L_{d1} K_{d3}, \quad I_{d2} = L_{d2} K_{d3}, \quad I_{d3} = L_{d1} K_{d1} + L_{d3} L_{d2}. \]

To examine the scrambling effect, the percent relative efficiencies with respect to direct method of the proposed estimators under three scrambled response models have been computed as

\[ E_{D1} = \frac{M[T_D]_{\min}}{M[\tilde{Y}_1]} \times 100, \quad E_{D2} = \frac{M[T_D]_{\min}}{M[\tilde{Y}_2]} \times 100, \quad E_{D3} = \frac{M[T_D]_{\min}}{M[\tilde{Y}]} \times 100. \quad (29) \]

From the data mentioned in Section 7, the percent relative efficiencies have been scrutinized for different choices of \( p = 0.1, 0.2, 0.3, 0.4, \ldots, 0.9 \) where \( p \in \{ \phi_y, \phi_x \} \). The results obtained under three different models are presented in Figure 1. The optimum value of fraction of sample to be drawn afresh at current occasion for proposed general class of estimator under the three scrambled response models for varying model parameter with respect to direct questioning method has been shown in Figure 1 and Figure 2 respectively.

![Figure 1](image_url) - Percent relative efficiencies with respect to direct method under three scrambled response models in two occasions successive sampling.
10. INTERPRETATION OF RESULTS AND EPILOGUE

The following interpretations can be drawn from the empirical and simulation studies.

1. Observations from Table 3:
   
   (a) It is observed that the proposed improved class of estimators turns out to be more efficient than the modified estimators by Singh and Pal (2015, 2017) under all the three considered scrambled response models.
   
   (b) The optimum value of fraction of fresh sample to be drawn exists for all the estimators under all considered models.
   
   (c) The improved class of estimator $\hat{\bar{Y}}$ proves to be more efficient for higher values of $\phi_x(\phi_y)$ in comparison to lower values when compared with the estimator $\hat{\bar{Y}}_{s1}$ and for the estimator $\hat{\bar{Y}}_{s2}$, proves to be more efficient for lower values of $\phi_x(\phi_y)$ in comparison to higher values.
   
   (d) The model $M_{PB1}$ yields more efficiency than $M_{PB2}$ and $M_{DP}$ for the estimator $\hat{\bar{Y}}_{s1}$. However, for the estimator $\hat{\bar{Y}}_{s2}$, the model $M_{PB2}$ yields more efficiency than $M_{PB1}$ and $M_{DP}$.

2. Observations from Table 4:
   
   (a) The proposed improved class of estimator is performing more efficiently than $\hat{Y}_{s1}$ and $\hat{Y}_{s2}$ respectively in terms of percent relative efficiency.
   
   (b) For Set-I and Set-II, $E_{s1j} \geq E_{s2j}$ $\forall$ $j = 1, 2, 3$. 

\[ Figure \ 2 \ - \ \text{Optimum value of fraction of sample drawn afresh for proposed improved class of estimator under three scrambled response models and under direct method.} \]
(c) In simulation, similar behaviour of proposed improved class of estimator is observed as in empirical studies.

3. From Figure 1, it is clear that the proposed improved class of estimators under scrambled response model $M_{DP}$ is performing better than $M_{PB1}$ when compared with direct method for varying model parameter, the model $M_{DP}$ efficiency decreases for higher values of model parameter which is in accordance with Diana and Perri (2010). However, more efficiency have been observed in $M_{PB2}$ when compared with direct method.

4. Figure 2 reflects that the optimum fraction of sample to be drawn afresh for improved class of estimator under $M_{DP}$ decreases with increasing value of model parameter. Therefore, the proposed improved class of estimators provides several advantages with the scrambled response models considered. The numerical applications through simplistic simulations reflects how the proposed improved class of estimator is fare in practice. Hence, it may be recommended for its practical use.

ACKNOWLEDGEMENTS

The authors are indebted to the referee for a careful reading of manuscript and valuable suggestions. The financial assistance from SERB, New Delhi is gratefully acknowledged. Authors sincerely acknowledged the free access to data by Statistical Abstracts of United states.

REFERENCES


Surveys related to sensitive issues are accompanied with social desirability response bias which flaw the validity of analysis. This problem became serious when sensitive issues are estimated on successive occasions. The scrambled response technique is an alternative solution as it preserve respondents anonymity. Therefore, the present article endeavours to propose an improved class of estimators for estimating sensitive population mean at current occasion using an innocuous variable in two occasion successive sampling. Detailed properties of the estimators are analysed. Optimum allocation to fresh and matched samples are obtained. Many existing estimators in successive sampling have been modified to work for sensitive population mean estimation under scrambled response technique. The proposed estimators has been compared with recent modified estimators. Theoretical considerations are integrated with empirical and simulation studies to ascertain the efficiency gain derived from the proposed improved class of estimators.

Keywords: Sensitive variable; Successive occasions; Scrambled response model; Population mean; Optimum matching fraction.