Item sum techniques for quantitative sensitive estimation on successive occasions

Kumari Priyanka^{1,a}, Pidugu Trisandhya^a

^aDepartment of Mathematics, Shivaji College, University of Delhi, India

Abstract

The problem of the estimation of quantitative sensitive variable using the item sum technique (IST) on successive occasions has been discussed. IST difference, IST regression, and IST general class of estimators have been proposed to estimate quantitative sensitive variable at the current occasion in two occasion successive sampling. The proposed new estimators have been elaborated under Trappmann *et al.* (*Journal of Survey Statistics and Methodology*, **2**, 58–77, 2014) as well as Perri *et al.* (*Biometrical Journal*, **60**, 155–173, 2018) allocation designs to allocate long list and short list samples of IST. The properties of all proposed estimators have been derived including optimum replacement policy. The proposed estimators have been mutually compared under the above mentioned allocation designs. The comparison has also been conducted with a direct method. Numerical applications through empirical as well as simplistic simulation has been used to show how the illustrated IST on successive occasions may venture in practical situations.

Keywords: sensitive variable, successive occasions, class of estimators, population mean, variance, bias, mean squared error, optimum matching fraction

1. Introduction

Data gathering on sensitive, incriminating, stigmatizing or too personal issues are an avowedly baffling task. The three possibilities: under reporting, over reporting and refusal to answer the sensitive questions are prevalent. Generally, socially undesirable characteristics such as drug addiction, tax evasion, plagiarism, criminal conviction, and unauthorized natural resource use are likely to be under reported whereas socially desirable characteristics such as energy conservation, reducing pollution, ecological and biological conservation, and participation in elections are likely to be over reported. However some may refuse to report because of social stigma or fear of privacy violations. All these phenomenon provoke non sampling errors that are difficult to deal with and can seriously damage the validity of the analysis. In order to overcome mis-reporting on sensitive issues, many data collection strategies have been developed to elicit a more honest response from respondents by increasing the obscurity of the survey process and ensuring privacy protection.

One important method of survey addressing sensitive issue is the indirect questioning method. The indirect questioning technique may be classified in three different categories: the randomized response technique (RRT), the item count technique (ICT), and the non RRT. The RRT was initiated by Warner (1965), the ICT was originally proposed by Miller (1984) for binary variables to estimate the prevalence of a stigmatizing behavior within the population. However the non RRT was initiated recently by Tian and Tang (2014).

¹ Corresponding author: Department of Mathematics, Shivaji College, University of Delhi, New Delhi 110 027, India.
E-mail: priyanka.ism@gmail.com

In this paper we focus on the generalization of second technique the ICT. The ICT is used in surveys that require the study of qualitative variable. However, Chaudhuri and Christofides (2013) proposed a generalization to ICT for estimating a quantitative sensitive variable. Trappmann *et al.* (2014) named this technique as item sum technique (IST). Perri *et al.* (2018) discussed the optimal sample size allocation in the IST. Further enhancement in IST literature can be seen in Hussian *et al.* (2015), Rueda *et al.* (2017).

In many fields of applied research it has been observed that data needs to be gathered for a sensitive variables which can change over time. The statistical tool recommended for such a situation is successive sampling. Jessen (1942) initiated the idea of sampling same population over time with partial replacement of units called successive sampling. However, an analysis of a sensitive variable in successive sampling has been initiated by Arnab and Singh (2013). They applied RRT on successive occasions. Additional literature addressing sensitive issue over successive occasions can be seen in Yu *et al.* (2015), Naeem and Shabbir (2016), Singh *et al.* (2017), Priyanka *et al.* (2018), and Priyanka and Trisandhya (2018). These researchers focused on a scrambled response technique or RRT to handle sensitive issues on successive occasions. However, the IST is now emerging as an alternative technique to deal with sensitive issues. Hence, an attempt has been made in the present work to use IST in successive sampling to estimate a sensitive population mean. To the best of our knowledge this is an initial attempt and will contribute another useful method in successive sampling literature.

Therefore, the IST difference, the IST regression and also the IST general class of estimators have been proposed in the present work to estimate a sensitive population mean in two occasion successive sampling. All these proposed estimators have been discussed under Trappmann *et al.* (2014) allocation and Perri *et al.* (2018) optimal allocation designs. Detail properties including optimum replacement strategies are elaborated. The proposed methods have been compared mutually as well as with the corresponding direct method. Numerical applications through empirical simulation and simplistic simulation show how the illustrated IST on successive occasions may develop in practical situations. Finally some concluding remarks have been forwarded.

2. The item sum technique

The well-known technique in sensitive characteristics estimation is the ICT; however, the ICT is generally applicable for dichotomous (qualitative) variables only. Hence, the ICT was generalized by Chaudhuri and Christofides (2013) that can be used to estimate the quantitative sensitive variable. Later Trappmann *et al.* (2014) named this generalized version of ICT as IST and used it to estimate some quantitative sensitive variable. The algorithm for the IST is as follows.

From a random sample (s), two random sub-samples (s_{ll} and s_{sl}) are generated. The sub-sample s_{ll} , is confronted with a long list (LL) of items containing the sensitive question and a number of innocuous/non-sensitive questions. However the respondents in sub-sample s_{sl} have been given a short list (SL) of items containing only the innocuous questions present in the LL sample. Respondents in each sample are asked to report the total score of all items given to them, without disclosing the individual scores for the items. The mean difference of the answers between the s_{ll} and s_{sl} is used as an unbiased estimator of the population mean of the sensitive variable. All sensitive and innocuous variables should be quantitative in nature and measured on the same scale as that of the sensitive variable in the IST.

The decisive point in the IST is how to split the total sample in to the LL sample and SL sample. Trappmann *et al.* (2014) allocated the same number of units to each sample irrespective of the vari-

Table 1: Response received under item sum technique (IST)

Occasion	Sample size	Response received	IST estimate
I	n	$z_{1i} = \begin{cases} x_i + t_{1i}, & \text{if } i \in s_{nll}, \\ t_{1i}, & \text{if } i \in s_{nsl}, \end{cases}$	$\hat{\bar{x}}_n = \bar{z}_{1nll} - \bar{t}_{1nsl}$
	m	$z_{1i} = \begin{cases} x_i + t_{1i}, & \text{if } i \in s_{mll}, \\ t_{1i}, & \text{if } i \in s_{msl}, \end{cases}$	$\hat{\bar{x}}_m = \bar{z}_{1mll} - \bar{t}_{1msl}$
II	m	$z_{2i} = \begin{cases} y_i + t_{2i}, & \text{if } i \in s_{mll}, \\ t_{2i}, & \text{if } i \in s_{msl}, \end{cases}$	$\hat{\bar{y}}_m = \bar{z}_{2mll} - \bar{t}_{2msl}$
	u	$z_{2i} = \begin{cases} y_i + t_{2i}, & \text{if } i \in s_{ull}, \\ t_{2i}, & \text{if } i \in s_{usl}, \end{cases}$	$\hat{\bar{y}}_u = \bar{z}_{2ull} - \bar{t}_{2usl}$

z_{ji} ; $j = 1, 2$ denote the observed response at first and second occasion respectively on the i^{th} observation.

$\bar{z}_{jll}$; $j = 1, 2$; $i \in \{n, m, u\}$ denote the mean of z_j in the long list (LL) samples.

$\bar{t}_{jisl}$; $j = 1, 2$; $i \in \{n, m, u\}$ denote the mean of t_1 and t_2 in the short list (SL) samples.

ation of items in the two lists. However, Perri *et al.* (2018) advocated the requirement of optimum allocation of LL and SL samples. The IST may be modified to deal with sensitive issues on successive occasions if the sensitive variable is also changing by time, which is often the scenario. For example if the sensitive variable is the amount spent per month on drugs such as cigarettes or pan masala by college students, then the non-sensitive variable may be taken as the total monthly pocket money received by them or the amount spent on purchasing books. Similarly, if the sensitive variable is the number of abortions, then the non-sensitive variable may be the number of children or total number of members in that family. The sensitive question together with non-sensitive questions will comprise of LL sample; however, only non-sensitive questions will comprise the SL sample. There may be any number of non-sensitive question with a sensitive question to be used for LL sample and the same non-sensitive questions to be used for the SL sample. But in this paper we considered one sensitive and one non-sensitive question case on successive occasions.

3. Proposed IST frame work in successive sampling design

Consider a finite population P consisting of N identifiable units for sampling over two successive occasions. Let x denote the quantitative sensitive variable at the first occasion which changes to y at second occasion. Similarly let t_1 be the non sensitive/innocuous variable at the first occasion which changes to t_2 at the second occasion. Assume that x_i, y_i, t_{1i} , and t_{2i} denotes the value of x, y, t_1 , and t_2 respectively on the unit $i \in P$. To estimate the population mean of quantitative sensitive variable $\bar{Y}$ at current occasion using the IST, the sampling design is proposed as:

At first occasion a sample of size n is drawn using simple random sample without replacement (SR-SWOR) which has been split to s_{nll} and s_{nsl} samples called the LL-sample and SL-sample respectively. Now, at the second occasion considering the partial overlap case, two independent samples are considered, one is a matched sample of size $m = n\lambda$ drawn as SRSWOR sub-sample from sample size n at first occasion and second is a fresh sample of size $u = (n - m) = n\mu$, which is drawn afresh at current occasion. Further, the samples of sizes m and u are split into corresponding LL-sample and SL-samples as s_{mll} , s_{msl} , s_{ull} , and s_{usl} respectively. The response obtained from the respondents on two occasions and the corresponding IST estimate based on different samples are presented in Table 1.

4. IST successive difference estimator

In order to utilize information available from previous occasion an IST difference type estimator $\mathbb{T}_{1m}$ is considered based on sample of size m retained from previous occasion and the estimator based on

sample of size u is the IST estimator $\mathbb{T}_{1u} = \hat{y}_u$. Combining the two estimators as the convex linear combinations, the final estimator for sensitive population mean at current occasion is given by

$$\mathbb{T}_1 = \phi_1 \mathbb{T}_u + (1 - \phi_1) \mathbb{T}_{1m}, \quad (4.1)$$

where $\mathbb{T}_u = \hat{y}_u$ and $\mathbb{T}_{1m} = \hat{y}_m + k(\hat{x}_n - \hat{x}_m)$; $\phi_1 \in [0, 1]$ and k is a scalar quantity to be chosen suitably.

5. IST successive regression estimator

The regression estimator is another estimator in survey sampling theory. Hence, the estimator for the matched portion of the sample have been chosen as a regression type estimator given by $\mathbb{T}_{2m}$. The final IST successive regression estimator to estimate the sensitive population mean at current occasion is given as

$$\mathbb{T}_2 = \phi_2 \mathbb{T}_u + (1 - \phi_2) \mathbb{T}_{2m}, \quad (5.1)$$

where $\mathbb{T}_{2m} = \hat{y}_m + \hat{b}(m_{ll})(\hat{x}_n - \hat{x}_m)$ with $\hat{b}(m_{ll}) = \{s_{z_1 z_2}(m_{ll})\} / \{s_{z_2}^2(m_{ll})\}$ and $\phi_2 \in [0, 1]$ is a scalar quantity to be chosen suitably.

6. IST successive general class of estimator

Many estimators such as ratio, product, exponential ratio, may be thought on similar lines for proposing an estimator based on matched sample of size m . Therefore, in order to generalize the frame work, an IST general class of estimator has been proposed, so that the IST difference, IST regression and others may be viewed as members of the proposed class of estimator. Hence, the final estimator in this case is given as

$$\mathbb{T}_3 = \phi_3 \mathbb{T}_u + (1 - \phi_3) \mathbb{T}_{3m}, \quad (6.1)$$

where $\mathbb{T}_{3m} = g(\hat{y}_m, \hat{x}_m, \hat{x}_n)$ is a function of $\hat{y}_m$, $\hat{x}_m$, and $\hat{x}_n$. Following Srivastava and Jhaji (1980), Tracy *et al.* (1996), and Priyanka and Trisandhya (2018), the function g is assumed such that it satisfies following conditions:

- (i) The point $(\hat{y}_m, \hat{x}_m, \hat{x}_n)$ assumes the value in a closed convex subset $\mathbb{R}^3$ of three dimensional real space containing the point $(\bar{Y}, \bar{X}, \bar{X})$.
- (ii) The function $g(\hat{y}_m, \hat{x}_m, \hat{x}_n)$ is continuous and bounded in $\mathbb{R}^3$.
- (iii) $g(\bar{Y}, \bar{X}, \bar{X}) = \bar{Y}$ and $g_1(\bar{Y}, \bar{X}, \bar{X}) = \partial g(\hat{y}_m, \hat{x}_m, \hat{x}_n) / \partial \hat{y}_m = 1$, i.e., first order partial derivative of g with respect to $\hat{y}_m$ at $g(\bar{Y}, \bar{X}, \bar{X}) = \bar{Y} \Rightarrow g_1(K) = \partial g(\cdot) / \partial \hat{y}_m|_K = 1$, where $K = (\bar{Y}, \bar{X}, \bar{X})$.
- (iv) The first and second order partial derivatives of $g(\hat{y}_m, \hat{x}_m, \hat{x}_n)$ exist and are continuous and bounded in $\mathbb{R}^3$.

7. Analysis of IST estimators on successive occasions

To elucidate the performances of proposed IST estimators, the bias, variance/mean squared error of the proposed estimators $\mathbb{T}_i$ ($i = 1, 2, 3$) has been calculated as

$$\begin{aligned} \mathbb{B}(\mathbb{T}_i) &= E(\mathbb{T}_i - \bar{Y}) \\ &= E[\phi_i(\mathbb{T}_u - \bar{Y}) + (1 - \phi_i)(\mathbb{T}_{im} - \bar{Y})] \\ &= \phi_i \mathbb{B}[\mathbb{T}_u] + (1 - \phi_i) \mathbb{B}[\mathbb{T}_{im}], \quad i = 1, 2, 3. \end{aligned}$$

Since, $\mathbb{T}_u$ is unbiased for $\bar{Y}$, so $\mathbb{B}(\mathbb{T}_u) = 0$. Therefore, the bias of estimator $\mathbb{T}_i$ is given as

$$\mathbb{B}(\mathbb{T}_i) = (1 - \phi_i)\mathbb{B}(\mathbb{T}_{im}), \quad i = 1, 2, 3. \quad (7.1)$$

The variance of the estimator $\mathbb{T}_i$ is computed as

$$\begin{aligned} \mathbb{V}(\mathbb{T}_i) &= E(\mathbb{T}_i - \bar{Y})^2 \\ &= E[\phi_i(\mathbb{T}_u - \bar{Y}) + (1 - \phi_i)(\mathbb{T}_{im} - \bar{Y})]^2 \\ &= \phi_i^2 \mathbb{V}(\mathbb{T}_u) + (1 - \phi_i)^2 \mathbb{V}(\mathbb{T}_{im}) + 2\phi_i(1 - \phi_i)\text{cov}(\mathbb{T}_u, \mathbb{T}_{im}), \quad i = 1, 2, 3. \end{aligned} \quad (7.2)$$

As $\mathbb{T}_u$ and $\mathbb{T}_{im}$ are based on two independent samples of sizes u and m respectively. So, $\text{cov}(\mathbb{T}_u, \mathbb{T}_{im}) = 0$. Therefore, the variance of estimator $\mathbb{T}_i$ becomes

$$\mathbb{V}(\mathbb{T}_i) = \phi_i^2 \mathbb{V}(\mathbb{T}_u) + (1 - \phi_i)^2 \mathbb{V}(\mathbb{T}_{im}). \quad (7.3)$$

It can be seen that, $\mathbb{V}(\mathbb{T}_i)$ in equation (7.3) is a function of ϕ_i . So, it has been optimized with respect to ϕ_i and optimum value of ϕ_i is obtained as:

$$\phi_{i\text{opt.}} = \frac{\mathbb{V}(\mathbb{T}_{im})}{\mathbb{V}(\mathbb{T}_u) + \mathbb{V}(\mathbb{T}_{im})}, \quad i = 1, 2, 3. \quad (7.4)$$

Substituting the optimum value of ϕ_i from equation (7.4) in equation (7.3) the optimum variance of the proposed estimator $\mathbb{T}_i$ is computed as

$$\mathbb{V}(\mathbb{T}_i)_{\text{opt.}} = \frac{\mathbb{V}(\mathbb{T}_u) \times \mathbb{V}(\mathbb{T}_{im})}{\mathbb{V}(\mathbb{T}_u) + \mathbb{V}(\mathbb{T}_{im})}, \quad i = 1, 2, 3. \quad (7.5)$$

7.1. Bias and variance of $\mathbb{T}_u$ and $\mathbb{T}_{im}$; $i = 1, 2, 3$

The estimator $\mathbb{T}_u$ is unbiased for $\bar{Y}$, hence, its variance is computed as

$$\mathbb{V}(\mathbb{T}_u) = \left(\frac{S_{z_2}^2}{u_{ll}} \right) + \left(\frac{S_{t_2}^2}{u_{sl}} \right) - \left(\frac{S_{z_2}^2 + S_{t_2}^2}{N} \right). \quad (7.6)$$

The estimator $\mathbb{T}_{1m}$ is also unbiased for $\bar{Y}$, so its variance is obtained as

$$\begin{aligned} \mathbb{V}(\mathbb{T}_{1m}) &= \left(\frac{1}{m_{ll}} \right) \left[k^2 S_{z_1}^2 - 2k\rho_{z_1 z_2} S_{z_1} S_{z_2} + S_{z_2}^2 \right] + \left(\frac{1}{m_{sl}} \right) \left[k^2 S_{t_1}^2 - 2kS_{t_1} S_{t_2} \rho_{t_1 t_2} + S_{t_2}^2 \right] \\ &\quad + \left(\frac{1}{n_{ll}} \right) \left[2k\rho_{z_1 z_2} S_{z_1} S_{z_2} - k^2 S_{z_1}^2 \right] + \left(\frac{1}{n_{sl}} \right) \left[2kS_{t_1} S_{t_2} \rho_{t_1 t_2} - k^2 S_{t_1}^2 \right] - \left(\frac{1}{N} \right) \left[S_{z_2}^2 + S_{t_2}^2 \right]. \end{aligned} \quad (7.7)$$

The above expression for variance of $\mathbb{T}_{1m}$ contain an unknown constant k , hence, it is optimized with respect to k and optimum value of k is obtained as

$$k_{\text{opt.}} = \frac{\left(\frac{1}{m_{ll}} - \frac{1}{n_{ll}} \right) S_{z_1} S_{z_2} \rho_{z_1 z_2} + \left(\frac{1}{m_{sl}} - \frac{1}{n_{sl}} \right) S_{t_1} S_{t_2} \rho_{t_1 t_2}}{\left(\frac{1}{m_{ll}} - \frac{1}{n_{ll}} \right) S_{z_1}^2 + \left(\frac{1}{m_{sl}} - \frac{1}{n_{sl}} \right) S_{t_1}^2}.$$

Now, as the estimator $\mathbb{T}_{2m}$ and $\mathbb{T}_{3m}$ are biased for $\bar{Y}$, hence the expression for their bias and mean squared error have been computed under the following transformations:

$$\begin{aligned}\bar{z}_{2ull} &= \bar{Z}_2(1 + e_0), & \bar{z}_{2mll} &= \bar{Z}_2(1 + e_1), & \bar{z}_{1mll} &= \bar{Z}_1(1 + e_2), & \bar{z}_{1nll} &= \bar{Z}_1(1 + e_3), \\ \bar{t}_{2ust} &= \bar{T}_2(1 + e_4), & \bar{t}_{2mst} &= \bar{T}_2(1 + e_5), & \bar{t}_{1mst} &= \bar{T}_1(1 + e_6), & \bar{t}_{1nst} &= \bar{T}_1(1 + e_7), \\ s_{z_2}^2(m_{ll}) &= S_{z_2}^2(1 + e_8), & s_{z_1 z_2}(m_{ll}) &= S_{z_1 z_2}(1 + e_9), \\ n_{rs} &= \frac{v_{rs}}{(v_{20})^{\frac{r}{2}}(v_{02})^{\frac{s}{2}}}, & v_{rs} &= \frac{1}{n-1} \sum (z_{1i} - \bar{z}_1)^r (z_{2i} - \bar{z}_2)^s\end{aligned}$$

such that, $E(e_i) = 0$; $|e_i| < 1$, where $i = 0, 1, 2, 3, 4, 5, 6, 7, 8$, and 9 .

Under the above transformations, retaining the terms up to first order of approximation, we have for bias and mean squared error of $\mathbb{T}_{2m}$ as

$$\mathbb{B}(\mathbb{T}_{2m}) = \left(\frac{1}{m_{ll}} - \frac{1}{n_{ll}} \right) \beta_{z_1 z_2} \left[S_{z_1} n_{03} - S_{z_1} \frac{n_{12}}{\rho_{z_1 z_2}} \right] \quad (7.8)$$

and

$$\begin{aligned}\mathbb{V}(\mathbb{T}_{2m}) &= \left(\frac{1}{m_{ll}} \right) \left[S_{z_1}^2 \beta_{z_1 z_2}^2 - 2 S_{z_1} S_{z_2} \beta_{z_1 z_2} \rho_{z_1 z_2} + S_{z_2}^2 \right] + \left(\frac{1}{m_{sl}} \right) \left[S_{t_2}^2 + \beta_{z_1 z_2}^2 S_{t_1}^2 - 2 \beta_{z_1 z_2} S_{t_1} S_{t_2} \rho_{t_1 t_2} \right] \\ &+ \left(\frac{1}{n_{ll}} \right) \left[2 S_{z_1} S_{z_2} \beta_{z_1 z_2} \rho_{z_1 z_2} - S_{z_1}^2 \beta_{z_1 z_2}^2 \right] + \left(\frac{1}{n_{sl}} \right) \left[2 S_{t_1} S_{t_2} \rho_{t_1 t_2} \beta_{z_1 z_2} - S_{t_1}^2 \beta_{z_1 z_2}^2 \right] - \left(\frac{1}{N} \right) \left[S_{z_2}^2 + S_{t_2}^2 \right]. \quad (7.9)\end{aligned}$$

Also, the bias and mean squared error of the estimator $\mathbb{T}_{3m}$ has been derived under above considered transformations as:

$$\mathbb{T}_{3m} = g(\hat{y}_m, \hat{x}_m, \hat{x}_n).$$

Expanding $g(\hat{y}_m, \hat{x}_m, \hat{x}_n)$ about the point $K = (\bar{Y}, \bar{X}, \bar{X})$ using Taylor series expansion, retaining terms up to first order of approximations, we have

$$\begin{aligned}\mathbb{T}_{3m} &= g \left[\bar{Y} + (\hat{y}_m - \bar{Y}), \bar{X} + (\hat{x}_m - \bar{X}), \bar{X} + (\hat{x}_n - \bar{X}) \right] \\ &= (\hat{y}_m + (\hat{x}_m - \bar{X})) G_2 + (\hat{x}_n - \bar{X}) G_3 + \left[(\hat{y}_m - \bar{Y})^2 G_{11} + (\hat{x}_m - \bar{X})^2 G_{22} + (\hat{x}_n - \bar{X})^2 G_{33} \right. \\ &\quad \left. + (\hat{y}_m - \bar{Y})(\hat{x}_m - \bar{X}) G_{12} + (\hat{y}_m - \bar{Y})(\hat{x}_n - \bar{X}) G_{13} + (\hat{x}_m - \bar{X})(\hat{x}_n - \bar{X}) G_{23} + \dots \right], \quad (7.10)\end{aligned}$$

where,

$$\begin{aligned}G_1 &= \left. \frac{\partial g}{\partial \hat{y}_m} \right|_K = 1, & G_2 &= \left. \frac{\partial g}{\partial \hat{x}_m} \right|_K, & G_3 &= \left. \frac{\partial g}{\partial \hat{x}_n} \right|_K, & G_{11} &= \left. \frac{1}{2} \frac{\partial^2 g}{\partial \hat{y}_m^2} \right|_K = 0, & G_{22} &= \left. \frac{1}{2} \frac{\partial^2 g}{\partial \hat{x}_m^2} \right|_K, \\ G_{33} &= \left. \frac{1}{2} \frac{\partial^2 g}{\partial \hat{x}_n^2} \right|_K, & G_{12} &= \left. \frac{1}{2} \frac{\partial^2 g}{\partial \hat{y}_m \partial \hat{x}_m} \right|_K, & G_{13} &= \left. \frac{1}{2} \frac{\partial^2 g}{\partial \hat{y}_m \partial \hat{x}_n} \right|_K, & G_{23} &= \left. \frac{1}{2} \frac{\partial^2 g}{\partial \hat{x}_m \partial \hat{x}_n} \right|_K.\end{aligned}$$

Bias and Mean squared error of the class of estimator $\mathbb{T}_{3m}$ to the first order approximations are ob-

tained as

$$\begin{aligned}\mathbb{B}(\mathbb{T}_{3m}) = & \frac{1}{m_{ll}} \left[S_{z_1}^2 G_{22} + \rho_{z_1 z_2} S_{z_1} S_{z_2} G_{12} \right] + \frac{1}{m_{sl}} \left[S_{t_1}^2 G_{22} + \rho_{t_1 t_2} S_{t_1} S_{t_2} G_{12} \right] \\ & + \frac{1}{n_{ll}} \left[S_{z_1}^2 G_{33} + \rho_{z_1 z_2} S_{z_1} S_{z_2} G_{13} + S_{z_1}^2 G_{23} \right] + \frac{1}{n_{sl}} \left[S_{t_1}^2 G_{33} + \rho_{t_1 t_2} S_{t_1} S_{t_2} G_{13} + S_{t_1}^2 G_{23} \right] \\ & - \frac{1}{N} \left[S_{t_1}^2 (G_{22} + G_{33} + G_{23}) + S_{t_1} S_{t_2} \rho_{t_1 t_2} (G_{12} + G_{13}) + S_{z_1}^2 (G_{22} + G_{33} + G_{23}) \right. \\ & \left. + S_{z_1} S_{z_2} \rho_{z_1 z_2} (G_{12} + G_{13}) \right]\end{aligned}\quad (7.11)$$

and

$$\begin{aligned}\mathbb{V}(\mathbb{T}_{3m}) = & \left(\frac{1}{m_{ll}} \right) \left[S_{z_1}^2 G_2^2 + 2\rho_{z_1 z_2} S_{z_1} S_{z_2} G_2 + S_{z_2}^2 \right] + \left(\frac{1}{m_{sl}} \right) \left[S_{t_1}^2 G_2^2 + 2S_{t_1} S_{t_2} \rho_{t_1 t_2} G_2 + S_{t_2}^2 \right] \\ & + \left(\frac{1}{n_{ll}} \right) \left[S_{z_1}^2 G_3^2 + 2\rho_{z_1 z_2} S_{z_1} S_{z_2} G_3 + 2G_2 G_3 S_{z_1}^2 \right] + \left(\frac{1}{n_{sl}} \right) \left[S_{t_1}^2 G_3^2 + 2G_3 S_{t_1} S_{t_2} \rho_{t_1 t_2} + 2G_2 G_3 S_{t_1}^2 \right] \\ & - \left(\frac{1}{N} \right) \left[S_{z_1}^2 G_2^2 + 2\rho_{z_1 z_2} S_{z_1} S_{z_2} G_2 + S_{t_1}^2 G_2^2 + 2S_{t_1} S_{t_2} \rho_{t_1 t_2} G_2 + S_{z_2}^2 + S_{t_2}^2 + S_{z_1}^2 G_3^2 \right. \\ & \left. + 2\rho_{z_1 z_2} S_{z_1} S_{z_2} G_3 + 2G_2 G_3 S_{z_1}^2 + S_{t_1}^2 G_3^2 + 2S_{t_1} S_{t_2} \rho_{t_1 t_2} G_3 + 2G_2 G_3 S_{t_1}^2 \right].\end{aligned}\quad (7.12)$$

Clearly, we can see that equation (7.12) is a function of G_2 and G_3 . So, after minimizing equation 15 by partially differentiating with respect to G_2 and G_3 respectively and equating to zero we get the optimized value of G_2 and G_3 as

$$\begin{aligned}G_{2\text{opt.}} = & - \frac{\left[\left(\frac{1}{m_{ll}} - \frac{1}{n_{ll}} \right) S_{z_1} S_{z_2} \rho_{z_1 z_2} + \left(\frac{1}{m_{sl}} - \frac{1}{n_{sl}} \right) S_{t_1} S_{t_2} \rho_{t_1 t_2} \right]}{\left(\frac{1}{m_{ll}} - \frac{1}{n_{ll}} \right) S_{z_1}^2 + \left(\frac{1}{m_{sl}} - \frac{1}{n_{sl}} \right) S_{t_1}^2}, \\ G_{3\text{opt.}} = & - \frac{\left[\left(\frac{1}{n_{ll}} - \frac{1}{N} \right) \left(S_{z_1} S_{z_2} \rho_{z_1 z_2} + S_{z_1}^2 G_2 \right) + \left(\frac{1}{n_{sl}} - \frac{1}{N} \right) \left(S_{t_1} S_{t_2} \rho_{t_1 t_2} + S_{t_1}^2 G_2 \right) \right]}{\left(\frac{1}{n_{ll}} - \frac{1}{N} \right) S_{z_1}^2 + \left(\frac{1}{n_{sl}} - \frac{1}{N} \right) S_{t_1}^2}.\end{aligned}$$

7.2. Allocating LL & SL Sample using Trappmann *et al.* (2014) approach and Perri *et al.* (2018) approach

Since, in IST a sample is split in to LL sample and SL sample. Trappmann *et al.* (2014) considered equal number of units in both the samples irrespective of variability of the items in the two lists. Applying his approach on successive occasions we have the following allocations:

$$n_{ll} = n_{sl} = \frac{n}{2}, \quad m_{ll} = m_{sl} = \frac{m}{2}, \quad \text{and} \quad u_{ll} = u_{sl} = \frac{u}{2}.$$

However, Perri *et al.* (2018) concluded that the estimates may be affected due to high variability of items in LL sample and SL sample. Hence, they proposed optimal sample size allocation to LL and SL samples by minimizing the variance of IST estimates under a budget constraints. Hence, modifying these ideas to work for allocating LL sample and SL sample on various samples at first and second

Table 2: Minimum variance

i	Est.	Minimum Variance under Trappmann <i>et al.</i> (2014) approach	
1	$\mathbb{T}_1$	$\mathbb{V}_t(\mathbb{T}_1)_{\min} = \left[\frac{\hat{\mu}_{1t}^2 J_{11} - \hat{\mu}_{1t} J_{12} + J_{13}}{\hat{\mu}_{1t}^2 J_{14} + \hat{\mu}_{1t} J_{15} + K_{11}} \right]$	with, $\hat{\mu}_{1t} = \min \left\{ \frac{-I_{12} + \sqrt{I_{12}^2 + I_{11} I_{13}}}{I_{11}}, \frac{-I_{12} - \sqrt{I_{12}^2 + I_{11} I_{13}}}{I_{11}} \right\} \in [0, 1]$
2	$\mathbb{T}_2$	$\mathbb{V}_t(\mathbb{T}_2)_{\min} = \left[\frac{\hat{\mu}_{2t}^2 J_{31} - \hat{\mu}_{2t} J_{32} + J_{33}}{\hat{\mu}_{2t}^2 J_{34} + \hat{\mu}_{2t} J_{35} + K_{31}} \right]$	with, $\hat{\mu}_{2t} = \min \left\{ \frac{-I_{32} + \sqrt{I_{32}^2 + I_{31} I_{33}}}{I_{31}}, \frac{-I_{32} - \sqrt{I_{32}^2 + I_{31} I_{33}}}{I_{31}} \right\} \in [0, 1]$
3	$\mathbb{T}_3$	$\mathbb{V}_t(\mathbb{T}_3)_{\min} = \left[\frac{\hat{\mu}_{3t}^2 J_{51} - \hat{\mu}_{3t} J_{52} + J_{53}}{\hat{\mu}_{3t}^2 J_{54} + \hat{\mu}_{3t} J_{55} + K_{51}} \right]$	with, $\hat{\mu}_{3t} = \min \left\{ \frac{-I_{52} + \sqrt{I_{52}^2 + I_{51} I_{53}}}{I_{51}}, \frac{-I_{52} - \sqrt{I_{52}^2 + I_{51} I_{53}}}{I_{51}} \right\} \in [0, 1]$
i	Est.	Minimum Variance under Perri <i>et al.</i> (2018) approach	
1	$\mathbb{T}_1$	$\mathbb{V}_p(\mathbb{T}_1)_{\min} = \left[\frac{\hat{\mu}_{1p}^2 J_{21} - \hat{\mu}_{1p} J_{22} + J_{23}}{\hat{\mu}_{1p}^2 J_{24} + \hat{\mu}_{1p} J_{25} + K_{21}} \right]$	with, $\hat{\mu}_{1p} = \min \left\{ \frac{-I_{22} + \sqrt{I_{22}^2 + I_{21} I_{23}}}{I_{21}}, \frac{-I_{22} - \sqrt{I_{22}^2 + I_{21} I_{23}}}{I_{21}} \right\} \in [0, 1]$
2	$\mathbb{T}_2$	$\mathbb{V}_p(\mathbb{T}_2)_{\min} = \left[\frac{\hat{\mu}_{2p}^2 J_{41} - \hat{\mu}_{2p} J_{42} + J_{43}}{\hat{\mu}_{2p}^2 J_{44} + \hat{\mu}_{2p} J_{45} + K_{41}} \right]$	with, $\hat{\mu}_{2p} = \min \left\{ \frac{-I_{42} + \sqrt{I_{42}^2 + I_{41} I_{43}}}{I_{41}}, \frac{-I_{42} - \sqrt{I_{42}^2 + I_{41} I_{43}}}{I_{41}} \right\} \in [0, 1]$
3	$\mathbb{T}_3$	$\mathbb{V}_p(\mathbb{T}_3)_{\min} = \left[\frac{\hat{\mu}_{3p}^2 J_{61} - \hat{\mu}_{3p} J_{62} + J_{63}}{\hat{\mu}_{3p}^2 J_{64} + \hat{\mu}_{3p} J_{65} + K_{61}} \right]$	with, $\hat{\mu}_{3p} = \min \left\{ \frac{-I_{62} + \sqrt{I_{62}^2 + I_{61} I_{63}}}{I_{61}}, \frac{-I_{62} - \sqrt{I_{62}^2 + I_{61} I_{63}}}{I_{61}} \right\} \in [0, 1]$

occasion assuming same budget allocation for each LL and SL samples we have:

$$n_{ll} = \frac{nS_{z_1}}{S_{z_1} + S_{t_1}} = n\beta_1 \text{ (say)}, \quad n_{sl} = \frac{nS_{t_1}}{S_{z_1} + S_{t_1}} = n\beta_2 \text{ (say)}, \quad u_{ll} = \frac{uS_{z_2}}{S_{z_2} + S_{t_2}} = u\beta_3 \text{ (say)},$$

$$u_{sl} = \frac{uS_{t_2}}{S_{z_2} + S_{t_2}} = u\beta_4 \text{ (say)}, \quad m_{ll} = \frac{mS_{z_1}}{S_{z_1} + S_{t_1}} = m\beta_1 \text{ (say)}, \quad \text{and} \quad m_{sl} = \frac{mS_{t_1}}{S_{z_1} + S_{t_1}} = m\beta_2 \text{ (say)}.$$

Using the assumptions of both the approaches the minimum variance of proposed estimators $\mathbb{T}_i$ has been obtained and are presented in Table 2. Where,

$$J_{j1} = K_{j0}K_{j3}f - K_{j0}K_{j4}f^2, \quad J_{j2} = K_{j1}K_{j3} - f(K_{j0}K_{j4}f - K_{j0}K_{j2} - K_{j0}K_{j3} + K_{j1}K_{j4}),$$

$$J_{j3} = K_{j1}K_{j2} + K_{j1}K_{j3} - fK_{j1}K_{j4}, \quad J_{j4} = f(K_{j0} + K_{j4}) - K_{j3}, \quad J_{j5} = K_{j2} + K_{j3} - K_{j1} - f(K_{j0} + K_{j4}),$$

$$I_{j1} = J_{j1}J_{j5} + J_{j2}J_{j4}, \quad I_{j2} = J_{j1}K_{j1} - J_{j4}J_{j3}, \quad I_{j3} = J_{j2}K_{j1} + J_{j3}J_{j5},$$

$$K_{j0} = S_{z_2}^2 + S_{t_2}^2; \quad \forall j = 1, 2, \dots, 6; \quad K_{j1} = 2S_{z_2}^2 + 2S_{t_2}^2; \quad \forall j = 1, 3, 5; \quad K_{j1} = \frac{S_{z_2}^2}{\beta_3} + \frac{S_{t_2}^2}{\beta_4}; \quad \forall j = 2, 4, 6;$$

$$K_{12} = 2(k_t^2 S_{z_1}^2 - 2k_t S_{z_1} S_{z_2} \rho_{z_1 z_2} + S_{z_2}^2 + k_t^2 S_{t_1}^2 - 2k_t S_{t_1} S_{t_2} \rho_{t_1 t_2} + S_{t_2}^2),$$

$$K_{13} = 2(2k_t S_{z_1} S_{z_2} \rho_{z_1 z_2} - k_t^2 S_{z_1}^2 + 2k_t S_{t_1} S_{t_2} \rho_{t_1 t_2} - k_t^2 S_{t_1}^2), \quad K_{14} = S_{z_2}^2 + S_{t_2}^2,$$

$$K_{22} = \frac{k_p^2 S_{z_1}^2 - 2k_p S_{z_1} S_{z_2} \rho_{z_1 z_2} + S_{z_2}^2}{\beta_1} + \frac{k_p^2 S_{t_1}^2 - 2k_p S_{t_1} S_{t_2} \rho_{t_1 t_2} + S_{t_2}^2}{\beta_2},$$

$$K_{23} = \frac{2k_p S_{z_1} S_{z_2} \rho_{z_1 z_2} - k_p^2 S_{z_1}^2}{\beta_1} + \frac{2k_p S_{t_1} S_{t_2} \rho_{t_1 t_2} - k_p^2 S_{t_1}^2}{\beta_2}, \quad K_{24} = S_{z_2}^2 + S_{t_2}^2,$$

$$k_t = \frac{S_{z_1} S_{z_2} \rho_{z_1 z_2} + S_{t_1} S_{t_2} \rho_{t_1 t_2}}{S_{z_1}^2 + S_{t_1}^2}, \quad k_p = \left(\frac{S_{z_1} S_{z_2} \rho_{z_1 z_2}}{\beta_1} + \frac{S_{t_1} S_{t_2} \rho_{t_1 t_2}}{\beta_2} \right) / \left(\frac{S_{z_1}^2}{\beta_1} + \frac{S_{t_1}^2}{\beta_2} \right),$$

$$K_{32} = 2 \left(\frac{S_{z_1 z_2}^2}{S_{z_2}^4} S_{z_1}^2 - 2 \frac{S_{z_1 z_2}}{S_{z_2}^2} S_{z_1} S_{z_2} \rho_{z_1 z_2} + \frac{S_{z_1}^2}{S_{z_2}^4} S_{t_1}^2 - 2 \frac{S_{z_1 z_2}}{S_{z_2}^2} S_{t_1} S_{t_2} \rho_{t_1 t_2} + S_{z_2}^2 + S_{t_2}^2 \right),$$

$$\begin{aligned}
K_{33} &= 2 \left(2S_{z_1} S_{z_2} \rho_{z_1 z_2} \frac{S_{z_1 z_2}}{S_{z_2}^2} - S_{z_1}^2 \frac{S_{z_1 z_2}^2}{S_{z_2}^4} + 2S_{t_1} S_{t_2} \rho_{t_1 t_2} \frac{S_{z_1 z_2}}{S_{z_2}^2} - S_{t_1}^2 \frac{S_{z_1 z_2}^2}{S_{z_2}^4} \right), \quad K_{34} = S_{z_2}^2 + S_{t_2}^2, \\
K_{42} &= \left(S_{z_1}^2 \frac{S_{z_1 z_2}^2}{S_{z_2}^4} - 2S_{z_1} S_{z_2} \rho_{z_1 z_2} \frac{S_{z_1 z_2}}{S_{z_2}^2} + S_{z_2}^2 \right) \frac{1}{\beta_1} + \left(S_{t_1}^2 \frac{S_{z_1 z_2}^2}{S_{z_2}^4} - 2 \frac{S_{z_1 z_2}}{S_{z_2}^2} S_{t_1} S_{t_2} \rho_{t_1 t_2} + S_{t_2}^2 \right) \frac{1}{\beta_2}, \\
K_{43} &= \left(2S_{z_1} S_{z_2} \rho_{z_1 z_2} \frac{S_{z_1 z_2}}{S_{z_2}^2} - S_{z_1}^2 \frac{S_{z_1 z_2}^2}{S_{z_2}^4} \right) \frac{1}{\beta_1} + \left(2S_{t_1} S_{t_2} \rho_{t_1 t_2} \frac{S_{z_1 z_2}}{S_{z_2}^2} - S_{t_1}^2 \frac{S_{z_1 z_2}^2}{S_{z_2}^4} \right) \frac{1}{\beta_2}, \\
K_{44} &= S_{z_2}^2 + S_{t_2}^2, \quad K_{52} = 2 \left(S_{z_1}^2 G_{2t}^2 + 2\rho_{z_1 z_2} S_{z_1} S_{z_2} G_{2t} + S_{t_1}^2 G_{2t}^2 + 2S_{t_1} S_{t_2} \rho_{t_1 t_2} G_{2t} + S_{z_2}^2 + S_{t_2}^2 \right), \\
K_{53} &= 2 \left[S_{z_1}^2 G_{3t}^2 + 2\rho_{z_1 z_2} S_{z_1} S_{z_2} G_{3t} + 2G_{2t} G_{3t} S_{z_1}^2 + S_{t_1}^2 G_{3t}^2 + 2G_{3t} S_{t_1} S_{t_2} \rho_{t_1 t_2} + 2G_{2t} G_{3t} S_{t_1}^2 \right], \\
K_{54} &= \left[S_{z_1}^2 G_{2t}^2 + 2\rho_{z_1 z_2} S_{z_1} S_{z_2} G_{2t} + S_{t_1}^2 G_{2t}^2 + 2S_{t_1} S_{t_2} \rho_{t_1 t_2} G_{2t} + S_{z_2}^2 + S_{t_2}^2 + S_{z_1}^2 G_{3t}^2 + 2\rho_{z_1 z_2} S_{z_1} S_{z_2} G_{3t} \right. \\
&\quad \left. + 2G_{2t} G_{3t} S_{z_1}^2 + S_{t_1}^2 G_{3t}^2 + 2G_{3t} S_{t_1} S_{t_2} \rho_{t_1 t_2} + 2G_{2t} G_{3t} S_{t_1}^2 \right], \\
K_{62} &= \frac{1}{\beta_1} \left(S_{z_1}^2 G_{2p}^2 + 2\rho_{z_1 z_2} S_{z_1} S_{z_2} G_{2p} + S_{z_2}^2 \right) + \frac{1}{\beta_2} \left(S_{t_1}^2 G_{2p}^2 + 2S_{t_1} S_{t_2} \rho_{t_1 t_2} G_{2p} + S_{t_2}^2 \right), \\
K_{63} &= \frac{1}{\beta_1} \left(S_{z_1}^2 G_{3p}^2 + 2\rho_{z_1 z_2} S_{z_1} S_{z_2} G_{3p} + 2G_{2p} G_{3p} S_{z_1}^2 \right) + \frac{1}{\beta_2} \left(S_{t_1}^2 G_{3p}^2 + 2G_{3p} S_{t_1} S_{t_2} \rho_{t_1 t_2} + 2G_{2p} G_{3p} S_{t_1}^2 \right), \\
K_{64} &= \left[S_{z_1}^2 G_{2p}^2 + 2\rho_{z_1 z_2} S_{z_1} S_{z_2} G_{2p} + S_{t_1}^2 G_{2p}^2 + 2S_{t_1} S_{t_2} \rho_{t_1 t_2} G_{2p} + S_{z_2}^2 + S_{t_2}^2 + S_{z_1}^2 G_{3p}^2 + 2\rho_{z_1 z_2} S_{z_1} S_{z_2} G_{3p} \right. \\
&\quad \left. + 2G_{2p} G_{3p} S_{z_1}^2 + S_{t_1}^2 G_{3p}^2 + 2G_{3p} S_{t_1} S_{t_2} \rho_{t_1 t_2} + 2G_{2p} G_{3p} S_{t_1}^2 \right], \\
G_{2t} &= -\frac{S_{z_1} S_{z_2} \rho_{z_1 z_2} + S_{t_1} S_{t_2} \rho_{t_1 t_2}}{S_{z_1}^2 + S_{t_1}^2}, \quad G_{3t} = -\frac{S_{z_1} S_{z_2} \rho_{z_1 z_2} + S_{z_1}^2 G_{2t} + S_{t_1} S_{t_2} \rho_{t_1 t_2} + S_{t_1}^2 G_{2t}}{S_{z_1}^2 + S_{t_1}^2}, \\
G_{2p} &= -\left(\frac{S_{z_1} S_{z_2} \rho_{z_1 z_2}}{\beta_1} + \frac{S_{t_1} S_{t_2} \rho_{t_1 t_2}}{\beta_2} \right) / \left(\frac{S_{z_1}^2}{\beta_1} + \frac{S_{t_1}^2}{\beta_2} \right), \\
G_{3p} &= -\left(\frac{S_{z_1} S_{z_2} \rho_{z_1 z_2} + S_{z_1}^2 G_{2p}}{\beta_1} + \frac{S_{t_1} S_{t_2} \rho_{t_1 t_2} + S_{t_1}^2 G_{2p}}{\beta_2} \right) / \left(\frac{S_{z_1}^2}{\beta_1} + \frac{S_{t_1}^2}{\beta_2} \right), \quad \text{and} \quad f = \frac{n}{N}.
\end{aligned}$$

8. Efficiency comparison

In order to compare various proposed IST estimators in successive sampling, the percent relative efficiencies have been computed for data considered in section 9 under Trappmann *et al.* (2014) as well as Perri *et al.* (2018) allocation designs as:

$$E_1 = \frac{\mathbb{V}_t(\mathbb{T}_1)_{\min}}{\mathbb{V}_p(\mathbb{T}_1)_{\min}} \times 100, \quad E_2 = \frac{\mathbb{V}_t(\mathbb{T}_2)_{\min}}{\mathbb{V}_p(\mathbb{T}_2)_{\min}} \times 100, \quad \text{and} \quad E_3 = \frac{\mathbb{V}_t(\mathbb{T}_3)_{\min}}{\mathbb{V}_p(\mathbb{T}_3)_{\min}} \times 100.$$

9. Numerical demonstration

Population Source: [Free access to data by Statistical Abstracts of United States]

To evaluate the performance of proposed IST successive sampling estimators, numerical illustrations has been supplemented using natural population. The population consists of $N = 51$ states. Let the aim be to estimate rate of abortion in year 2004. Therefore, for IST successive sampling frame work

Table 3: Empirical results

$\hat{\mu}_{1p}$	$\hat{\mu}_{1t}$	$\hat{\mu}_{2p}$	$\hat{\mu}_{2t}$	$\hat{\mu}_{3p}$	$\hat{\mu}_{3t}$	E_1	E_2	E_3
0.8424	0.8359	0.7617	0.7626	*	*	114.9334	115.8558	-

*' = indicates that the optimum value of fraction of sample to be drawn afresh do not exist; '-' = denote corresponding percent relative efficiency cannot be computed.

we consider:

y = Rate of abortions in the year 2004

x = Rate of abortions in the year 2000

t_1 = Number of residents in the year 2000

t_2 = Number of residents in the year 2004

Clearly the rate of abortion is sensitive; however, rate of residents is non-sensitive. Hence, the data is suitable to be applied for IST frame work. Since same study variable "rate of abortion" has been observed for two different years 2000 and 2004, therefore, the considered data is suitable for IST successive sampling frame work. The numerical calculations have been performed on the data with results represented in Table 3.

The value of E_i ($i = 1, 2$) are observed to be more than 100, this indicates that optimum allocation design by Perri *et al.* (2018) is preferable over Trappmann *et al.* (2014) design. Therefore, the further numerical analysis has been done using Perri *et al.* (2018) allocation design.

10. Simulation study

An extensive simulation study has been done using Monte Carlo simulation for the data mentioned in Section 9. The 5,000 different Monte Carlo replications have been observed. The process is also repeated for a different combination of constants termed as sets. The variance/mean squared error of the proposed estimators $\mathbb{T}_1$, $\mathbb{T}_2$, and $\mathbb{T}_3$ has been computed under Perri *et al.* (2018); in addition, the allocation design and are denoted by $\mathbb{V}_p(\mathbb{T}_1)$, $\mathbb{V}_p(\mathbb{T}_2)$, and $\mathbb{V}_p(\mathbb{T}_3)$ respectively. The percent of relative efficiencies for the IST successive difference and regression estimators with respect to IST successive general class of estimator have been computed as:

$$E_{s1} = \frac{\mathbb{V}_p(\mathbb{T}_1)}{\mathbb{V}_p(\mathbb{T}_3)} \times 100 \quad \text{and} \quad E_{s2} = \frac{\mathbb{V}_p(\mathbb{T}_2)}{\mathbb{V}_p(\mathbb{T}_3)} \times 100.$$

The procedure have been repeated for three different combinations of constants, termed as different sets given as:

I : $n = 24$, $u = 04$, $m = 20$

II : $n = 24$, $u = 08$, $m = 16$

III : $n = 24$, $u = 10$, $m = 14$.

Figure 1 and Figure 2 summarizes the outcomes of the simulation results.

11. Direct method

Estimators that use IST are less efficient than estimators obtained using direct questioning. Hence, in order to identify the amount of loss we compare the proposed class of estimator $\mathbb{T}_3$ with respect to

Figure 1: Simulated percent relative efficiency of the proposed IST general class of estimator with respect to proposed IST difference estimator for three different sets.

Figure 2: Simulated percent relative efficiency of the proposed IST general class of estimator with respect to proposed IST regression estimator for three different sets.

corresponding direct method. The estimator under direct questioning method is proposed as

$$\mathbb{T}_D = \chi \mathbb{T}_{uD} + (1 - \chi) \mathbb{T}_{3mD}, \quad \chi \in [0, 1], \quad (11.1)$$

where

$$\mathbb{T}_{uD} = \bar{y}_u, \quad (11.2)$$

$$\mathbb{T}_{3mD} = d(\bar{y}_m, \bar{x}_m, \bar{x}_n), \quad (11.3)$$

where $\mathbb{T}_{3mD}$ follow similar regularity conditions as stated in Section 6. The, minimum mean squared error of the class of estimator $\mathbb{T}_D$ is obtained as

$$\mathbb{V}(\mathbb{T}_D)_{\min} = \left[\frac{-\hat{\mu}_D^2 J_{d1} + \hat{\mu}_D J_{d2} + J_{d3}}{\hat{\mu}_D^2 J_{d4} - \hat{\mu}_D J_{d5} + K_{d1}} \right] \quad (11.4)$$

with

$$\hat{\mu}_D = \min \left\{ \frac{I_{d2} + \sqrt{I_{d2}^2 - I_{d1}I_{d3}}}{I_{d1}}, \frac{I_{d2} - \sqrt{I_{d2}^2 - I_{d1}I_{d3}}}{I_{d1}} \right\} \quad \text{such that } \hat{\mu}_D \in [0, 1], \quad (11.5)$$

where,

$$\begin{aligned} J_{d1} &= K_{d1}K_{d3}f^2, & J_{d2} &= f(K_{d1}K_{d3} - K_{d1}K_{d2} + fK_{d1}K_{d3}), & J_{d3} &= K_{d1}K_{d2} - fK_{d1}K_{d3}, \\ J_{d4} &= f(K_{d1} + K_{d3}), & J_{d5} &= K_{d1} - K_{d2} + f(K_{d1} + K_{d3}), & I_{d1} &= J_{d1}J_{d5} - J_{d2}J_{d4}, \\ I_{d2} &= J_{d3}J_{d4} + J_{d1}K_{d1}, & I_{d3} &= J_{d2}K_{d1} + J_{d3}J_{d5}, & K_{d1} &= S_y^2, & K_{d2} &= S_y^2 + S_x S_y D_2^2 + 2S_y^2 D_2 \rho_{yx}, \\ K_{d3} &= S_y^2 + S_x^2 D_2^2 + 2S_y S_x \rho_{yx} D_2, & \text{and } D_2 &= \frac{-S_y S_x \rho_{yx}}{S_x^2}. \end{aligned}$$

Further the simulated ratio of the mean squared error of $\mathbb{T}_D$ and $\mathbb{T}_3$ have been computed by considering 5,000 different samples using Monte Carlo simulation study for different sets and results are presented in Figure 3 and Figure 4 respectively

$$\text{Ratio} = \frac{\mathbb{V}(\mathbb{T}_D)}{\mathbb{V}_p(\mathbb{T}_3)}. \quad (11.6)$$

Remark 1. The IST is used with an expectation to receive a true response as in IST directly; however, the response to sensitive questions are not being asked. Standard bias reduction methods like Jackknife, the Bootstrap and methods that use approximations of the bias function through asymptotic expansions of the bias that follow Kosmidis (2014) may be used if some false responses are still received that show bias in the estimator after explaining the IST efficiently to respondents.

12. Interpretations of results

The following interpretations can be drawn from empirical and simulation results:

1. It has been observed that IST is feasible in successive sampling to handle sensitive issues on successive occasions.
2. From Table 3, it is clear that E_1 and E_2 both are coming out to be greater than 100, this implies that optimum allocation design by Perri *et al.* (2018) is more efficient than allocation by Trappmann *et al.* (2014) design in two occasion successive sampling. It is to be noted that for the considered data the optimum fraction to be drawn afresh do not exist for IST successive general class of estimator, so corresponding efficiency cannot be computed. Hence, in order to check the validity of IST successive general class of estimator simulation has been carried out with several choices of parameters.

Figure 3: Ratio of mean squared error of T_3 (under optimum allocation design) with respect to direct method under IST in two occasion successive sampling for Set-I.

3. Simulation results in Figure 1 and Figure 2, justify that E_{s_1} and E_{s_2} are greater than 100 for all three considered sets. This indicates that IST successive general class of estimators is more efficient than IST successive regression and IST, a successive difference estimators. However, $E_{s_1} > E_{s_2}$ indicate that the IST successive difference estimator is better than the IST successive regression estimator. Also, the simulated percent relative efficiency increases as ϕ increases and is in accordance with the theory of successive sampling.
4. From Figure 3 and Figure 4, it is observed that as ϕ increases the simulated ratio of the mean squared error of the direct method and IST successive general class of estimator increases. The values indicate a loss in precision of the IST general class of estimator over the direct method. The issues under consideration are sensitive; therefore, a direct method is unsuitable because the privacy of respondents need to be considered; therefore, despite of loss in precision IST need to be preferred over direct method for sensitive issues on successive occasions.

13. Conclusion

The IST on successive occasions enables an estimation of the population mean of stigmatized quantitative variable using innocuous information, that reduce the social desirability response bias and provides privacy. Out of the three proposed IST successive estimators, the IST successive general class of estimator under both Trappmann *et al.* (2014) allocation designs as well as Perri *et al.* (2018) allocation designs have been proven to be more efficient than the other two. The optimum allocation of LL and SL samples by Perri *et al.* (2018) have been found to be more fruitful in successive sampling than allocation due to Trappmann *et al.* (2014). While comparing with direct method, certain

Figure 4: Ratio of mean squared error of T_3 (under optimum allocation design) with respect to direct method under IST in two occasion successive sampling for Set-II.

amount of loss in precision is observed but that is realistic as the survey issues are sensitive so there may be chances of complete refusal or partial refusal if we apply direct method. However, using IST on successive occasion atleast estimation of sensitive issues are possible. Therefore it can be concluded that the proposed IST successive estimators provide comfort and satisfaction to the respondents in terms of privacy protection as well as a methodological advancement in literature related to successive sampling dealing with sensitive issues. Hence, the IST successive class of estimators may be recommended for practical use by survey practitioners.

Acknowledgements

The authors are thankful to the reviewers and editors for their valuable suggestions that improved an earlier version of the paper. The authors are also thankful to SERB, New Delhi, India for providing the financial assistance to carry out the present work. The authors also sincerely acknowledge free access to data from the Statistical Abstracts of United States.

References

- Arnab R and Singh S (2013). Estimation of mean of sensitive characteristics for successive sampling, *Communications in Statistics - Theory and Methods*, **42**, 2499–2524.
- Chaudhuri A and Christofides TC (2013). *Indirect Questioning in Sample Surveys*, Springer-Verlag, Berlin, Heidelberg, De.
- Hussian Z, Shabbir N, and Shabbir J (2015). An alternative item sum technique for improved estimators of population mean in sensitive surveys, *Hacettepe University Bulletin of Natural Sciences*

- and *Engineering Series B: Mathematics and Statistics*, **46**, 1–30.
- Jessen RJ (1942). Statistical investigation of a sample survey for obtaining farm facts, *Iowa Agriculture and Home Economics Experiment Station. Research Bulletin*, **26**, 1–104.
- Kosmidis I (2014). Bias in parametric estimation: reduction and useful side-effects, *WIREs Computational Statistics*, **6**, 185–196.
- Miller JD (1984). *A new survey technique for studying deviant behavior* (Ph.D. thesis), The George Washington University, Washington DC.
- Naeem N and Shabbir J (2016). Use of scrambled responses on two occasions successive sampling under non-response, *Hacettepe University Bulletin of Natural Sciences and Engineering Series B: Mathematics and Statistics*, **46**. Available from: <http://www.hjms.hacettepe.edu.tr/uploads/32c18b80-275f-4b5a-a28d-6a5890eecac3.pdf>
- Perri PF, Rueda Garcia M, and Cobo Rodriguez B (2018). Multiple sensitive estimation and optimal sample size allocation in the item sum technique, *Biometrical Journal*, **60**, 155–173.
- Priyanka K and Trisandhya P (2018). A composite class of estimators using scrambled response mechanism for sensitive population mean in successive sampling, *Communications in Statistics - Theory and Methods*, DOI: 10.1080/03610926.2017.1422762
- Priyanka K, Trisandhya P, and Mittal R (2018). Dealing sensitive characters on successive occasions through a general class of estimators using scrambled response techniques, *Metron*, **76**, 203–230.
- Rueda Garcia M, Perri PF, and Cobo Rodriguez B (2017). Advances in estimation by the item sum technique using Auxiliary information in complex surveys, *Advances in Statistical Analysis*, **102**, 455–478.
- Singh GN, Suman S, Khetan M, and Paul C (2017). Some estimation procedures of sensitive character using scrambled response techniques in successive sampling, *Communications in Statistics - Theory and Methods*, DOI:10.1080/03610926.2017.1327073.
- Srivastava SK and Jhaji HS (1980). A class of estimators using auxiliary information for estimating finite population variance, *Sankhya*, **C42**, 87–96.
- Tracy DS, Singh HP, and Singh R (1996). An alternative to the ratio-cum-product estimator in sample surveys, *Journal of Statistical Planning and Inference*, **53**, 375–397.
- Trappmann M, Krumpal I, Kirchner A, and Jann B (2014). Item sum: a new technique for asking quantitative sensitive questions, *Journal of Survey Statistics and Methodology*, **2**, 58–77.
- Tian GL and Tang ML (2014). *Incomplete Categorical Data Design: Non-Randomized Response Techniques for Sensitive Questions in Surveys*, Chapman & Hall/CRC, FL.
- Warner SL (1965). Randomized response: a survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association*, **60**, 63–69.
- Yu B, Jin Z, Tian J, and Gao G (2015). Estimation of sensitive proportion by randomized response data in successive sampling, *Computational and Mathematical Methods in Medicine*, **2015**, 1–6.