Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 72(2) 2018 147–156

Assessment of Scrambled Response on Second Call in Two-Occasion Successive Sampling under Non-Response

G.N. Singh, M. Khetan and S. Suman
Indian School of Mines, Dhanbad

Received 16 November 2016; Revised 31 January 2018; Accepted 23 June 2018

SUMMARY

This paper is an attempt to analyze the effect of modified Hansen and Hurwitz (1946) technique to estimate the current population mean in case of quantitative sensitive variable in two-occasion successive sampling. The problem of evaluating the degree of privacy protection is also considered. Randomized linear model and their special cases have been considered on current occasion and properties of the proposed estimation strategy have been examined. Simulation studies are carried out to evaluate the performances of the proposed estimator with respect to estimators under (i) complete response and(ii) under Hansen and Hurwitz (1946) technique. Analytical and numerical comparisons of some well-known scrambled models have been carried out in terms of efficiency and privacy protection. Results have proved the effectiveness of the suggested estimation strategy in real life situations.

Keywords: Non-response, Successive sampling, Variance, Scrambled response, Privacy protection.

AMS Subject Classification: 62D05

1. INTRODUCTION

In many surveys, where study character is of sensitive nature such as related to gambling, alcoholism, sexual behavior, abortion, drug taking, tax evasion, illegal income, etc. may cause embarrassment or fear of social disapproval to the respondents. Obtaining valid and truthful information depends on the nature of the cooperation of the respondents and their willingness depends on the confidentiality of the responses. In this situation, to collect reliable data, Pollock and Bek (1976), Eichhorn and Hayre (1983), Bar-Lev *et al*. (2004), Saha (2007) and Diana and Perri (2010)worked on scrambled randomized response (SRR) methods. Sometimes, the survey researchers may have to face a realistic situations where characteristics of interest may be of sensitive nature for some respondents but may not be for others. To deal with such situations, Diana *et al.* (2014) proposed modified Hansen and Hurwitz (1946) technique in sample surveys.

The above mentioned works discussed for single occasion surveys, but longitudinal surveys focus on studying and analyzing the trends and dynamics of characteristics under study multiple times because one time survey on characteristics may not supply the desired information. For example, a survey on drug supply may be of interest in the following ways: (a) the average supply of a drug (say, cocaine) for a particular year; (b) the change in average supply of the drug over two different years; or (c) simultaneously to know both (a) and (b). For such types of situations, successive sampling proves to be more reliable in generating precise estimates over different occasions.

Very few attempts namely Arnab and Singh (2013) and Yu *et al.* (2014) have been found which dealt with the sensitive issues on successive occasions, while using randomized response technique. In successive sampling related to sensitive issues and motivated with the above arguments, in this paper, we have assessed the modified Hansen and Hurwitz (1946) technique to

Corresponding author: G.N. Singh
E-mail address: gnsingh_ism@yahoo.com

estimate the current population mean in two-occasion successive sampling under non-response. The goal is to encourage greater cooperation from the respondents and reduce the inclination towards responding their responses in a misleading fashion because of the above type of mentioned reason. Properties of the proposed estimator are examined under the randomized mechanism and their empirical performances are compared with other estimators. Some scrambled models have been considered to see the behaviors of the proposed estimator regarding efficiency and privacy protection. Results are interpreted and suitable recommendations have been made.

2. REVIEW OF HANSEN AND HURWITZ (1946) TECHNIQUE ON TWO SUCCESSIVE OCCASIONS

Let $U=(U_1,U_2,---,U_N)$ be a finite population of size *N*, which has been sampled over two occasions. The character under study is denoted by $x(y)$ on the first (second)occasion respectively. We assume that non-response occurs on both occasions, so the population can be divided into two classes, those who will respond at the first attempt and those who will not respond at the first attempt. Let $N_1^{'}$ and $N_2^{'}$ be the sizes of response and non-response classes respectively on the first occasion, similarly N_1 and N_2 be the sizes of response and non-response classes respectively on the second occasion. A random sample s_n of *n* units is drawn on the first occasion by using simple random sampling (without replacement) schemeand a random sub sample s_m of $m = n\lambda$ units are retained (matched) from the sample selected on the first occasion for its use on the second occasion. Let out of *m* matched units, m_1 units respond and m_2 units do not respond and out of m_2 non-responding units, a sub-sample of m_{2h} units selected for direct interview on both occasion. Similarly, let out of $u = (n\text{-}m)$ non-matched units on first occasion, $u_1^{'}$ units respond and $u_2^{'}$ units do not respond, and out of $u_2^{'}$ non-responding units, a subsample of $u_{h2}^{'}$ is selected for direct interview on first occasion. On the current (second) occasion a simple random sample (without replacement) s_u of $u=(n\text{-}m)=n\mu$ units are drawn afresh from the entire population, so that the sample size on the current (second) occasion is also *n*. Here λ and $\mu\ (\mu+\lambda=1)$ are the fractions of matched and fresh samples, respectively on the current (second) occasion. We assume that in the fresh sample of *u* units on the current (second) occasion, u_1 units respond and u_2 units do not respond. Let u_{2h} be the size of the sub sample drawn from the non-responding units in the fresh sample (s_u) on the current (second) occasion. The following notations are considered for the further use:

$\bar{X}$, $\bar{Y}$: The population means of the variables *x* and *y* respectively.

$\bar{x}_u^* = \dfrac{u_1^{'}\bar{x}_{u_1} + u_2^{'}\bar{x}_{u_{h2}}}{u}$, $\bar{x}_m^* = \dfrac{m_1\bar{x}_{m_1} + m_2\bar{x}_{m_{h2}}}{m}$: Hansen and Hurwitz (1946) estimators of the population mean $\bar{X}$ on the first occasion based on sample size u and m respectively.

$\bar{y}_u^* = \dfrac{u_1\bar{y}_{u_1} + u_2\bar{y}_{u_{h2}}}{u}$, $\bar{y}_m^* = \dfrac{m_1\bar{y}_{m_1} + m_2\bar{y}_{m_{h2}}}{m}$: Hansen and Hurwitz (1946) estimators of the population mean $\bar{Y}$ on the second occasion based on sample size u and m respectively. $S_x^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{X})^2, S_y^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})^2$: The population variances of the variables *x* and *y* respectively.

$S_{2x}^2 = \dfrac{1}{N_2^{'}-1}\sum_{i=1}^{N_2^{'}}(x_i-\bar{X}_2)^2, S_{2y}^2 = \dfrac{1}{N_2-1}\sum_{i=1}^{N_2}(y_i-\bar{Y}_2)^2$: The population variances of the variables *x* and *y* for non-response class respectively.

$\mu\left(=\dfrac{u}{n}\right)$, $\lambda\left(=\dfrac{m}{n}\right)$: The fractions of fresh and matched samples respectively.

$W = \dfrac{N_2}{N}$: Proportion of non-responding units in population.

$$f_2 = \frac{m_2}{m_{h_2}} = \frac{u_2^{'}}{u_{h_2}^{'}} = \frac{u_2}{u_{h_2}}.$$

3. MODIFIED HANSEN HURWITZ (1946) TECHNIQUE ON TWO OCCASIONS

When the characteristic under study becomes sensitive in nature, an assumption of Hansen Hurwitz (1946) technique i.e. all u_{2h} and m_{2h} units respond may not hold well, and if they do, their responses may not be in truthful manner. It is a serious problem in many surveys comprising the human population of misleading reporting and negation to respond. Even when the interviewers do their best to guarantee confidentiality, subjects can be sceptical and may be reluctant to supply truthful answers. To overcome this situation some modifications was carried out by Dianna *et al.* (2014) in Hansen and Hurwitz (1946) technique and later on it was used by Nameen and Sabbir (2017) in successive sampling on two occasion.

It is assumed that on the second occasion, persons who refuse to respond on first attempt due to their belief that the characteristic of interest is of sensitive nature are subject to scramble response on second attempt. The scrambled responses are used in the second phase of non-response to suggest response truthfully and secure the privacy of respondents.

For second call on the second occasion, let T_2 be the scrambled response and V_1 and V_2 be two scrambled variables, both are mutually independent having known mean (μ_{v_1}, μ_{v_2}) and variance $(\sigma^2_{v_1}, \sigma^2_{v_2})$.

We use the randomized linear model on second occasion as

$$T_2 = V_1 Y + V_2 \tag{1}$$

$$E_R(T_2) = \mu_{v_1} Y + \mu_{v_2} \tag{2}$$

$$V_R(T_2) = \sigma^2_{v_1} Y^2 + \sigma^2_{v_2}, \text{ since } \sigma_{v_1 v_2} = 0 \tag{3}$$

where (E_R, V_R) are the expectation and variance under randomization mechanism. We are assuming that the interviewer is completely unaware of the values generated by the respondents from the scrambled distributions V_1, V_2 and this assumption provides greater confidence among respondents about their privacy protection. Let $\hat{y}_i$ denote the transformation of randomized response of the i^{th} unit on current occasion whose expectation under the randomization mechanism coincides with the true response y_i

$$\hat{y}_i = \frac{t_{2i} - \mu_{v_2}}{\mu_{v_1}}$$

$$V_R(\hat{y}_i) = \frac{\sigma^2_{v_1} y_i^2 + \sigma^2_{v_2}}{\mu^2_{v_1}} = \psi_{2i} \tag{4}$$

The modified Hansen and Hurwitz (1946) estimators are defined as

$$\hat{\bar{y}}^{**}_m = \frac{m_1 \bar{y}_{m_1} + m_2 \hat{\bar{y}}_{m_{h_2}}}{m}$$

where $\hat{\bar{y}}_{m_{h_2}} = \sum_{i=1}^{m_{h_2}} \hat{y}_i / m_{h_2}$, which is an unbiased estimator of $\bar{Y}$ with variance

$$V(\hat{\bar{y}}^{**}_m) = \left(\frac{1}{m} - \frac{1}{N}\right) S^2_y + \frac{(f_2 - 1)W}{m} S^2_{2y} + \frac{f_2 W}{m N_2} \sum_{i=1}^{N_2} \psi_{2i} \tag{5}$$

Similarly $\hat{\bar{y}}^{**}_u = \frac{u_1 \bar{y}_{u_1} + u_2 \hat{\bar{y}}_{u_{h_2}}}{u}$

where $\hat{\bar{y}}_{u_{h_2}} = \sum_{i=1}^{u_{h_2}} \hat{y}_i / u_{h_2}$, which is an unbiased estimator of $\bar{Y}$ with variance

$$V(\hat{\bar{y}}^{*}_u) = \left(\frac{1}{u} - \frac{1}{N}\right) S^2_y + \frac{(f_2 - 1)W}{u} S^2_{2y} + \frac{f_2 W}{u N_2} \sum_{i=1}^{N_2} \psi_{2i} \tag{6}$$

See Diana *et al.* (2014).

The estimates of the variances of estimators $\hat{\bar{y}}^{**}_m$ and $\hat{\bar{y}}^{*}_u$ are given below:

$$\hat{V}(\hat{\bar{y}}^{**}_m) = \left(\frac{1}{m} - \frac{1}{N}\right) \hat{S}^2_y + \frac{(f_2 - 1)W}{m} \hat{S}^2_{2y} + \frac{f_2 W}{m N_2} \sum_{i=1}^{N_2} \psi_{2i}$$

and

$$\hat{V}(\hat{\bar{y}}^{*}_u) = \left(\frac{1}{u} - \frac{1}{N}\right) \hat{S}^2_y + \frac{(f_2 - 1)W}{u} \hat{S}^2_{2y} + \frac{f_2 W}{u N_2} \sum_{i=1}^{N_2} \psi_{2i}$$

4. FORMULATION OF ESTIMATOR

When study variable sensitive in nature, we proposed an estimation strategy to estimate the current population mean $\bar{Y}$ in two occasion successive sampling. Following Patterson (1950) and Hansen *et al.* (1953) we have proposed the estimator T^{**} as

$$T^{**} = \delta_1 \bar{x}^*_u + \delta_2 \bar{x}^*_m + \delta_3 \bar{y}^{**}_m + \delta_4 \bar{y}^{**}_u \tag{7}$$

where $(\delta_1, \delta_2, \delta_3, \delta_4)$ are suitably chosen scalars.

4.1 Properties of proposed estimator

The modified Hansen-Hurwitz estimator is unbiased. Further, following Artes and Garcia (2001), Garcia and Artes (2002) and Singh and Kumar (2008) and Singh *et al.* (2014), we have

$$E\left(\bar{x}_u^*\right) = E\left(\bar{x}_m^*\right) = \bar{X} \text{ and } E\left(\bar{y}_u^{**}\right) = E\left(\bar{y}_m^{**}\right) = \bar{Y} \quad (8)$$

Thus, from equations (7) and (8) we find that

$$E\left(T^{**}\right) = \left(\delta_1 + \delta_2\right)\bar{X} + \left(\delta_3 + \delta_4\right)\bar{Y} \quad (9)$$

For $E\left(T^{**}\right) = \bar{Y}$, we must have

$$\left(\delta_1 + \delta_2\right) = 0 \;\Rightarrow \delta_2 = -\delta_1$$

and $\left(\delta_3 + \delta_4\right) = 1 \;\Rightarrow \delta_4 = 1 - \delta_3$

Substituting the values of $\delta_2\ (= -\delta_1)$ and $\delta_4 (= 1 - \delta_3)$ in equation (9), we have the unbiased version of the estimator T^{**} as

$$T^{**} = \delta_1\left(\bar{x}_u^* - \bar{x}_m^*\right) + \delta_3 \bar{y}_m^{**} + \left(1 - \delta_3\right)\bar{y}_u^{**} \quad (10)$$

The covariance type terms are of order N^{-1}, hence for large population size, they are ignored. Hence,

$$Cov\left(\bar{y}_m^{**}, \bar{y}_u^{**}\right) = Cov\left(\bar{x}_u^*, \bar{x}_m^*\right) = Cov\left(\bar{x}_u^*, \bar{y}_m^{**}\right) = Cov\left(\bar{x}_u^*, \bar{y}_u^{**}\right) = Cov\left(\bar{x}_m^*, \bar{y}_u^{**}\right) = 0 \quad (11)$$

Thus we have the following theorem:

Theorem 1. The variance of T^{**} is obtained as

$$V\left(T^{**}\right) = \left[\frac{\delta_1^2 \eta_1^*}{\mu\lambda} + \frac{\delta_3^2 \eta_2^{**}}{\lambda} + \frac{\left(1 - \delta_3\right)^2 \eta_2^{**}}{\mu} - \frac{2\delta_1 \delta_3 \eta_3^*}{\lambda}\right]\frac{1}{n}, \quad (12)$$

where

$$\eta_1^* = S_x^2 + (f_2 - 1)WS_{2x}^2,$$

$$\eta_2^{**} = S_y^2 + (f_2 - 1)WS_{2y}^2 + \frac{f_2 W}{N_2}\sum_{i=1}^{N_2}\psi_{2i}$$

$\eta_3^* = \rho_{yx} S_x S_y + (f_2 - 1)W\rho_{2yx} S_{2x} S_{2y}$, and (λ and μ) are the fractions of matched and fresh samples for the estimator T^{**}.

Proof: The variance of T^{**} is given as

$$\begin{aligned} V\left(T^{**}\right) = {} & \delta_1^2\left[V\left(\bar{x}_u^*\right) + V\left(\bar{x}_m^*\right) - 2Cov\left(\bar{x}_u^*, \bar{x}_m^*\right)\right] + \delta_3^2 V\left(\bar{y}_m^{**}\right) \\ & + \left(1 - \delta_3\right)^2 V\left(\bar{y}_u^{**}\right) + 2\delta_1\delta_3\left[Cov\left(\bar{x}_u^*, \bar{y}_m^{**}\right) - Cov\left(\bar{x}_m^*, \bar{y}_m^{**}\right)\right] \\ & + 2\delta_1\left(1 - \delta_3\right)\left[Cov\left(\bar{x}_u^*, \bar{y}_u^{**}\right) - Cov\left(\bar{x}_m^*, \bar{y}_u^{**}\right)\right] \\ & + 2\delta_1\left(1 - \delta_3\right)Cov\left(\bar{y}_m^{**}, \bar{y}_u^{**}\right) \end{aligned}$$

Following the results of equation (11) we have

$$\begin{aligned} V\left(T^{**}\right) = {} & \delta_1^2\left[V\left(\bar{x}_u^*\right) + V\left(\bar{x}_m^*\right)\right] + \delta_3^2 V\left(\bar{y}_m^{**}\right) \\ & + \left(1 - \delta_3\right)^2 V\left(\bar{y}_u^{**}\right) - 2\delta_1\delta_3 Cov\left(\bar{x}_m^*, \bar{y}_m^{**}\right) \quad (13) \end{aligned}$$

where

$$V(\bar{x}_u^*) = \frac{1}{u}S_x^2 + \frac{(f_2 - 1)W}{u}S_{2x}^2, V(\bar{x}_m^*) = \frac{1}{m}S_x^2 + \frac{(f_2 - 1)W}{m}S_{2x}^2$$

$$V(\bar{y}_m^{**}) = \frac{1}{m}S_y^2 + \frac{(f_2 - 1)W}{m}S_{2y}^2 + \frac{f_2 W}{mN_2}\sum_{i=1}^{N_2}\psi_{2i}, V(\bar{y}_u^{**}) = \frac{1}{u}S_y^2 + \frac{(f_2 - 1)W}{u}S_{2y}^2 + \frac{f_2 W}{uN_2}\sum_{i=1}^{N_2}\psi_{2i}$$

$$Cov(\bar{x}_m^*, \bar{y}_m^{**}) = \frac{1}{m}\rho_{yx} S_x S_y + \frac{1}{m}(f_2 - 1)W\rho_{2yx} S_{2x} S_{2y},$$

substituting the above expressions in equation (13), we have

$$V\left(T^{**}\right) = \begin{bmatrix} \delta_1^2\left[\frac{1}{u}\left\{S_x^2 + (f_2 - 1)WS_{2x}^2\right\} + \frac{1}{m}\left\{S_x^2 + (f_2 - 1)WS_{2x}^2\right\}\right] \\ +\delta_3^2\left[\frac{1}{m}\left\{S_y^2 + (f_2 - 1)WS_{2y}^2 + \frac{f_2 W}{N_2}\sum_{i=1}^{N_2}\psi_{2i}\right\}\right] \\ +\left(1 - \delta_3\right)^2\left[\frac{1}{u}\left\{S_y^2 + (f_2 - 1)WS_{2y}^2 + \frac{f_2 W}{N_2}\sum_{i=1}^{N_2}\psi_{2i}\right\}\right] \\ -2\delta_1\delta_3\left[\frac{1}{m}\left\{\rho_{yx} S_x S_y + (f_2 - 1)W\rho_{2yx} S_{2x} S_{2y}\right\}\right] \end{bmatrix}$$

$$\begin{aligned} V\left(T^{**}\right) = {} & \delta_1^2\left[\frac{1}{u}\eta_1^* + \frac{1}{m}\eta_1^*\right] + \delta_3^2\left[\frac{1}{m}\eta_2^{**}\right] \\ & + \left(1 - \delta_3\right)^2\left[\frac{1}{u}\eta_2^{**}\right] - 2\delta_1\delta_3\left[\frac{1}{m}\eta_3^*\right] \end{aligned}$$

substituting the values of u and m in terms of μ and λ respectively in the above expression, we have

$$V\left(T^{**}\right)=\delta_1^2\left[\frac{1}{n\mu}\eta_1^*+\frac{1}{n\lambda}\eta_1^*\right]+\delta_3^2\left[\frac{1}{n\lambda}\eta_2^{**}\right]$$
$$+\left(1-\delta_3\right)^2\left[\frac{1}{n\mu}\eta_2^{**}\right]-2\delta_1\delta_3\left[\frac{1}{n\lambda}\eta_3^*\right]$$

$$V\left(T^{**}\right)=\delta_1^2\frac{1}{n}\left[\frac{\lambda+\mu}{\mu\lambda}\eta_1^*\right]+\delta_3^2\left[\frac{1}{n\lambda}\eta_2^{**}\right]$$
$$+\left(1-\delta_3\right)^2\left[\frac{1}{n\mu}\eta_2^{**}\right]-2\delta_1\delta_3\left[\frac{1}{n\lambda}\eta_3^*\right]$$

as we know that $\lambda+\mu=1$, so we have

$$V\left(T^{**}\right)=\left[\frac{\delta_1^2\eta_1^*}{\mu\lambda}+\frac{\delta_3^2\eta_2^{**}}{\lambda}+\frac{\left(1-\delta_3\right)^2\eta_2^{**}}{\mu}-\frac{2\delta_1\delta_3\eta_3^*}{\lambda}\right]\frac{1}{n}$$

Since the variance of the proposed estimator T^{**} is the functions of unknown constants δ_1 and δ_3 therefore, $V(T^{**})$ is minimized with respect to δ_1 and δ_3, and subsequently we get the optimum values of δ_1 and δ_3 and finally, optimum variance of the estimator T^{**} is obtained as

$$\delta_{1_{opt}}=\frac{\lambda\mu\eta_3^*\eta_2^{**}}{\left(\eta_1^*\eta_2^{**}-\mu^2\eta_3^{*2}\right)},\ \delta_{3_{opt}}=\frac{\lambda\eta_1^*\eta_2^{**}}{\left(\eta_1^*\eta_2^{**}-\mu^2\eta_3^{*2}\right)}$$

and

$$V\left(T^{**}\right)_{opt}=\frac{\eta_2^*\left(\eta_1^*\eta_2^{**}-\mu\eta_3^{*2}\right)}{\left(\eta_1^*\eta_2^{**}-\mu^2\eta_3^{*2}\right)}\frac{1}{n}. \tag{14}$$

4.2 Optimum replacement strategy

The idea of longitudinal surveys is mainly concerned with obtaining efficient estimates with minimal cost. To determine the optimum value of μ (fraction of sample to be drawn afresh on second occasion) so that the population mean $\overline{Y}$ may be estimated with maximum precision, we minimize variance of T^{**} given in equation (14) with respect to μ which results in quadratic equation in μ, the quadratic equation and respective solutions of μ say $\hat{\mu}$ are given below:

$$\mu^2\eta_3^{*2}-2\mu\eta_1^*\eta_2^{**}+\eta_1^*\eta_2^{**}=0 \tag{15}$$

$$\hat{\mu}=\frac{\eta_1^*\eta_2^{**}\pm\sqrt{\eta_1^*\eta_2^{**}\left(\eta_1^*\eta_2^{**}-\eta_3^{*2}\right)}}{\eta_3^{*2}} \tag{16}$$

From equation (16), it is obvious that the real values of $\hat{\mu}$ exist, if, the quantity under square root is greater than or equal to zero. For any value of correlation ρ_{yx}, which satisfy the condition of real solution; two real values of $\hat{\mu}$ are possible. Hence, while choosing the values of $\hat{\mu}$, it should be remembered that $0\le\hat{\mu}\le 1$. All other values of $\hat{\mu}$ are inadmissible. If both the values of $\hat{\mu}$ are admissible, lower one will be the best choice. Substituting the admissible value of $\hat{\mu}$ say $\mu^{(0)}$ from equation (16) into equation (14) respectively, we have the optimum variance of T^{**}, which are shown below:

$$V\left(T^{**}\right)_{opt}=\frac{\eta_2^{**}\left(\eta_1^*\eta_2^{**}-\mu^{(0)}\eta_3^{*2}\right)}{\left(\eta_1^*\eta_2^{**}-\mu^{(0)2}\eta_3^{*2}\right)}\frac{1}{\mathrm{n}}. \tag{17}$$

5. SOME SPECIAL CASES

In this section, we considered the special cases of the randomized linear model defined in equation (1) for different choices of V_1 and V_2. Putting the different value of V_1 and V_2 weget four known models which are additive, multiplicative and mixed in nature. We assess the efficiencies and privacy protection of the proposed estimator under these models. The models are given below:

S.N.	Model name	Special cases	Randomized linear model
1	M_1: Pollock and Bek additive model	When $V_1=1$	$T_{21}=Y+V_2$
2	M_2: Eichhorn and Hayre multiplicative model	When $V_2=0$	$T_{22}=V_1Y$
3	M_3: Saha mixed model	When $V_2=V_1V_2$	$T_{23}=V_1(Y+V_2)$
4	M_4: Diana *et al.* model	When $V_1=\alpha+(1-\alpha)V_1$ and $V_2=\alpha V_2$	$T_{24}=\begin{bmatrix}\alpha(V_2+Y)\\+(1-\alpha)V_1Y\end{bmatrix}$ where α lies in the interval (0, 1) to be chosen suitably.

Proceeding on the similar line as discussed for the proposed estimator T^{**}, the optimum variance of the estimator T^{**} under these four models arederived as

S.N.	Model name	MSE of the proposed estimator T^{**}
1	M_1: Pollock and Bek additive model	$\mathrm{M}\left(T^{**}_{M_1}\right)_{\text{opt}} = \frac{\eta_2^{**}\left(\eta_1^*\varepsilon_2^{**} - \mu^{(0)}\eta_3^{*2}\right)}{\left(\eta_1^*\varepsilon_2^{**} - \mu^{(0)2}\eta_3^{*2}\right)}\frac{1}{\text{n}}.$
2	M_2: Eichhorn and Hayre multiplicative model	$\mathrm{M}\left(T^{**}_{M_2}\right)_{\text{opt}} = \frac{\eta_2^{**}\left(\eta_1^*\vartheta_2^{**} - \mu^{(0)}\eta_3^{*2}\right)}{\left(\eta_1^*\vartheta_2^{**} - \mu^{(0)2}\eta_3^{*2}\right)}\frac{1}{\text{n}}.$
3	M_3: Saha mixed model	$\mathrm{M}\left(T^{**}_{M_3}\right)_{\text{opt}} = \frac{\eta_2^{**}\left(\eta_1^*\kappa_2^{**} - \mu^{(0)}\eta_3^{*2}\right)}{\left(\eta_1^*\kappa_2^{**} - \mu^{(0)2}\eta_3^{*2}\right)}\frac{1}{\text{n}}.$
4	M_4: Diana *et al.* model	$\mathrm{M}\left(T^{**}_{M_4}\right)_{\text{opt}} = \frac{\eta_2^{**}\left(\eta_1^*\nu_2^{**} - \mu^{(0)}\eta_3^{*2}\right)}{\left(\eta_1^*\nu_2^{**} - \mu^{(0)2}\eta_3^{*2}\right)}\frac{1}{\text{n}}.$

where

$$\varepsilon_2^{**} = S_y^2 + (f_2-1)WS_{2y}^2 + f_2W\sigma_{v_2}^2,\ \vartheta_2^{**} = S_y^2 + (f_2-1)WS_{2y}^2 + f_2W\left(\frac{\sigma_{v_1}^2\mu_{2,y}}{\mu_{v_1}^2}\right),$$

$$\kappa_2^{**} = S_y^2 + (f_2-1)WS_{2y}^2 + f_2W\left[\frac{\sigma_{v_1}^2\mu_{2,y} + 2\mu_{v_2}\sigma_{v_1}^2\bar{Y}_2 + \sigma_{v_2}^2\mu_{v_1}^2 + \sigma_{v_1}^2(\mu_{v_2}^2 + \sigma_{v_2}^2)}{\mu_{v_1}^2}\right],$$

and

$$\nu_2^{**} = S_y^2 + (f_2-1)WS_{2y}^2 + f_2W\left[\frac{(1-\alpha)^2\sigma_{v_1}^2\mu_{2,y} + \alpha^2\sigma_{v_2}^2}{\left\{\alpha + (1-\alpha)\mu_{v_1}\right\}^2}\right].$$

6. EFFICIENCY COMPARISON

6.1 Efficiency comparison with the similar estimator under complete response

The estimator T is defined under the similar circumstances as the estimator T^{**} but under complete response and it is given as

$$T = \gamma_1\bar{x}_u + \gamma_2\bar{x}_m + \gamma_3\bar{y}_m + \gamma_4\bar{y}_u \tag{18}$$

where $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ are suitably chosen scalars.

Proceeding on the similar line as discussed for the proposed estimator T^{**}, the optimum variance of the estimator T is derived as

$$\mathrm{V}(T)_{\text{opt}} = \frac{S_y^2\left[S_x^2S_y^2 - \mu'\left(\rho_{yx}S_yS_x\right)^2\right]}{\left[S_x^2S_y^2 - \mu'^2\left(\rho_{yx}S_yS_x\right)^2\right]}\frac{1}{n}. \tag{19}$$

where

$$\hat{\mu}' = \frac{S_x^2S_y^2 \pm \sqrt{S_x^2S_y^2\left(S_x^2S_y^2 - \left(\rho_{yx}S_yS_x\right)^2\right)}}{\left(\rho_{yx}S_yS_x\right)^2}$$

(fraction of the fresh sample for the estimator T),

$$\gamma_{1_{\text{opt}}} = \frac{\lambda'\mu'\bar{Y}\eta_3\eta_2}{\bar{X}\left(\eta_1\eta_2 - \mu'^2\eta_3^2\right)} \text{ and } \gamma_{3\text{opt}} = \frac{\lambda'\eta_1\eta_2}{\left(\eta_1\eta_2 - \mu'^2\eta_3^2\right)}$$

6.2 Efficiency comparison with the similar estimator under non- response

We consider the following linear unbiased combination of all the available information as an estimator of the current population mean $\bar{Y}$,

$$T^* = \beta_1\bar{x}_u^* + \beta_2\bar{x}_m^* + \beta_3\bar{y}_m^* + \beta_4\bar{y}_u^* \tag{20}$$

where $(\beta_1, \beta_2, \beta_3, \beta_4)$ are suitably chosen constants.

The expression of the optimum variance of the estimator T^* is given by:

$$\mathrm{V}\left(T^*\right)_{\text{opt}} = \frac{\eta_2^*\left(\eta_1^*\eta_2^* - \mu^{0(1)}\eta_3^{*2}\right)}{\left(\eta_1^*\eta_2^* - \mu^{0(1)2}\eta_3^{*2}\right)}\frac{1}{n}. \tag{21}$$

where

$$\mu^{0(1)} = \frac{\eta_1^*\eta_2^* \pm \sqrt{\eta_1^*\eta_2^*\left(\eta_1^*\eta_2^* - \eta_3^{*2}\right)}}{\eta_3^{*2}},$$

$$\eta_2^* = S_y^2 + (f_2-1)WS_{2y}^2,\ \beta_2 = 1-\beta_1,\ \beta_4 = 1-\beta_3,$$

$$\beta_{1_{opt}} = \frac{\bar{Y}\lambda^{(1)}\mu^{(1)}\eta_3^*\eta_2^*}{\bar{X}\left(\eta_1^*\eta_2^* - \mu^{(1)2}\eta_3^{*2}\right)} \text{ and}$$

$$\beta_{3_{opt}} = \frac{\lambda^{(1)}\eta_1^*\eta_2^*}{\left(\eta_1^*\eta_2^* - \mu^{(1)2}\eta_3^{*2}\right)}.$$

6.3 Efficiency comparison with the estimator (without matching from previous occasion) under non-response

The estimator ζ is defined under the similar circumstances as the estimator T^{**} when the information from previous occasion is not used and it is given as

$$\zeta = \lambda\bar{y}_m^{**} + \mu\bar{y}_u^{**} \tag{22}$$

The optimum variance of the estimator ζ is obtained in the similar way as discussed in section 4. and is given as

$$V(\zeta)_{opt} = \left[S_y^2 + (f_2 - 1)WS_{2y}^2 + \frac{f_2 W}{N_2} \sum_{i=1}^{N_2} \psi_{2i} \right] \frac{1}{n} \quad (23)$$

The percent relative efficiency in the precision of the proposed estimator T^{**} with respect to estimators ζ under their respective optimality conditions is given as

$$E = \frac{V(\zeta)_{opt}}{V(T^{**})_{opt}} \times 100. \quad (24)$$

Numerical illustrations:

Let us consider $N = 1,00,000$, $n = 500$ and 20% of the non-response rate. We assume that the study variable X on the first occasion follows Gamma probability distribution $\gamma(2.2,\ 3.5)$. Further, the study variable Y on the second occasion is explained by a model as $y_i = Rx_i + \in_i x_i^g$, where $\in \sim N(0,1)$, R=2.0 and g=1.5, which is related to the study variable X. The scrambled variables V_1 and V_2 are generated independently from U(0,1).

Table 1 presents the values of μ and MSE of the estimator T, T^* and T^{**} for different values of f_2.

Table 1. Optimum values of μ and mean square errors of the estimators T, T^* and T^{**}

f_2	2		3		4		5	
Estimator	μ	MSE	μ	MSE	μ	MSE	μ	MSE
T	0.5881	**2.3816**	0.5881	2.3816	0.5881	2.3816	0.5881	2.3816
T^*	0.5688	2.9496	0.5565	3.5125	0.5480	4.0729	0.5417	4.6316
T^{**}	**0.5487**	4.0161	0.5375	5.1045	0.5305	6.1898	0.5257	2736

Table 2.Presents the values of μ and percent relative efficiency of the proposed estimator T^{**} with the estimator ς.

Table 2. Optimum values of μ and percent relative efficiency E of the estimator T^{**} with respect to ς

f_2	**2**		**3**		**4**		**5**	
W	μ	E	μ	E	μ	E	μ	E
0.20	0.5487	**109**	0.5375	107	0.5305	106	**0.5257**	105

When characteristic under study is sensitive in nature on second occasion, the variance of the proposed estimator T^{**} is quite higher than the estimator T^* under non-response in two-occasion successive sampling. So, the proposed estimator is less efficient than the estimator T^* in terms of efficiency. But on the other hand, an important aspect is to know by using the randomized device how much privacy of the respondents is protected? In practice, the respondents are more concerned with high confidentiality than the researchers who are generally more interested in efficacious results. The proposed estimation procedure takes care of confidentiality of respondents as well as provides the reasonable degree of precision in estimates. Generally the privacy is more protected on the cost of enhanced variance in estimation procedures. The privacy and efficiency move in opposite directions, i.e. the low privacy protection results in high efficiency and vice-versa. Practically, it is more important to find a reasonable compromise between efficiency and privacy protection to obtain the truthful response from the respondents. Diana and Perri (2010) suggested the multiple correlation coefficients as a normalized privacy protection measure in the case of simple random sampling. Under two-occasion successive sampling, the same normalized measure is defined as follows:

$$\tau = 1 - \rho_{y.xt}^2, \quad (25)$$

where

$$\rho_{y.xt}^2 = \frac{\rho_{yx}^2 + \rho_{yt}^2 - 2\rho_{yx}\rho_{yt}\rho_{xt}}{1 - \rho_{xt}^2}$$

When $\tau = 1$, it specifies the maximum privacy protection, $\tau = 0$, it means that the privacy protection is almost negligible.

From equation (25), we compute the value of a normalized measure $\tau = 0.3228$, which indicates that a fair degree of privacy is protected if one uses the modified Hansen and Hurwitz (1946) technique suggested in this work.

Table 3. presents the values of μ, MSE and privacy protection τ of the proposed estimator T^{**} under models M_1, M_2 and M_3 for different values of f_2.

Table 3. Optimum values of μ and mean square errors of the estimator T^{**} under different models M_1, M_2 and M_3

f_2		2		3		4		5	
Estimator	τ	μ	**MSE**	μ	**MSE**	μ	**MSE**	μ	**MSE**
$T^{**}_{M_1}$	0.0529	0.5671	**3.0127**	0.5549	3.6067	0.5464	4.1980	0.5402	4.7876
$T^{**}_{M_2}$	0.2915	0.5524	3.7570	0.5409	4.7174	0.5335	5.6748	0.5284	6.6305
$T^{**}_{M_3}$	**0.3414**	**0.5458**	4.2486	0.5350	5.4518	0.5283	6.6521	0.5238	7.8509

Table 4. presents the values of μ, E and privacy protection τ of the proposed estimator T^{**} under model M_4 for different value of f_2 and α.

7. INTERPRETATIONS OF RESULTS

The following interpretations may be read out from *Tables 1-4:*

(1) From Table 1 it is clear that

(a) MSEs of the proposed estimator T^{**} and estimator T^* increase when we increase the values of f_2. This indicates that the size of the sub sample plays an important role in the performance of the proposed estimator.

(b) Optimum values of μ of the proposed estimator T^{**} and estimator T^* decrease when we increase the values of f_2. This behavior shows that for the larger size of the sub sample one may require to draw the smaller sample on the current occasion which reduces the cost of the survey.

(c) $M(T) < M(T^*) < M(T^{**})$ i.e. estimator T having the minimum MSE than other estimator, this behavior is on the expected line as the efficiencies of the estimators will increase when there is complete response.

(d) $\mu(T^{**}) < \mu(T^*) < \mu(T)$ i.e. optimum value of μ is lowest of the proposed estimator which is a highly desirable result.

From *Table 1*, it is concluded that the proposed estimator is better than the estimators T and T^* in terms of cost of the survey in two-occasion successive sampling.

(2) From Table 2 it is clear that

(a) Efficiencies of the proposed estimator T^{**} decrease when we increase the values of f_2.

(b) Optimum values of μ decrease when we increase the values of f_2.

From Table 2, it is concluded that the proposed estimator is superior to the estimator which does not use the information from the previous occasion under non-response. This behavior indicates that usefulness of the two-occasion successive sampling than the single occasion sampling.

(3) From Table 3 it is clear that

(a) MSEs of the proposed estimator T^{**} under models M_1, M_2 and M_3 increase when we increase the values of f_2.

Table 4. Optimum values of μ and mean square errors of the estimator T^{**} under model M_4

f_2		2		3		4		5	
α	τ	μ	**MSE**	μ	**MSE**	μ	**MSE**	μ	**MSE**
0	**0.2915**	**0.5524**	3.7570	0.5409	4.7174	0.5335	5.6748	0.5284	6.6305
0.1	0.2429	0.5569	3.4929	0.5449	4.3230	0.5372	5.1502	0.5317	5.9758
0.2	0.1951	0.5602	3.3166	0.5481	4.0600	0.5401	4.8004	0.5344	5.5392
0.3	0.1516	0.5627	3.1987	0.5505	3.8841	0.5423	4.5666	0.5364	5.2474
0.4	0.1150	0.5645	3.1205	0.5523	3.7674	0.5440	4.4115	0.5379	5.0539
0.5	0.0857	0.5657	3.0690	0.5535	3.6906	0.5451	4.3095	0.5390	4.9267
0.6	0.0664	0.5665	3.0380	0.5542	3.6444	0.5458	4.2481	0.5397	4.8501
0.7	0.0549	0.5669	3.0207	0.5547	3.6186	0.5462	4.2138	0.5400	4.8073
0.8	0.0498	0.5671	3.0132	0.5548	3.6075	0.5464	4.1990	0.5402	4.7889
0.9	0.0496	0.5671	3.0129	0.5549	3.6069	0.5464	4.1982	0.5402	4.7879
1	0.0529	0.5671	**3.0127**	0.5549	3.6067	0.5464	4.1980	0.5402	4.7876

(b) Optimum values of μ of the proposed estimator T^{**} under models M_1, M_2 and M_3 decrease when we increase the values of f_2. This behavior shows that for larger size of the sub sample one may requires to draw the smaller sample on the current occasion which reduces the cost of the survey.

(c) $M(T^{**}_{M_3}) > M(T^{**}_{M_2}) > M(T^{**}_{M_1})$ i.e. MSEs of the proposed estimator T^{**} have the lowest MSE under model M_1 while it has more MSEs under models M_2 and M_3.

(d) $\mu(T^{**}_{M_1}) > \mu(T^{**}_{M_2}) > \mu(T^{**}_{M_3})$ i.e. the optimum value of μ for the estimator T^{**} under model M_1 is maximum while it has lesser values under the models M_2 and M_3.

(e) $\tau(T^{**}_{M_3}) > \tau(T^{**}_{M_2}) > \tau(T^{**}_{M_1})$ i.e. the value of τ for the estimator T^{**} under model M_1 is minimum while it has larger values under the models M_2 and M_3.

From *Table 3,* it is concluded that the proposed estimator T^{**} under model M_1 have minimum values of MSE and privacy protection (τ) and maximum optimum value of fraction of fresh sample on current occasion. Further, the proposed estimator T^{**} under model M_3 have maximum values of MSE and privacy protection (τ) and minimum optimum value of fraction of fresh sample on current occasion.

(4) From Table 4 it is clear that

(a) For fixed value of α, MSE of the proposed estimator T^{**} increases with the increasing values of f_2. This behavior is similar to that discussed in 1 (a).

(b) For fixed value of α, optimum values of μ of the proposed estimator T^{**} decrease with the increasing values of f_2. This behavior is similar to that discussed in 3 (b).

(c) For fixed value of f_2, MSE of the proposed estimator T^{**} increases when we increase the values of α.

(d) For fixed value of f_2, optimum value of μ of the proposed estimator T^{**} decreases when we increase the values of α.

(e) When $\alpha = 0$ then $M(T^{**}_{M_4}) = M(T^{**}_{M_2})$ and $\mu(T^{**}_{M_4}) = \mu(T^{**}_{M_2})$

(f) When $\alpha = 1$ then $M(T^{**}_{M_4}) = M(T^{**}_{M_1})$ and $\mu(T^{**}_{M_4}) = \mu(T^{**}_{M_1})$.

(g) Privacy protection of the proposed estimator T^{**} under model M_4 decreases when we increase the value of α except for $\alpha = 1$.

(h) When $\alpha = 0$, the estimator T^{**} have minimum privacy protection with $\tau_{M_4} = 0.0529$, while for $\alpha = 1$, the estimator T^{**} have maximum privacy protection with $\tau_{M_4} = 0.2915$.

8. CONCLUSIONS AND RECOMMENDATIONS

From above analyses, we conclude that when study variable is sensitive in nature on current occasion, the modified Hansen and Hurwitz (1946) technique is preferable. The variance of the proposed estimator is slightly high, however, the privacy protection of respondents is also high which is very much desirable in randomizes response techniques. By using the trust worthy randomized mechanisms, there are some compromise between loss in efficiency and privacy protection of respondents. Hence the proposed estimator is preferable when the study variable is sensitive in nature. It is a good choice for the perspective of cost of the survey and privacy protection. Further, Proposed estimator under model M_3 is also more effective than any other model for the above type of situation. Finally, looking on the nice behaviors of the proposed estimator one may recommend them to the survey statisticians and practitioners for their practical applications.

ACKNOWLEDGEMENTS

Authors are thankful to the Indian Institute of Technology (Indian School of Mines), Dhanbad for providing financial assistance and necessary infrastructure to carry out the present research work. Authors are also thankful to the reviewer for their valuable suggestions which improved the quality of the paper.

REFERENCES

Artes, E. and Garcia, A.V. (2001): Estimating the current mean in successive sampling using a product estimate, *Conference on Agricultural and Environmental Statistical Application in Rome, XLIII-1, XLIII-2.*

Bar-Lev, S.K., Bobovitch, E., Boukai, B. (2004): A note on randomized response models for quantitative data. *Metrika,* **60(3)**, 255–260.

Choudhary, R.K., Bathla, H.V.L. and Sud, U.C. (2004): On Non-response in Sampling on Two occasions. *J. Ind. Soc. Agril. Statist.,* **58(3)**, 331-343.

Cochran, W.G. (1977): Sampling Techniques, *3rd edition, John Wiley & Sons, New York.*

Das, A.K. (1982): Estimation of population ratio on two occasions. *J. Ind. Soc. Agril. Statist.,* **34**, 1-9.

Diana, G. and Perri, P.F. (2010): New scrambled response models for estimating the mean of a sensitive quantitative character. *J. App. Statist.,* **37(11)**, 1875-1890.

Diana, G., Riaz, S., and Shabbir, J. (2014): Hansen and Hurwitz estimator with scrambled response on the second call. *J. App. Statist.,* **41(3)**, 596-611.

Eckler, A.R. (1955): Rotation Sampling. *Annals Math. Statist.,* **26**, 664-685.

Eichhorn, B., Hayre, L.S. (1983): Scrambled randomized response methods for obtaining sensitive quantitative data. *J. Statist. Plg. Inf.,* **7(4)**, 307-316.

Garcia, A.V. and Artes, E.M. (2002): Improvement on estimating of current population ratio in successive sampling, *Brazilian J. Probability Statist.,* **16(2)**, 107-122.

García Luengo A.V. and Oña Casado I. (2011): The non-response in the change of mean and the sum of mean for current occasion in sampling on two occasions. *Chilean J. Statist.,* **2(1)**, 63-84.

Hansen, M.H. and Hurwitz, W.N. (1946): The problem of the non-response in sample surveys *J. Amer. Statist. Assoc.,* **41**, 517-529.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953): Sample surveys methods and theory,1, 2, *John Wiley and Sons, Inc., New York and London.*

Jessen, R.J. (1942): Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agricultural Experiment Station Road Bulletin No. 304, Ames, USA, 1-104.*

Naeem, Noreen, and Javid Shabbir: Use of Scrambled Responses on Two Occasions Successive Sampling under Non-Response. (Accepted in *Hacettepe J. Math. Statist.* 2017)

Patterson, H.D. (1950): Sampling on successive occasions with partial replacement of units. *J. Roy. Statist. Soc.. B,* **12**, 241-255.

Pollock, K.H., Bek, Y. (1976): A comparison of three randomized response models for quantitative data. *J. Amer. Statist. Assoc.*, **71**, 884-886.

Raghunath, A. and Singh, S. (2013): Estimation of mean of sensitive characteristics for successive sampling. *Comm. Statist.-Theory Methods*, **42(14)**, 2499-2524.

Saha, A. (2007): A simple randomized response technique in complex surveys. *Metron*, LXV **(1)**, 59-66.

Singh, G.N. and Singh, V.K. (2001): On the use of auxiliary information in successive sampling. *J. Ind. Soc. Agril. Statist.,* **54(1)**, 1-12.

Singh, G.N. and Homa, F. (2013): Effective rotation patterns in successive sampling over two occasions. *J. Statist. Theory Practice,* **7(1)**, 146-155.

Singh, G.N. and Karna J.P. (2014): Estimation of population mean under non response in two- occasion rotation patterns. *Comm. Statist.-Theory Methods,* **43(16)**, 3429-3442.

Singh, G.N., Khetan, M. and Maurya, S. (2016): Estimation under non-response when it occurs on both occasions in two-occasion successive sampling. *Comm. Statist.-Theory Methods*, **45(13)**, 3939-3951.

Singh, H.P. and Kumar, S. (2010): Estimation of population product in presence of non-response in successive sampling. *Statistical Papers,* **51(4)**, 975-996.

Singh, V.K., Singh, G.N. and Shukla, D. (1991): An efficient family of ratio cum difference type estimators in successive sampling over two occasions. *J. Scien. Res.,* **41C**, 149-159.

Yu, B., Jin, Z. Tian, J. and Gao, Ge (2015): Estimation of sensitive proportion by randomized response data in successive sampling. *Computational and Mathematical Methods in Medicine,* DOI: 10.1155/2015/172918.