Use of Scrambled Responses on Two Occasions Successive Sampling under Non-Response

Noreen Naeem *† and Javid Shabbir ‡

Abstract

In this paper, we deal with a problem of non-response on two successive occasions when the study character becomes sensitive in nature on second occasion. Estimators are formulated by considering two cases of non-response, (i) when non-response on both occasions, (ii) when non-response on current occasion only. Expressions for mean squared errors (MSEs) are derived under large sample approximation and the optimum replacement strategies are also discussed. A numerical study is carried out in support of the proposed technique.

Keywords: Successive occasions, rotation pattern, scrambled response, mean squared error (MSE), optimum replacement policy, non-response.

2000 AMS Classification: AMS

1. Introduction

Now a days, sample surveys are not limited to one time observations. In many studies, especially in sociological and economic research, the character under study depends on time and it changes frequently with the passage of time according to its nature. In such situations, samples are selected on successive occasions (weekly, monthly, seasonally or annually), with the partial replacement of units, is called rotation sampling or repeated sampling on successive occasions.

Theory of successive sampling starts with the work of [8]. Later on [17] and [9] have extended his work. [11], [16], [10], [12], [6] have also contributed in the development of the theory of successive sampling.

In application, the surveyed units on successive occasions can be influenced with the problem of non-response. To deal with the problem of non-response on two occasions [2] have proposed a minimum variance linear unbiased estimator using the [7] technique. [15] and [1] have also discussed the theory under non-response on successive occasions.

When the characteristic under study becomes sensitive in nature under non-response, a serious difficulty in many surveys comprising the human population is misleading reporting and negation to respond. The randomized mechanism is designed to embolden the cooperation and honest comebacks to questions.

*Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan.
Email: noreen.qau90@gmail.com

†Corresponding Author.

‡Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan.
Email: javidshabbir@gmail.com

2. Review of Hansen and Hurwitz (HH) technique on two successive occasions

Consider a finite population Ω of size N. Let the study characteristics Y_h $(h = 1, 2)$ be sampled over two successive occasions. Let the population is divided into two classes; those who respond in the first attempt and those who do not respond. The sizes of these two classes are denoted by N_1 and N_2, respectively. On first occasion, a simple random sample of size n units are selected from N units. From which $m = n\lambda$ units are retained on the second occasion and an independent sample (unmatched with first) of $u = n\mu$ units are selected from the remaining population. We assume that in match portion of the sample m_1 units respond and m_2 do not. Similarly, u_1 units respond and u_2 units do not respond in the unmatched portion of the sample. Let $m_{h_2} = m_2/k$, $k > 1$, denotes the subsample of matched portion from the non-respondent group on two occasions for collecting information by personal interview. Similarly, for unmatch portion of the sample, u_{h_2} denotes the subsample from the non-response class on both occasions. Let σ^2 be the population variance and σ_2^2 be the population variance concerning to non-response class. Let ρ be the population correlation coefficient and ρ_2 be the population correlation coefficient pertaining to non-response class.
$S^2_{y_h}$: the population variance of the study variable for $h = (1, 2)$ occasion, $S^2_{y_h} = \frac{1}{N-1}\sum_{i=1}^{N}(y_{hi} - \bar{Y}_h)^2$, $(h = 1, 2)$
$S^2_{y_h}(2)$: the population variance of the study variable from the non-response class, $S^2_{y_h}(2) = \frac{1}{N_2-1}\sum_{i=1}^{N_2}(y_{hi} - \bar{Y}_h)^2$, $(h = 1, 2)$.

Now, we define the estimators for two occasions as follows

(a) When non-response is on both occasions

$$\bar{y}_{NR1} = \phi\bar{y}_{Rm_1} + (1-\phi)\bar{y}_{Ru_1}$$
$$\bar{y}_{Rm_1} = \bar{y}^*_{2m} + b^*_{y_2y_1m}(\bar{y}^*_{1n} - \bar{y}^*_{1m})$$
$$\bar{y}_{Ru_1} = \bar{y}^*_{2u}$$

(b) When non-response is on current occasion only

$$\bar{y}_{NR2} = \phi\bar{y}_{Rm_2} + (1-\phi)\bar{y}_{Ru_2}$$
$$\bar{y}_{Rm_2} = \bar{y}^*_{2m} + b_{y_2y_1m}(\bar{y}_{1n} - \bar{y}_{1m})$$
$$\bar{y}_{Ru_2} = \bar{y}^*_{2u}$$

where $\bar{y}^*_{hm}$, $(h = 1, 2)$ and $\bar{y}^*_{2u}$ are the Hansen-Hurwitz estimators;

$$\bar{y}^*_{hm} = \frac{m_1\bar{y}_{hm_1} + m_2\bar{y}_{hm_{h_2}}}{m}, \quad \bar{y}^*_{2u} = \frac{u_1\bar{y}_{2u_1} + u_2\bar{y}_{2u_{h_2}}}{u} \quad (h = 1, 2);$$

$b^*_{y_2y_1m}$ and $b_{y_2y_1m}$ are sample regression coefficients and ϕ is the constant.
Since the Hansen-Hurwitz estimator is an unbiased of the population mean, therefore, the estimators $\bar{y}_{NR1}$ and $\bar{y}_{NR2}$ are also unbiased to estimate the current occasion mean $\bar{Y}_2$. The expressions for their mean square errors (MSEs) and optimum replacement strategies are given by:

$$(2.1) \qquad MSE(\bar{y}_{NR1}) = \frac{E\{nE + u_1(F^* - G^*)\}}{n^2E + u_1^2(F^* - G^*)},$$

and

$$(2.2)\qquad MSE(\bar{y}_{NR2}) = \frac{E(nE - u_{2opt}F)}{n^2E - u_{2opt}^2F},$$

where $u_{1opt} = n\left[\frac{-E\pm\sqrt{E^2-E(F^*-G^*)}}{F^*-G^*}\right]$, $u_{2opt} = n\left[\frac{E\pm\sqrt{E^2-EF}}{F}\right]$ are the optimum replacement values and $E = S_{y_2}^2 + W_2(k-1)S_{y_2}^2(2)$, $F = S_{y_2}^2\rho_{y_1y_2}^2$, $F^* = \beta_{y_2y_1}^2\{S_{y_1}^2 + W_2(k-1)S_{y_1}^2(2)\}$ and $G^* = 2\beta_{y_2y_1}\{\rho_{y_2y_1}S_{y_2}S_{y_1} + W_2(k-1)\rho_{y_2y_1}(2)S_{y_2}S_{y_1}(2)\}$ are the substitutions to simplify the expressions.

3. Modification of HH technique on two occasions

When nature of the study characteristic becomes sensitive, it is quite difficult to make sure that all m_{h_2} and u_{h_2} units respond. If they do, then their responses are truthfull. To overcome this situation, some modifications are made with Hansen-Hurwitz model. On second occasion, we get response from the people in first attempt, but on second call people hesitate to reply the questions asked by the suveyer under sensitive characteristic. The scrambled responses are used on the second phase of non-response to evoke responses 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 means (μ_{v_1}, μ_{v_2}) and variances $(\sigma_{v_1}, \sigma_{v_2})$.
The randomized linear model on the current occasion can be written as:

$$(3.1)\qquad T_2 = V_1Y_2 + V_2$$

with

$$(3.2)\qquad E_R(T_2) = \mu_{v_1}Y_2 + \mu_{v_2}$$

$$(3.3)\qquad V_R(T_2) = \sigma_{v_1}^2Y_2^2 + \sigma_{v_2}^2, \qquad \text{since } \sigma_{v_1v_2} = 0$$

where $E_R(T_2)$ is the expected value and $V_R(T_2)$ is the variance under randomized mechanism. Here the assumption is followed that the surveyer is completely naive about those values generated from the scrambling distributions V_1 and V_2 by the respondents. This assumption build up the greater confidence among the people about the protection of their privacy.

Now let $\hat{y}_{2i}$ be the suitable transformation on second occasion of scrambled response t_{2i}, whose expectation matches with the true response y_{2i} under the randomization mechanism

$$(3.4)\qquad \hat{y}_{2i} = \frac{t_{2i} - \mu_{v_2}}{\mu_{v_1}}$$

with variance,

$$(3.5)\qquad V_R(\hat{y}_{2i}) = \frac{\sigma_{v_1}^2y_{2i}^2 + \sigma_{v_2}^2}{\mu_{v_1}^2} = \theta_{2i}$$

Now, the modified Hansen-Hurwitz estimator on the current (second) occasion can be written as:

(a) For matched portion:

$$\hat{\bar{y}}^*_{2m} = \frac{m_1 \bar{y}_{2m_1} + m_2 \hat{\bar{y}}_{2m_{h_2}}}{m}, \tag{3.6}$$

where $\hat{\bar{y}}_{2m_{h_2}} = \sum_{i=1}^{m_{h_2}} \hat{y}_{2i}/m_{h_2}$, with variance (followed by [4])

$$Var(\hat{\bar{y}}^*_{2m}) = \frac{1}{m} S^2_{y_2} + \frac{W_2(k-1)}{m} S^2_{y_2}(2) + \frac{kW_2}{mN_2} \sum_{i=1}^{N_2} \theta_{2i} \tag{3.7}$$

(b) For un-matched portion

$$\hat{\bar{y}}^*_{2u} = \frac{u_1 \bar{y}_{2u_1} + u_2 \hat{\bar{y}}_{2u_{h_2}}}{u}, \tag{3.8}$$

where $\hat{\bar{y}}_{2u_{h_2}} = \sum_{i=1}^{u_{h_2}} \hat{y}_{2i}/u_{h_2}$, with varaince (followed by [4])

$$Var(\hat{\bar{y}}^*_{2u}) = \frac{1}{u} S^2_{y_2} + \frac{W_2(k-1)}{u} S^2_{y_2}(2) + \frac{kW_2}{uN_2} \sum_{i=1}^{N_2} \theta_{2i} \tag{3.9}$$

Both estimators $\hat{\bar{y}}^*_{2m}$ and $\hat{\bar{y}}^*_{2u}$ are unbiased (see [4]).

4. Proposed estimators and their properties

The following estimators are proposed to capture the effect of rotation pattern under two cases of non-response:

Case-1: When non-response exists on both occasion

$$\bar{y}_{RS_1} = \phi \bar{y}^*_{Rm_1} + (1-\phi) \bar{y}^*_{Ru_1}, \tag{4.1}$$

where

$$\bar{y}^*_{Rm_1} = \hat{\bar{y}}^*_{2m} + b^*_{y_2 y_1 m}(\bar{y}^*_{1n} - \bar{y}^*_{1m}), \tag{4.2}$$

$$\bar{y}^*_{Ru_1} = \hat{\bar{y}}^*_{2u}. \tag{4.3}$$

Case-2: When non-response exists on current occasion only

$$\bar{y}_{RS_2} = \phi \bar{y}^*_{Rm_2} + (1-\phi) \bar{y}^*_{Ru_2}, \tag{4.4}$$

where

$$\bar{y}^*_{Rm_2} = \hat{\bar{y}}^*_{2m} + b_{y_2 y_1 m}(\bar{y}_1 - \bar{y}_{1m}), \tag{4.5}$$

$$\bar{y}^*_{Ru_2} = \hat{\bar{y}}^*_{2u}. \tag{4.6}$$

ϕ are the weights whose values are to be determined under certain criterion.
We use the following symbols to obtain the expressions of MSEs under large sample approximation.
Let

$$\hat{e}^*_{02m} = \frac{\hat{\bar{y}}^*_{2m} - \bar{Y}_2}{\bar{Y}_2}, \qquad \hat{e}^*_{02u} = \frac{\hat{\bar{y}}^*_{2u} - \bar{Y}_2}{\bar{Y}_2}, \qquad e^*_{01n} = \frac{\bar{y}^*_{1n} - \bar{Y}_1}{\bar{Y}_1}, e^*_{01m} = \frac{\bar{y}^*_{1m} - \bar{Y}_1}{\bar{Y}_1},$$

$$e_{01n} = \frac{\bar{y}_{1n} - \bar{Y}_1}{\bar{Y}_1}, \qquad e_{01m} = \frac{\bar{y}_{1m} - \bar{Y}_1}{\bar{Y}_1}$$

such that $E(\hat{e}^*_{02m}) = E(\hat{e}^*_{02u}) = E(e^*_{01n}) = E(e^*_{01m}) = E(e_{01n}) = E(e_{01m}) = 0$,

$$
\begin{aligned}
E(\hat{e}^{*2}_{02m}) &= \frac{1}{m}C^2_{y_2} + \frac{W_2(k-1)}{m}C^2_{y_2}(2) + \frac{kW_2}{\bar{Y}_2^{\,2} m N_2}\sum_{i=1}^{N_2}\theta_{2i}, \quad E(e^2_{01n}) = \frac{1}{n}C^2_{y_1},\\
E(\hat{e}^{*2}_{02u}) &= \frac{1}{u}C^2_{y_2} + \frac{W_2(k-1)}{u}C^2_{y_2}(2) + \frac{kW_2}{\bar{Y}_2^{\,2} u N_2}\sum_{i=1}^{N_2}\theta_{2i}, \quad E(e^2_{01m}) = \frac{1}{m}C^2_{y_1},\\
E(e^{*2}_{01n}) &= E(e^*_{01n}e^*_{01m}) = \frac{1}{n}C^2_{y_1} + \frac{W_2(k-1)}{n}C^2_{y_1}(2)\\
E(e^{*2}_{01m}) &= \frac{1}{m}C^2_{y_1} + \frac{W_2(k-1)}{m}C^2_{y_1}(2), \quad E(e_{01n}e_{01m}) = E(e_{01n}e^*_{01m}) = \frac{1}{n}C^2_{y_1},\\
E(\hat{e}^*_{02m}e^*_{01n}) &= \frac{1}{n}\rho_{y_1y_1}C_{y_1}C_{y_2} + \frac{W_2(k-1)}{n}\rho_{y_2y_1}(2)C_{y_1}(2)C_{y_2}(2)\\
E(\hat{e}^*_{02m}e^*_{01m}) &= \frac{1}{m}\rho_{y_1y_1}C_{y_1}C_{y_2} + \frac{W_2(k-1)}{m}\rho_{y_2y_1}(2)C_{y_1}(2)C_{y_2}(2)\\
E(e_{01n}\hat{e}^*_{02m}) &= \frac{1}{n}\rho_{y_1y_2}C_{y_1}Cy_2, \quad E(e_{01m}\hat{e}^*_{02m}) = \frac{1}{m}\rho_{y_1y_2}C_{y_1}Cy_2.
\end{aligned}
$$

Since the estimators $\bar{y}^*_{Rm_1}$, $\bar{y}^*_{Ru_1}$, $\bar{y}^*_{Rm_2}$ and $\bar{y}^*_{Ru_2}$ are unbiased of population means, as they are the linear combinations of Hansen-Hurwitz estimators. Therefore, expressions for their Biases and MSEs are given by

$$
\begin{aligned}
Bias(\bar{y}_{RS_1}) &= \phi Bias(\bar{y}^*_{Rm_1}) + (1-\phi)Bias(\bar{y}^*_{Ru_1}) = 0\\
Bias(\bar{y}_{RS_2}) &= \phi Bias(\bar{y}^*_{Rm_2}) + (1-\phi)Bias(\bar{y}^*_{Ru_2}) = 0
\end{aligned}
$$

$$MSE(\bar{y}_{RS_1}) = \phi^2 MSE(\bar{y}^*_{Rm_1}) + (1-\phi)^2 MSE(\bar{y}^*_{Ru_1}) \tag{4.7}$$

$$MSE(\bar{y}_{RS_2}) = \phi^2 MSE(\bar{y}^*_{Rm_2}) + (1-\phi)^2 MSE(\bar{y}^*_{Ru_2}), \tag{4.8}$$

where

$$
\begin{aligned}
MSE(\bar{y}^*_{Rm_1}) &= E(\bar{y}^*_{Rm_1} - \bar{Y}_2)^2\\
&= \frac{1}{m}\{S^2_{y_2} + W_2(k-1)S^2_{y_2}(2) + \frac{kW_2}{N_2}\sum_{i=1}^{N_2}\theta_{2i}\} + \left(\frac{1}{m} - \frac{1}{n}\right)\beta^2_{y_2y_1}\{S^2_{y_1}\\
&\quad + W_2(k-1)S^2_{y_1}(2)\} - \left(\frac{1}{m} - \frac{1}{n}\right)2\beta_{y_2y_1}\{\rho_{y_2y_1}S_{y_2}S_{y_1}\\
&\quad + W_2(k-1)\rho_{y_2y_1}(2)S_{y_2}S_{y_1}(2)\}, \qquad (4.9)\\
MSE(\bar{y}^*_{Ru_1}) &= E(\bar{y}^*_{Ru_1} - \bar{Y}_2)^2\\
&= \frac{1}{u}\{S^2_{y_2} + W_2(k-1)S^2_{y_2}(2) + \frac{kW_2}{N_2}\sum_{i=1}^{N_2}\theta_{2i}\} \qquad (4.10)
\end{aligned}
$$

Similarly,

$$MSE(\bar{y}^*_{Rm_2}) = E(\bar{y}^*_{Rm_2} - \bar{Y}_2)^2$$

$$= \frac{1}{m}\left[S^2_{y_2} + W_2(k-1)S^2_{y_2}(2) + \frac{kW_2}{N_2}\sum_{i=1}^{N_2}\theta_{2i}\right] - \left(\frac{1}{m} - \frac{1}{n}\right)S^2_{y_2}\rho^2_{y_2y_1}, \tag{4.11}$$

$$MSE(\bar{y}^*_{Ru_2}) = E(\bar{y}^*_{Ru_2} - \bar{Y}_2)^2$$

$$= \frac{1}{u}\left[S^2_{y_2} + W_2(k-1)S^2_{y_2}(2) + \frac{kW_2}{N_2}\sum_{i=1}^{N_2}\theta_{2i}\right] \tag{4.12}$$

Replacing Eq.(4.9) and Eq.(4.10) in Eq.(4.7), we get

$$MSE(\bar{y}_{RS_1}) = \phi^2\Bigg[\frac{1}{m}\{S^2_{y_2} + W_2(k-1)S^2_{y_2}(2) + \frac{kW_2}{N_2}\sum_{i=1}^{N_2}\theta_{2i}\} + \left(\frac{1}{m} - \frac{1}{n}\right)\beta^2_{y_2y_1}\{S^2_{y_1}$$

$$+ W_2(k-1)S^2_{y_1}(2)\} - \left(\frac{1}{m} - \frac{1}{n}\right)2\beta_{y_2y_1}\{\rho_{y_2y_1}S_{y_2}S_{y_1} + W_2(k-1)\rho_{y_2y_1}(2)S_{y_2}S_{y_1}(2)\}\Bigg]$$

$$+ (1-\phi)^2\left[\frac{1}{u}\{S^2_{y_2} + W_2(k-1)S^2_{y_2}(2) + \frac{kW_2}{N_2}\sum_{i=1}^{N_2}\theta_{2i}\}\right]$$

$$MSE(\bar{y}_{RS_1}) = \phi^2\left[\left(\frac{n}{mu}\right)E^* + \left(\frac{u}{mn}\right)F^* - \left(\frac{u}{mn}\right)G^*\right] - \frac{2\phi}{n}E^* + \frac{1}{u}E^* \tag{4.13}$$

Differentiate Eq.(4.13) w.r.t. ϕ, we get

$$\phi_{opt} = \frac{E^*nm}{n^2E^* + u^2(F^* - G^*)} \tag{4.14}$$

Now, replacing Eq.(4.11) and Eq.(4.12) in Eq.(4.8), we have

$$MSE(\bar{y}_{RS_2}) = \phi^2\left[\frac{1}{m}\left\{S^2_{y_2} + W_2(k-1)S^2_{y_2}(2) + \frac{kW_2}{N_2}\sum_{i=1}^{N_2}\theta_{2i}\right\} - \left(\frac{1}{m} - \frac{1}{n}\right)S^2_{y_2}\rho^2_{y_2y_1}\right]$$

$$+ (1-\phi)^2\frac{1}{u}\left[S^2_{y_2} + W_2(k-1)S^2_{y_2}(2) + \frac{kW_2}{N_2}\sum_{i=1}^{N_2}\theta_{2i}\right]$$

$$= \phi^2\left[\frac{n^2E^* - u^2F^*}{mnu}\right] - \frac{2\phi}{u}E^* + \frac{1}{u}E^* \tag{4.15}$$

Differentiate Eq.(4.15) w.r.t. ϕ, we get

$$\phi_{opt} = \frac{nmE^*}{n^2E^* - u^2F^*}. \tag{4.16}$$

Replacing respective optimum values of ϕ in Eq.(4.13) and Eq.(4.15), after simplification we have

$$MSE(\bar{y}_{RS_1}) = \frac{E^*\{nE^* + u(F^* - G^*)\}}{n^2E^* + u^2(F^* - G^*)}, \tag{4.17}$$

$$MSE(\bar{y}_{RS_2}) = \frac{(nE^{*2} - uE^*F)}{(n^2E^* - u^2F)}, \tag{4.18}$$

where

$$E^* = S_{y_2}^2 + W_2(k-1)S_{y_2}^2(2) + \frac{kW_2}{N_2}\sum_{i=1}^{N_2}\theta_{2i}, \quad F^* = \beta_{y_2y_1}^2\{S_{y_1}^2 + W_2(k-1)S_{y_1}^2(2)\}$$

$$G^* = 2\beta_{y_2y_1}\{\rho_{y_2y_1}S_{y_2}S_{y_1} + W_2(k-1)\rho_{y_2y_1}(2)S_{y_2}S_{y_1}(2)\} \quad \text{and} \quad F = S_{y_2}^2\rho_{y_1y_2}^2,$$

The optimum replacement policy of u for two cases of non-response is obtained such that Eq.(4.17) and Eq.(4.18) are minimized w.r.t. u. The optimum values of u are;

(i) *When non-response is on both occasions*

$$u_{1opt} = n\left[\frac{-E^* \pm \sqrt{E^{*2} - E^*(F^* - G^*)}}{(F^* - G^*)}\right] \tag{4.19}$$

(ii) *When non-response is on current occasion only*

$$u_{2opt} = n\left[\frac{E^* \pm \sqrt{E^{*2} + E^*F}}{F}\right], \tag{4.20}$$

Hence, the optimum MSEs for both cases of non-response are:

$$MSE(\bar{y}_{RS_1})_{opt} = \frac{E^*\{nE^* + u_{1opt}(F^* - G^*)\}}{n^2E^* + u_{1opt}^2(F^* - G^*)}, \tag{4.21}$$

and

$$MSE(\bar{y}_{RS_2})_{opt} = \frac{(nE^{*2} - u_{2opt}E^*F)}{(n^2E^* - u_{2opt}^2F)} \tag{4.22}$$

5. Efficiency of proposed estimator

Generally under the problem of non-response, the variance of estimators is quite high than the variance under simple random sampling. When the study characteristic is of sensitive nature, due to randomized mechanism, it can be seen theoretically that variance of the estimator is more high than the variance of Hansen-Hurwitz non-response estimator and SRS mean estimator, while conducting surveys on repeated occasions. For such situations, in terms of efficiency, we can say that our proposed estimator is less efficient, as it results in higher variance. But besides efficiency of the estimators, our main objective is to maximize the privacy protection of repsondents, to their responses on sensitive issues (information). To evoke true responses, we have to make some compromise between efficiecny loss of estimators and privacy disclosure of respondents. Due to randomized mechnaism, high value of mean squared error indicates the high privacy protection of respondents.

As authors suggested, while selecting the samples on two occasions, non-response may occur due to lack of contact or unavailability of repondents to provide desired information. Further, on second phase, even if the surveys for sensitive characterisitics are conducted, people who do not respond on first phase, are enthusiastic to response directly or use the randomization device. For this purpose, the optimal randomized response (ORR) procedure is implemented by refining the randomization stage that allows the respondents revealing the truth without randomizing the actual response, under direct questionning. In this case, second phase of non-response on current occasion can be written as:

$$\hat{y}_{2i} = I_{2i}y_{2i} + (1 - I_{2i})t_{2i}, \tag{5.1}$$

where I_{2i} denotes the indicator variable possesing value one if the i^{th} repondent is willing to response the true vaue y_{2i} and possessing value zero if the randomized response is used on current (second) occasion. By replacing the above transformation in Equations Eq.(3.6) and Eq.(3.8), we get an unbiased estimator of population mean $\bar{Y}_2$ and the variance in Equation Eq.(3.5) becomes $(1 - I_{2i})\theta_{2i}$ instead of θ_{2i}. This randomized response device not only reduce the variance of proposed estimator but also ensuring the maximum privacy protection to respondents.

The above discussion depicts only one-sided picture, that the high score of mean squared error of proposed estimator indicates less efficiency of estimator and high privacy protection of respondent. Now the question arise, how the suggested mechanism helps to maximize the repondent's confidence regarding their sensitive inofrmation and upto what extent their priavcy is protected via proposed technique. Practically, it is necessary to find a compromise solution between efficiency loss and privacy protection. A few literature is available to cope with this scenerio. [3] suggested the multiple correlation coefficient as a normalized privacy protection measure. Under two occasion successive sampling, the normalized measure is defined as follow:

$$\begin{aligned} \tau &= 1 - \rho^2_{y_2.y_1t_2}, \\ &= 1 - \frac{\rho^2_{y_2y_1} + \rho^2_{y_2t_2} - 2\rho_{y_2y_1}\rho_{y_2t_2}\rho_{y_1t_2}}{1 - \rho^2_{y_1t_2}} \end{aligned} \tag{5.2}$$

where $\rho_{y_ht_2}, (h = 1, 2)$ represents the correlation between the randomized model and study varaible on both occasions.
Now, when $\tau = 1$, it specifies the maximum privacy protection, when $\tau = 0$ it means that the privacy protection declines with less cooperation anticipated from respondents.

Now, we study the behaviour of MSEs of our proposed estimators for two cases of non-response and compare with the MSEs of Hansen-Hurwitz estimators under two occasions given in Eq.(2.1) and Eq.(2.2).

6. Numerical illustration

We consider two data sets for make valid comparison.
Population 1: The data set given in [4] is as follow:
Consider a population N of size 1,00,000 with 40% weight of missing values. The variable on the first occasion Y_1 follows $G(a, b)$ with pararmeters $(a = 2.2, b = 3.5)$ and the variable on the second occasion Y_2 which is related to the variable on the second occasion Y_1 is explained by a model as $y_{2i} = R_iy_{1i} + \epsilon_iy^g_{1i}$, where ϵ follows $N(0, 1)$, $R = 2.0$ and $g = 1.5$. The scrambled variables V_1 and V_2 are generated indepedentally from $U(0, 1)$. Also consider the simple random sample of size $n = 500$ without replacement.
Population 2: The data set of Bank Robberies for the year 2010 and 2011 from Bank Crime Statistics, Federal Bureau of Investigation (Bank Crime Statistics) is

considered. The descriptive information is as follow: $N = 54$, $n = 15$, with 40% weight of missing values
Year 2010 : $\bar{Y_1} = 103.5$ $S^2_{y_1} = 18338.9$
Year 2011 : $\bar{Y_2} = 92.7$ $S^2_{y_1} = 13524.6$

We find MSEs of our proposed estimators for different values of k and compare with the MSEs of Hansen-Hurwitz estimators on two occasions. In Table 1, MSEs values are found using Eq.(4.21) and Eq.(2.1) when non-response is on both occasions and in Table 2, MSE values are computed when non-response is on current occasion only using Eq.(4.22) and Eq.(2.2).

Table 1. MSE values, when non-response is on both occasions

Model		k	2	3	4	5
Modified HH ($\bar{y}_{RS_1}$)	Population 1	MSE	**349.0704**	**351.5563**	**353.5841**	**355.1573**
		μ_{opt}	0.2940	0.3110	0.4248	0.5089
	Population 2	MSE	**11548.43**	**11687.78**	**11826.34**	**11964.27**
		μ_{opt}	0.5077	0.5586	0.5694	0.5889
HH ($\bar{y}_{NR_1}$)	Population 1	MSE	345.9726	347.410	348.4773	349.2699
		μ_{opt}	0.2843	0.3984	0.4087	0.5196
	Population 2	MSE	11409.49	11578.11	11619.87	11794.91
		μ_{opt}	0.6595	0.6143	0.5897	0.5740

Table 2. MSE values, when non-response is only on current occasion

Model		k	2	3	4	5
Modified HH ($\bar{y}_{RS_2}$)	Population 1	MSE	**4.6956**	**6.2363**	**7.7769**	**9.3175**
		μ_{opt}	0.5063	0.5048	0.5038	0.5032
	Population 2	MSE	**1096.06**	**1457.465**	**1812.32**	**2164.323**
		μ_{opt}	0.6181	0.5839	0.5653	0.5536
HH ($\bar{y}_{NR_2}$)	Population 1	MSE	3.8283	4.9356	6.0425	7.1496
		μ_{opt}	0.5078	0.5060	0.5049	0.5042
	Population 2	MSE	866.1033	1125.802	1376.86	1624.191
		μ_{opt}	0.6595	0.6143	0.5897	0.5740

The bold values in Tables 1 and 2 represent the mean squared error of the proposed estimator under non-response with randomized mechanism. We have seen that in Tables 1 and 2, the mean estimator under non-response with randomized mechansim, i.e. the modified Hansen-Hurwitz (HH) estimator, yields the higher mean squared error than the mean estimator without randomization, whether non-response is on both occasion or only on current occasion. These values are maximium as compared to the mean squared errors of Hansen-Hurwitz (HH) model. Also their MSEs increase with increase in the value of k. Further using Eq.(5.2), we compute the value of normalized measure for both data sets. For population 1 τ_1=0.269 and for population 2 $\tau_2 = 0.397$. Both values are greater than 0, which indicates that atleast some of the respondent's privacy is protected due to randomization technique.

Hence the proposed estimators are more preferable, when the study characteristics is sensitive in nature. It is a good choice for the perspective of privacy protection.

7. Conclusion

We have proposed a modifed version of the Hansen-Hurwitz estimator, when the study characteristic becomes sensitve in nature on two successive occasions. It is assumed for the Hansen-Hurwitz technique, that all units respond on the second call truthfully. When nature of the study characteristic is sensitive, this assumption violates. To overcome this situation and obtain the answers in truthfull manner, the randomized mechanism is used on the second call of non-response on current occasion. Regression type estimators are proposed for estimation of second (current) occasion mean under two cases of non-response and their variances are used as a tool to measure not only efficiency but also the privacy protection of respondents. By using the trustworthy randomized mechanism, there is some compromise between efficiency loss and privacy protection. To measure confidentiality, a normalized privacy protection measure is used with indication of value one showing maximum privacy protection and zero with minimum privacy protection. The objective of conducting this study is to give some contributions for the collection of sensitive information by using randomized mechansim to obtain the truthful responses and ensure the respondents regarding their protection of privacy, when information on two successive occasions have been collected.

Acknowledgements
The authors wish to thank anonymous referees for their careful reading and helpful suggestions which led to improvement over an earlier version of this paper. Thanks to the Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan for support and motivation to carry out this research.

References

[1] Bandyopadhyay, A. and Singh, G. N. *Estimation of population mean in presence of non response in two-occasion successive sampling*, Recent Advances in Information Technology, 109-116, 2014.

[2] Choudhary, R., Bathla, H., and Sud, U. *On non-response in sampling on two occasions*, Journal of the Indian Society of Agricultural Statistics, **58** (3), 331-343, 2004.

[3] Diana, G. and Perri, P. F. *New scrambled response models for estimating the mean of a sensitive quantitative character*, Journal of Applied Statistics, **37** (11), 1875-1890.

[4] Diana, G., Riaz, S., and Shabbir, J. *Hansen and hurwitz estimator with scrambled response on the second call*, Journal of Applied Statistics, **41**(3), 596-611, 2014.

[5] FBI. *Violations of the Federal Bank Robbery and Incidental Crime Statute*, Federal Bereau of Investigation, 2011.

[6] Feng, S. and Zou, G. *Sample rotation method with auxiliary variable*, Communications in Statistics-Theory and Methods, **26** (6), 1497-1509, 1997.

[7] Hansen, M. H. and Hurwitz, W. N. *The problem of non-response in sample surveys.* Journal of the American Statistical Association, **41** (236), 517-529, 1946.

[8] Jessen, R. J. *Statistical investigation of a sample survey for obtaining farm facts*, Retrospective Theses and Dissertations, 1943.

[9] Patterson, H. *Sampling on successive occasions with partial replacement of units*, Journal of the Royal Statistical Society, **12** (2), 241-255, 1950.

[10] Prabhu-Ajgaonkar, S. *The theory of univariate sampling on successive occasions under the general correlation pattern 1, 2*, Australian Journal of Statistics, **10** (2), 56-63, 1968.

[11] Rao, J. and Graham, J. E. *Rotation designs for sampling on repeated occasions.* Journal of the American Statistical Association, **59** (306), 492-509, 1964.

[12] Sen, A. *Some theory of sampling on successive occasions*, Australian Journal of Statistics, **15** (2), 105-110, 1973.

[13] Singh, G. N., Majhi, D., Maurya, S., and Sharma, A. *Some effective rotation patterns in estimation of population mean in two-occasion successive sampling.* Communications in Statistics-Theory and Methods, **44** (12), 2571-2585, (2015).

[14] Singh, G. N. and Sharma, A. *An alternative rotation patterns in two occasions successive sampling*, International Journal of Mathematics and Statistics. **15**(3), 9-22, 2014.

[15] Singh, G. N. and Karna, J. P. *Some imputation methods to minimize the effect of non-response in two occasion rotation patterns*, Communications in Statistics-Theory and Methods, **39**(18), 3264-3281, 2010.

[16] Tikkiwal, B. *Theory of multiphase sampling from a finite or an infinite population on successive occasions 1, 2*, Revue de l'Institut International de Statistique, 247-263, 1967.

[17] Yates, F. *Sampling methods for censuses and surveys*, Charles Griffin & Co. Ltd., London, (1949)