Journal of Modern Mathematics and Statistics 13 (1): 1-14, 2019
ISSN: 1994-5388

Improved Item Sum Technique for Estimating Population Mean with Sensitive Variables

[1]Owoidighe E. Idim, [1]Ekaette I. Enang and [2]Damilola B. Adekanmbi
[1]Department of Statistics, University of Calabar, Calabar, Nigeria
[2]Department of Statistics, Ladoke Akintola University of Technology, Ogbomosho, Nigeria

Abstract: Getting correct answers to sensitive questions from respondents and estimating the population parameters on variables that are sensitive in nature is still a problem in survey sampling. This study proposes a new sampling design to estimate the population mean of sensitive variable. It compares analytically and numerically the variance of the proposed estimator to some existing estimators and establish its greater efficiency. The process was extended to Searl's method of estimation, estimation method which utilizes priori information and estimation process using auxiliary information. Comparisons of their variances/mean square errors give valid results that agree with established facts in the literature

Key words: Sensitive variable, Iitem Sum Technique (IST), auxiliary information, efficiency, method, comparisons

INTRODUCTION

The challenge of getting truthful or correct answers to sensitive questions from respondents in most surveys has remained with researchers over the years. This is because humans naturally tend to hide, distort, vague, underreport or even refuse to respond to questions bearing sensitive or stigmatizing characteristics like tax evasion, cheating, buying of stolen properties and such other attitudes that potentially violate social norms. The reason for the unfortunate choice of such answers stems from the fact that these violations are often formally or informally sanctioned. In most cases, respondents tend to systematically under report norm violation and over report norm-conforming activities. These situationscan introduce considerable bias to the estimation of the unknown population parameters and lower the overall data quality of survey studies. The main challenge, therefore, is how to get truthful answers to questions bearing sensitive or stigmatizing characteristics from respondents.

In an attempt to contend with this situation, survey designers have over years developed various data collection strategies in a bid to elicit more honest answers from respondents. Warner (1965) introduced the Randomized Response Technique (RRT) with the main aim of estimating the true proportion of sensitive characteristics in the population while protecting the privacy of respondents. This technique was found to reduce the evasive answering bias and increases response rate. Horvitz *et al*. (1967), Kuk (1990), Kim and Warde (2004), Gjestvang and Singh (2006) and many other research suggested different modifications and conducted theoretical investigations tothe properties of the (Warner, 1965) RRT. Although, RRT has received a considerable wide attention in many research areas like physical and social sciences because of its advantages, there are several difficulties and limitations associated with RRTs. Surveys conducted through RRT require much time and cost to complete. Geurts (1980) considered the financial limitation of the RRT and also reported that larger sample sizes are needed to construct the confidence interval as compared to the direct questioning method. Hubbard *et al*. (1989) pointed out that making a decision on the choice of the randomization device best for obtaining information on sensitive or stigmatizing characteristic also posed a major challenge. Chaudhuri and Christofides (2007) criticized that RRT is confined with respondent's skill to understand and handle the device as an ingenious respondent may understand that his/her response can be traced back to his/her real status provided he/she understands the mathematical logic behind the randomization device. These limitations and difficulties made researchers to introduce alternatives techniques in the literature. These alternative techniques include the Unmatched Count Technique (UCT) (Smith *et al*., 1974), the nominative technique (Miller, 1985), the Three Card method (Droitcour *et al*., 2001), Item Count Technique (ICT) (Holbrook and Krosnick, 2009), Item Sum Technique (IST)

Corresponding Author: Owoidighe E. Idim, Department of Statistics, University of Calabar, Calabar, Nigeria

(Trappmann *et al*., 2013), one sample version item sum technique (Hussain *et al*., 2015) among others. This resarech proposes an alternative sampling strategy for estimating the population mean of variables with sensitive nature.

Item Sum Technique (IST) (Trappmann *et al*., 2013): Trappmann *et al*. (2013) proposed a quantified version of Item Count Technique (ICT) and named it item sum technique. Here the survey respondents are randomly divided into two subsamples. Each member of the first subsample is presented with a list containing g+1 items with g of those items related to non-sensitive characteristics (T_i) and one related to Sensitive characteristic (S) while each of the participants in the second subsample is presented with all the g non-sensitive items. All sensitive and non-sensitive items are quantitative in nature. Respondents in both subsamples are then asked to report the total score applicable to them without reporting the individual scores on each of the items. An unbiased estimator of population mean of sensitive item, say μ_s from the IST data can be estimated by the mean difference between the two subsample sand is given by:

$$\hat{\mu}_{s1} = \bar{y}_1 - \bar{y}_2 \tag{1}$$

where, $\bar{y}_1$ and $\bar{y}_2$ are the sample means of the first and second sub-samples, respectively. The variance of $\hat{\mu}_{s1}$ with variance:

$$\text{Var}(\hat{\mu}_{s1}) = \frac{\sigma_s^2}{n_1} + \frac{n\sum_{i=1}^{g}\sigma_{ti}^2}{n_1 n_2} \tag{2}$$

And:

$$\text{Var}(\hat{\mu}_{s1}) = \frac{1-f_1}{n_1}\sigma_s^2 + \frac{n-n_1f_2-n_2f_1}{n_1n_2}\sum_{g}^{g}\sigma_{ti}^2 \tag{3}$$

If sampling is done with and without replacement, respectively.

Item sum technique (Hussain *et al*., 2015): Hussain *et al*. (2015) proposed an IST without two subsamples. Suppose μ_s is the population mean of the sensitive variable of interest. μ_s is estimated by the method as follows: A simple random sample of size n is selected from the population. Each of the participants in the sample is provided with a list of g items. The ith item is an addition of queries about a stigmatizing Sensitive (S) and non-stigmatizing (T_i) variables. The respondents are directed to report only the total score of all items. Both the non-stigmatizing and the stigmatizing (sensitive) variables are quantitative in nature.

It is assumed that all (T_i) and (S) variables are unrelated to each other and the distribution of non-sensitive (T_i) variables are known to the interviewer. They proposed an unbiased estimator of the form:

$$\hat{\mu}_{s2} = \frac{\bar{y}-\sum_{i=1}^{g}\mu_{ti}}{g} \tag{4}$$

Where:

μ_{ti} = The population mean of the ith non-sensitive variable
I = 1, 2, ..., g
$\bar{y}$ = The sample mean of reported response

The variance of the estimator is given by:

$$\text{Var}(\hat{\mu}_{s2}) = \frac{\sigma_s^2}{n} + \frac{\sum_{i=1}^{g}\sigma_{ti}^2}{ng^2} \tag{5}$$

And when sample is drawn without replacement, the variance of $\hat{\mu}_{s2}$ is given by:

$$\text{Var}(\hat{\mu}_{s2}) = \frac{1-f}{ng^2}\left(g^2\sigma_s^2 + \sum_{i=1}^{g}\sigma_{ti}^2\right) \tag{6}$$

Where:

σ_s^2 = The population variance of sensitive variable
σ_{ti}^2 = The population variance of non sensitive variables
f = n/N

MATERIALS AND METHODS

The proposed sampling strategy: After considering the models proposed by Trappmann *et al.* (2013) and Hussain *et al*. (2015), we propose a new sampling design and an improved estimator obtained from a linear combination of the two existing estimators. The procedure is as described as.

Suppose, we have U = {U_1, U_2, ..., U_N} to be a finite human population of size N. Let a random sample of size n be drawn from the population with or without replacement. Initially, each respondent in the sample is served with a list containing g items. The jth item consist of questions about one Sensitive (S) variable and a non sensitive (T_i) variable. Both the sensitive and non sensitive variables are quantitative in nature and the respondents are directed to report only the total score of all items. Both the sensitive and non sensitive variables are unrelated to each other and the distribution of the non sensitive variables is assumed known to the interviewer. For instance, the respondents in the sample may be asked questions like:

- Last digit of your cell phone number+number of times you smoked shisha last month
- Date on your birth day was+number of times you smoked shisha last month
- Last digit of your CNIC (Computerized National Identity Card) number+number of times you smoked shisha last month
- Number of hours you watched TV last day+Number of times you smoked shisha last month

After this, the sample is then randomly divided into two subsamples of sizes n_1 and n_2. Members of the first subsample are served with a list containing g+1 items with of those items related to non sensitive T_i attributes and one Sensitive (S) attribute while members of the second subsample are served with a list containing the g non sensitive (T_i) variables only. All sensitive and non sensitive items are quantitative in nature and the respondents in the two subsamples are asked to report only the total score applicable to them without reporting each individual scores. Using the case of harmful type of tobacco (shisha) for instance, respondents in the first subsample may be asked the following questions:

- Last digit of your cell phone number is?
- Date on your birth day was?
- Last digit of your CNIC (Computerized National Identity Card) number is?
- Number of hours you watched TV last day?
- Number of times you smoked shisha last month?

While the respondents in the second subsample are asked every other question except the last one. The proposed estimators is given by:

$$\bar{y}_{sp} = \alpha_1\hat{\mu}_{s1} + \alpha_2\hat{\mu}_{s2} \tag{7}$$

Or more generally as:

$$\bar{y}_{sp} = \alpha_1(\bar{y}_1 - \bar{y}_2) + \alpha_2\left(\frac{\bar{y} - \sum_{i=1}^{g}\mu_{ti}}{g}\right) \tag{8}$$

Where:

$\alpha_1 + \alpha_2$ = 1, $\bar{y}_1$

$\bar{y}_2$ = As defined in Eq. 1 $\bar{y}$

μ_{ti} = As defined in Eq. 4

i = 1, 2, ..., g

The proposed estimator, $\bar{y}_{sp}$ is an unbiased estimator of the population mean of sensitive variable μ_s with optimum variance given as:

$$V_{opt}(\bar{y}_{sp}) = \frac{\sigma_s^2}{n} + \frac{\sum_{i=1}^{g}\sigma_{ti}^2}{ng^2} - \frac{\left(\frac{\sigma_s^2}{n} + \frac{\sum_{i=1}^{g}\sigma_{ti}^2}{ng^2}\right)^2}{\frac{n+n_1}{nn_1}\sigma_s^2 + \frac{n^2g^2+n_1n_2}{nn_1n_2g^2}\sum_{i=1}^{g}\sigma_{ti}^2} \tag{9}$$

When sample is drawn without replacement, the variance of $\bar{y}_{sp}$ is given by:

$$V_{opt}(\bar{y}_{sp}) = \frac{1-f}{ng^2}\left(g^2\sigma_s^2 + \sum_{i=1}^{g}\sigma_{ti}^2\right) - \frac{\left\{\frac{1-f}{ng^2}\left(g^2\sigma_s^2 + \sum_{i=1}^{g}\sigma_{ti}^2\right)\right\}^2}{\frac{n+n_1-nf_1-n_1f_2}{nn_1}\sigma_s^2 + \frac{n^2g^2+n_1n_2-n_1n_2f-nn_1g^2f_2-nn_2g^2f_1}{nn_1n_2g^2}\sum_{i=1}^{g}\sigma_{ti}^2} \tag{10}$$

Or:

$$V_{opt}(\bar{y}_{sp}) = Var(\hat{\mu}_{s2}) - \frac{[Var(\hat{\mu}_{s2})]}{Var(\hat{\mu}_{s1}) + Var(\hat{\mu}_{s2})}$$

Where:

σ_s^2 = The population variance of the sensitive variable

σ_{ti}^2 = The population variance of the non sensitive variable

$$f = \frac{n}{N}, f_1 = \frac{n_1}{N} \text{ and } f_2 = \frac{n_2}{N}$$

Efficiency comparison: Here we present the efficiency comparison of the proposed estimator, $\bar{y}_{sp}$ with other estimators. The proposed estimator will be more efficient than other existing estimators if the following condition holds:

$$Var(\hat{\mu}_{si}) - Var(\bar{y}_{sp}) > 0, \quad i = 1, 2 \tag{11}$$

Trappmann *et al.* (2013) estimator, $\hat{\mu}_{s1}$ with the proposed estimator $\bar{y}_{sp}$:

Consider; $Var(\hat{\mu}_{s1}) - Var(\bar{y}_{sp}) \geq 0$

$$\frac{\sigma_s^2}{n_1} + \frac{n\sum_{i=1}^{g}\sigma_{ti}^2}{n_1n_2} - \frac{\sigma_s^2}{n} - \frac{\sum_{i=1}^{g}\sigma_{ti}^2}{ng^2} + \frac{\left(\frac{\sigma_s^2}{n} + \frac{\sum_{i=1}^{g}\sigma_{ti}^2}{ng^2}\right)^2}{\frac{(n+n_1)}{nn_1}\sigma_s^2 + \frac{(n^2g^2+n_1n_2)}{nn_1n_2g^2}\sum_{i=1}^{g}\sigma_{ti}^2} \geq 0$$

$$\frac{n_2}{nn_1}\sigma_s^2+\frac{\left(n^2g^2-n_1n_2\right)}{nn_1n_2g^2}\sum_{i=1}^{g}\sigma_{ti}^2+\frac{\left(\frac{\sigma_s^2}{n}+\frac{\sum_{i=1}^{g}\sigma_{ti}^2}{ng^2}\right)^2}{\frac{(n+n_1)}{nn_1}\sigma_s^2+\frac{\left(n^2g^2+n_1n_2\right)}{nn_1n_2g^2}\sum_{i=1}^{g}\sigma_{ti}^2}\geq 0 \quad (12)$$

Since, all the terms in the right hand side of Eq. 11 are all positive, then theinequality always holds and we can infer that the proposed estimator, $\bar{y}_{sp}$ is more efficient than $\hat{\mu}_{s1}$.

When sampling without replacement: We have:

$$\text{Var}(\hat{\mu}_{s1})\text{-Var}(\hat{\mu}_{s2})+\frac{\left(\text{Var}(\hat{\mu}_{s2})\right)^2}{\text{Var}(\hat{\mu}_{s1})+\text{Var}(\hat{\mu}_{s2})}\geq 0 \quad (13)$$

And taking a close look at the difference, we have $\text{Var}(\hat{\mu}_{s1})\text{-Var}(\hat{\mu}_{s2})$:

$$\frac{\sigma_s^2}{n_1}+\left(\frac{n^2}{n_1n_2}+\frac{1}{Ng^2}-\frac{1}{ng^2}-\frac{2}{N}\right)\sum_{i=1}^{g}\sigma_{ti}^2\geq 0$$

Which holds since, N is large and $g\geq 2$. This implies that the inequality in Eq. 13 holds and we can therefore, infer that the proposed estimator, $\bar{y}_{sp}$ is more efficient than $\hat{\mu}_{s1}$ when sampling is done without replacement.

Hussain *et al.* (2015) estimator, $\hat{\mu}_{s2}$ with the proposed estimator, $\bar{y}_{sp}$:

Consider; $\text{Var}(\hat{\mu}_{s2})\text{-Var}(\bar{y}_{sp})\geq 0$

$$\frac{\sigma_s^2}{n}+\frac{\sum_{i=1}^{g}\sigma_{ti}^2}{ng^2}-\frac{\sigma_s^2}{n}-\frac{\sum_{i=1}^{g}\sigma_{ti}^2}{ng^2}+\frac{\left(\frac{\sigma_s^2}{n}+\frac{\sum_{i=1}^{g}\sigma_{ti}^2}{ng^2}\right)^2}{\frac{(n+n_1)}{nn_1}\sigma_s^2+\frac{(n^2g^2+n_1n_2)}{nn_1n_2g^2}\sum_{i=1}^{g}\sigma_{ti}^2}\geq 0$$

$$\Rightarrow\frac{\left(\frac{\sigma_s^2}{n}+\frac{\sum_{i=1}^{g}\sigma_{ti}^2}{ng^2}\right)^2}{\frac{(n+n_1)}{nn_1}\sigma_s^2+\frac{\left(n^2g^2+n_1n_2\right)}{nn_1n_2g^2}\sum_{i=1}^{g}\sigma_{ti}^2}\geq 0 \quad (14)$$

Since, the term in the RHS of the above inequality is positive, the inequality always holds. We therefore, infer that the proposed estimator, $\bar{y}_{sp}$ is more efficient than $\hat{\mu}_{s1}$.

When sampling without replacement: We have:

$$\text{Var}(\hat{\mu}_{s2})\text{-}\left\{\text{Var}(\hat{\mu}_{s2})\text{-}\frac{\left(\text{Var}(\hat{\mu}_{s2})\right)^2}{\text{Var}(\hat{\mu}_{s1})+\text{Var}(\hat{\mu}_{s2})}\right\}\geq 0$$
$$\Rightarrow\frac{\left(\text{Var}(\hat{\mu}_{s2})\right)^2}{\text{Var}(\hat{\mu}_{s1})+\text{Var}(\hat{\mu}_{s2})}\geq 0 \quad (15)$$

Equation 15 always holds, since, variance is a positive quantity and as such we can also infer that the proposed estimator $\bar{y}_{sp}$ is more efficient than $\hat{\mu}_{s2}$ when sampling is done without replacement.

Percentage relative efficiency: The Percentage Relative Efficiency (PRE) of the proposed estimator, $\bar{y}_{sp}$ with respect to other estimators is defined as:

$$\text{PRE}(\bar{y}_{sp},\hat{\mu}_{si})=\frac{\text{Var}(\hat{\mu}_{si})}{\text{Var}(\bar{y}_{sp})}*100 \quad (16)$$
$$i=1,2$$

Using the definition in Eq. 16 above and with the data provided by Hussain *et al* (2015) PRE of the proposed estimator, $\bar{y}_{sp}$ with respect to Trappmann *et al.* (2013); Hussain *et al.* (2015) estimators represented by $\hat{\mu}_{s1}$ and $\hat{\mu}_{s2}$, respectively have been obtained for the different fixed values of the parameters involved and the results are presented in Table 1-9. From the results, it can clearly be seen that the proposed estimator $\bar{y}_{sp}$ is always more efficient than the estimators $\hat{\mu}_{s1}$ and $\hat{\mu}_{s2}$.

Some alternative classes of estimators for μ_s

Searl's method of estimation: By application of Searl's (1964) method of estimation, we define a family of estimators for the population mean of sensitive variable μ_s as follows:

$$\bar{y}_{s\lambda}=\lambda\bar{y}_{sp} \quad (17)$$

Where:

λ = A constant to be suitably chosen by the interviewer or researcher

$\bar{y}_{sp}$ = The proposed estimator defined in Eq. 8 above

The bias of $\bar{y}_{s\lambda}$ is given by:

$$\text{Bias}(\bar{y}_{s\lambda})=(\lambda\text{-}1)\mu_s \quad (18)$$

And the mean square error, MSE is given by:

$$\text{MSE}(y_{s\lambda})=\lambda^2V_{opt}(\bar{y}_{sp})+(\lambda\text{-}1)^2\mu_s^2 \quad (19)$$

Efficiency comparison: The proposed estimator $\bar{y}_{s\lambda}$is relatively more efficient than $\bar{y}_{sp}$ if:

Table 1: PRE of $\bar{y}_{sp}$ with respect to $\hat{\mu}_{s1}$ and $\hat{\mu}_{s2}$ for n = 50, $n_1 = n_2 = 25$ and different valuesn of g, σ_s^2 with $\sum_{i=1}^{g}\sigma_{ti}^2 = 0.1$

	$\sum_{i=1}^{g}\sigma_{ti}^2 = 0.1$							
	g = 2		g = 3		g = 5		g = 10	
σ_s^2	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$
0.25	427.2727	130.5556	444.6809	129.0123	454.3307	128.2222	458.5657	127.8889
0.5	366.6667	137.5000	373.9130	136.5079	377.7778	136.0000	379.4411	135.7857
1	334.1463	142.7083	337.3626	142.1296	339.0438	141.8333	339.7602	141.7083
2	317.2840	146.0227	318.7845	145.7071	319.5609	145.5455	319.8901	145.4773
3	311.5702	147.2656	312.5461	147.0486	313.0493	146.9375	313.2622	146.8906
4	308.6757	147.9167	309.4183	147.7513	309.7902	147.6667	309.9475	147.6310
5	306.9652	148.3173	307.5388	148.1838	307.8337	148.1154	307.9584	148.0865
10	303.4913	149.1422	303.7736	149.0741	303.9184	149.0392	303.9796	149.0245
20	301.7478	149.5668	301.8878	149.5325	301.9596	149.5149	301.9899	149.5074
50	300.6997	149.8257	300.7554	149.8119	300.7839	149.8048	300.7960	149.8018
100	300.3499	149.9127	300.3777	149.9057	300.3920	149.9022	300.3980	149.9007
200	300.1750	149.9563	300.1889	149.9528	300.1960	149.9510	300.1990	149.9503

Table 2: PRE of $\bar{y}_{sp}$ with respect to $\hat{\mu}_{s1}$ and $\hat{\mu}_{s2}$ for n = 50, $n_1 = n_2 = 25$ and different valuesn of g, σ_s^2 with $\sum_{i=1}^{g}\sigma_{ti}^2 = 1$

	$\sum_{i=1}^{g}\sigma_{ti}^2 = 1$							
	g = 2		g = 3		g = 5		g = 10	
σ_s^2	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$
0.25	1000.0000	111.1111	1346.1540	108.0247	1651.7240	106.4444	1830.7690	105.0247
0.5	766.6667	115.0000	918.1818	112.2222	1025.9260	110.8000	1080.3920	110.2000
1	580.0000	120.8333	640.0000	118.5185	676.9231	117.3333	694.0594	116.8333
2	455.5556	128.1250	478.9474	126.3889	492.1569	125.5000	498.0100	125.1250
3	407.6923	132.5000	421.4286	131.1111	428.9474	130.4000	432.2259	130.1000
4	382.3529	135.4167	391.8919	134.2593	397.0297	133.6667	399.2519	133.4167
5	366.6667	137.5000	373.9130	136.5079	377.7778	136.0000	379.4411	135.7857
10	334.1463	142.7083	337.3626	142.1296	339.0438	141.8333	339.7602	141.7083
20	317.2840	146.0227	318.7845	145.7071	319.5609	145.5455	319.8901	145.4773
50	306.9652	148.3173	307.5388	148.1838	307.8337	148.1154	307.9584	148.0865
100	303.4913	149.1422	303.7736	149.9007	303.9184	149.0392	303.9796	149.0245
200	301.7478	149.5668	301.8878	149.9503	301.9596	149.5149	301.9899	149.5074

Table 3: PRE of $\bar{y}_{sp}$ with respect to $\hat{\mu}_{s1}$ and $\hat{\mu}_{s2}$ for n = 50, $n_1 = n_2 = 25$ and different valuesn of g, σ_s^2 with $\sum_{i=1}^{g}\sigma_{ti}^2 = 100$

	$\sum_{i=1}^{g}\sigma_{ti}^2 = 100$							
	g = 2		g = 3		g = 5		g = 10	
σ_s^2	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$
0.25	1686.138	106.3046	3625.184	102.8367	9523.529	101.0612	32140.00	100.3121
0.5	1672.549	106.3591	3553.589	102.8955	9011.111	101.1222	26833.33	100.3741
1	1646.154	106.4677	3419.266	103.0127	8140.000	101.2438	20200.00	100.4975
2	1596.296	106.6832	3181.356	103.2453	6833.333	101.4851	13566.67	100.7426
3	1550.000	106.8966	2977.165	103.4756	5900.000	101.7241	10250.00	100.9852
4	1506.897	107.1078	2800.000	103.7037	5200.000	101.9608	8260.000	101.2255
5	1466.667	107.3171	2644.828	103.9395	4655.556	102.1951	6933.333	101.4634
10	1300.000	108.3333	2089.474	105.0265	3100.000	103.3333	3918.1818	102.6190
20	1077.778	110.2273	1514.286	107.0707	1933.333	105.4545	2195.2381	104.7727
50	766.6667	115.0000	918.1818	112.2222	1025.926	110.8000	1080.3922	110.2000
100	580.0000	120.8333	640.0000	118.5185	676.9231	117.3333	694.0594	116.8333
200	455.5556	128.1250	478.9474	126.3889	492.1569	125.5000	498.0100	125.1250

$$\text{MSE}(\bar{y}_{s\lambda})\text{-}V_{opt}(\bar{y}_{sp}) \leq 0$$

$$(\lambda^2\text{-}1)V_{opt}(\bar{y}_{sp}) + (\lambda\text{-}1)^2\mu_s^2 \leq 0$$

It can be shown that $\text{MSE}(\bar{y}_{s\lambda})\text{-}V_{opt}(\bar{y}_{sp}) \leq 0$ if and only if:

$$\frac{\mu_s^2\text{-}V_{opt}(\bar{y}_{sp})}{\mu_s^2 + V_{opt}(\bar{y}_{sp})} < \lambda \leq 1 \qquad (20)$$

Table 4: PRE of $\bar{y}_{sp}$ with respect to $\hat{\mu}_{s1}$ and $\hat{\mu}_{s2}$ for n = 50, n_1 = 20, n_2 = 30 and different valuesn of g, σ_s^2 with $\sum_{i=1}^{g}\sigma_{ti}^2 = 0.1$

	$\sum_{i=1}^{g}\sigma_{ti}^2 = 0.1$							
	g = 2		g = 3		g = 5		g = 10	
σ_s^2	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$
0.25	478.7879	126.4000	498.9362	125.0667	510.1050	124.3840	515.0066	124.0960
0.5	417.4603	131.5000	426.0870	130.6667	430.6878	130.2400	432.6680	130.0600
1	384.5528	135.1429	388.4615	134.6667	390.5046	134.4229	391.3753	134.3200
2	367.4897	137.3846	369.3370	137.1282	370.2927	136.9969	370.6980	136.9415
3	361.7080	138.2105	362.9151	138.0351	363.5375	137.9453	363.8010	137.9074
4	358.7992	138.6400	359.6953	138.5067	360.1565	138.4384	360.3516	138.4096
5	357.0481	138.9032	357.7605	138.7957	358.1268	138.7406	358.2817	138.7174
10	353.5328	139.4426	353.8846	139.3880	354.0650	139.3600	354.1413	139.3482
20	351.7686	139.7190	351.9434	139.6915	352.0329	139.6774	352.0707	139.6714
50	350.7080	139.8870	350.7776	139.8760	350.8133	139.8703	350.8283	139.8679
100	350.3541	139.9434	350.3888	139.9379	350.4067	139.9350	350.4142	139.9338
200	350.1771	139.9717	350.1944	139.9689	350.2033	139.9675	350.2071	139.9669

Table 5: PRE of $\bar{y}_{sp}$ with respect to $\hat{\mu}_{s1}$ and $\hat{\mu}_{s2}$ for n = 50, n_1 = 20, n_2 = 30 and different valuesn of g, σ_s^2 with $\sum_{i=1}^{g}\sigma_{ti}^2 = 1$

	$\sum_{i=1}^{g}\sigma_{ti}^2 = 1$							
	g = 2		g = 3		g = 5		g = 10	
σ_s^2	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$
0.25	1058.333	110.4348	1426.9230	107.5362	1752.2990	106.0522	1942.949	105.4261
0.5	822.2222	113.8462	986.3636	111.2821	1103.0860	109.9692	1162.092	109.4154
1	633.3333	118.7500	700.0000	116.6667	741.0256	115.6000	760.0660	115.1500
2	507.4074	124.5455	534.2105	123.0303	549.3464	122.2545	556.0531	121.9273
3	458.9744	127.8571	475.0000	126.6667	483.7719	126.0571	487.5969	125.8000
4	433.3333	130.0000	444.5946	129.0196	450.6601	128.5176	453.2835	128.3058
5	417.4603	131.5000	426.0870	130.6667	430.6878	130.2400	432.6680	130.0600
10	384.5528	135.1429	388.4615	134.6667	390.5046	134.4229	391.3753	134.3200
20	367.4897	137.3846	369.3370	137.1282	370.2927	136.9969	370.6980	136.9415
50	357.0481	138.9032	357.7605	138.7957	358.1268	138.7406	358.2817	138.7174
100	353.5328	139.4426	353.8846	139.3880	354.0650	139.3600	354.1413	139.3482
200	351.7686	139.7190	351.9434	139.6915	352.0329	139.6774	352.0707	139.6714

Table 6: PRE of $\bar{y}_{sp}$ with respect to $\hat{\mu}_{s1}$ and $\hat{\mu}_{s2}$ for n = 50, n_1 = 20, n_2 = 30 and different valuesn of g, σ_s^2 with $\sum_{i=1}^{g}\sigma_{ti}^2 = 100$

	$\sum_{i=1}^{g}\sigma_{ti}^2 = 100$							
	g = 2		g = 3		g = 5		g = 10	
σ_s^2	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$
0.25	1752.640	106.0509	3772.982	102.7226	9918.628	101.0185	33483.33	100.2996
0.5	1738.888	106.1017	3699.282	102.7783	9387.037	101.0768	27961.11	100.3589
1	1712.179	106.2028	3561.009	102.8893	8483.333	101.1928	21058.33	100.4771
2	1661.728	106.4032	3316.101	103.1094	7127.778	101.4229	14155.55	100.7115
3	1614.881	106.6012	3105.905	103.3268	6159.524	101.6503	10704.16	100.9430
4	1571.264	106.7969	2923.529	103.5417	5433.333	101.8750	8633.333	101.1719
5	1530.555	106.9903	2763.793	103.7540	4868.519	102.0971	7252.777	101.3981
10	1361.904	107.9245	2192.105	104.7799	3254.762	103.1698	4115.151	102.4906
20	1137.037	109.6429	1600.000	106.6667	2044.444	105.1429	2322.222	104.5000
50	822.2222	113.8462	986.3636	111.2821	1103.086	109.9692	1162.091	109.4154
100	633.3333	118.7500	700.0000	116.6667	741.0256	115.6000	760.0660	115.1500
200	507.4074	124.5455	534.2105	123.0303	549.3464	122.2545	556.0531	121.9273

Where vopt ($\bar{y}_{sp}$) is defined by Eq. 9.

Optimum estimator amongst the family of estimators, $\bar{\mathbf{y}}_{s\lambda}$: To find the optimum value of λ which minimizes the Mean Square Error (MSE) of, $\bar{y}_{s\lambda}$ we differentiate Eq. 19 with respect to λ and equal to 0. Thus:

$$\lambda_{opt} = \frac{\mu_s^2}{\mu_s^2 + V_{opt}(\bar{y}_{sp})} \quad (21)$$

Thus, the optimum estimator, say, $\bar{y}_{s\lambda opt}$ is given by:

$$\bar{y}_{s\lambda opt} = \lambda_{opt}\bar{y}_{sp} \quad (22)$$

Table 7: PRE of $\bar{y}_{sp}$ with respect to $\hat{\mu}_{s1}$ and $\hat{\mu}_{s2}$ for n = 50, n_1 = 30, n_2 = 20 and different valuesn of g, σ_s^2 with $\sum_{i=1}^{g}\sigma_{ti}^2 = 0.1$

$\sum_{i=1}^{g}\sigma_{ti}^2 = 0.1$	g = 2		g = 3		g = 5		g = 10	
σ_s^2	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$
0.25	403.0303	133.0000	419.1489	131.3333	428.0840	130.4800	432.0053	130.1200
0.5	338.0952	142.0000	344.5652	140.8889	348.0159	140.3200	349.5010	140.0800
1	303.2520	149.2000	306.0440	148.5333	307.5033	148.1920	308.1252	148.0480
2	285.1852	154.0000	286.4641	153.6296	287.1257	153.4400	287.4063	153.3600
3	279.0634	155.8462	279.8893	155.5897	280.3151	155.4585	280.4954	155.4031
4	275.9834	156.8235	276.5928	156.6275	276.9064	156.5271	277.0391	156.4847
5	274.1298	157.4286	274.6120	157.2698	274.8601	157.1886	274.9650	157.1543
10	270.4073	158.6829	270.6437	158.6016	270.7650	158.5600	270.8163	158.5424
20	268.5393	159.3333	268.6563	159.2922	268.7163	159.2711	268.7416	159.2622
50	267.4163	159.7313	267.4628	159.7148	267.4866	159.7063	267.4967	159.7027
100	267.0416	159.8653	267.0648	159.8570	267.0767	159.8528	267.0817	159.8510
200	266.8541	159.9326	266.8657	159.9284	266.8717	159.9263	266.8742	159.9254

Table 8: PRE of $\bar{y}_{sp}$ with respect to $\hat{\mu}_{s1}$ and $\hat{\mu}_{s2}$ for n = 50, n_1 = 30, n_2 = 20 and different valuesn of g, σ_s^2 with $\sum_{i=1}^{g}\sigma_{ti}^2 = 1$

$\sum_{i=1}^{g}\sigma_{ti}^2 = 1$	g = 2		g = 3		g = 5		g = 10	
σ_s^2	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$
0.25	1016.6670	110.9091	1369.2300	107.8788	1680.4600	106.3273	1862.8200	105.6727
0.5	766.6667	115.0000	918.1818	112.2222	1025.9260	110.8000	1080.3920	110.2000
1	566.6667	121.4286	625.0000	119.0476	660.8974	117.8286	677.5578	117.3143
2	433.3333	130.0000	455.2632	128.1481	467.6471	127.2000	473.1343	126.8000
3	382.0513	135.4545	394.6429	133.9394	401.5351	133.1636	404.5404	132.8364
4	354.9020	139.2308	363.5135	137.9487	368.1518	137.2923	370.1579	137.0154
5	338.0952	142.0000	344.5652	140.8889	348.0159	140.3200	349.5010	140.0800
10	303.2520	149.2000	306.0440	148.5333	307.5033	148.1920	308.1252	148.0480
20	285.1852	154.0000	286.4641	153.6296	287.1257	153.4400	287.4063	153.3600
50	274.1294	157.4286	274.6120	157.2698	274.8601	157.1886	274.9650	157.1543
100	270.4073	158.6829	270.6437	158.6016	274.8601	158.5600	270.8163	158.5424
200	268.5393	159.3333	268.6563	159.2922	268.7163	159.2711	268.7416	159.2622

Table 9: PRE of $\bar{y}_{sp}$ with respect to $\hat{\mu}_{s1}$ and $\hat{\mu}_{s2}$ for n = 50, n_1 = 30, n_2 = 20 and different valuesn of g, σ_s^2 with $\sum_{i=1}^{g}\sigma_{ti}^2 = 100$

$\sum_{i=1}^{g}\sigma_{ti}^2 = 100$	g = 2		g = 3		g = 5		g = 10	
σ_s^2	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s1}$	$\bar{y}_{sp}$ with $\hat{\mu}_{s2}$
0.25	1751.815	106.0539	3771.149	102.7239	9913.725	101.0190	33466.67	100.2997
0.5	1737.255	106.1078	3695.694	102.7811	9377.778	101.0778	27933.33	100.3593
1	1708.974	106.2151	3554.128	102.8951	8466.667	101.1952	21016.67	100.4781
2	1655.556	106.4286	3303.389	103.1217	7100.000	101.4286	14100.00	100.7143
3	1605.952	106.6403	3088.189	103.3465	6123.809	101.6601	10641.67	100.9486
4	1559.770	106.8504	2901.470	103.5696	5391.667	101.8898	8566.667	101.1811
5	1516.667	107.0588	2737.931	103.7908	4822.222	102.1176	7183.333	101.4118
10	1338.095	108.0769	2152.632	104.8718	3195.238	103.2308	4039.394	102.5385
20	1100.000	110.0000	1546.429	106.9136	1975.000	105.3333	2242.857	104.6667
50	766.6667	115.0000	918.1818	112.2222	1025.924	110.8000	1080.392	110.2000
100	566.6667	121.4286	625.0000	119.0476	660.8974	117.8286	677.5578	117.3143
200	433.3333	130.0000	455.2632	128.1481	467.6471	127.2000	473.1343	126.8000

It bias is given by:

$$\text{Bias}\left(\bar{y}_{s\lambda opt}\right) = \left(\lambda_{opt}\text{-}1\right)\mu_s \quad (23)$$

And the Mean Square Error, MSE is given by:

$$\text{MSE}\left(\bar{y}_{s\lambda opt}\right) = \frac{\mu_s^2 V_{opt}\left(\bar{y}_{sp}\right)}{\mu_s^2 + V_{opt}\left(\bar{y}_{sp}\right)} \quad (24)$$

The Relative Efficiency (RE) of the optimum estimator $\bar{y}_{s\lambda opt}$ with respect to the proposed estimator $\bar{y}_{sp}$ is given by $RE = \frac{V_{opt}(\bar{y}_{sp})}{MSE(\bar{y}_{s\lambda opt})}$:

$$RE = 1 + \frac{V_{opt}(\bar{y}_{sp})}{\mu_s^2} \tag{25}$$

Since, RE > 1, it is clear that the optimum estimator $\bar{y}_{s\lambda opt}$ is always more efficient than the proposed estimator, $\bar{y}_{sp}$.

Estimation method which utilizes priori information: The use of prior knowledge about a population parameter has been proven to improve the precision and efficiency of estimation of the study variable. Bayesian method of estimation that utilizes prior information in the form of prior distribution is well known for this. In some cases, prior information are used alongside with sample information to get a more precise Statistically efficient estimate of the parameter of interest. In the light of Thompson (1968) and Mathur and Singh (2008), we define another estimator for μ_s. Let μ_{s0} be the prior estimate or guessed value of the population mean of sensitive variable, μ_s then we define a class new of estimators as:

$$\bar{y}_{sk} = K\bar{y}_{sp} + (1-K)\mu_{s0},\ 0 < K \le 1 \tag{26}$$

where, K is a constant specified by the researcher according to his/her belief in the priorestimate μ_{s0}. If K is close to zero, it shows his/her strong belief in $\mu_{s0}$and if K is close to 1 it indicates strong belief in μ_s.The bias of $\bar{y}_{sk}$ is given by:

$$\text{Bias}(\bar{y}_{sk}) = (1-K)\mu_s w \tag{27}$$

And the Mean Square Error, MSE is given by:

$$MSE(\bar{y}_{sk}) = w^2(1-K)^2\mu_s^2 + K^2V_{opt}(\bar{y}_{sp}) \tag{28}$$

Where $w = \mu_{s0}/\mu_s - 1$.

Efficiency comparison: The suggested estimator $\bar{y}_{sk}$ is more efficient than the estimator $\bar{y}_{sp}$, if:

$$MSE(\bar{y}_{sk}) - V_{opt}(\bar{y}_{sp}) \le 0$$

It can be shown that $MSE(\bar{y}_{sk}) - V_{opt}(\bar{y}_{sp}) \le 0$ if and only if:

$$\frac{w^2\mu_s^2 - V_{opt}(\bar{y}_{sp})}{w^2\mu_s^2 + V_{opt}(\bar{y}_{sp})} < K \le 1 \tag{29}$$

Where $V_{opt}(\bar{y}_{sp})$ is given by Eq. 9.

Optimum estimator of $\bar{y}_{sk}$: To obtain the optimum value of K that minimizes the MSE ($\bar{y}_{sk}$), we differentiate Eq. 28 with respect to K. Thus:

$$K_{opt} = \frac{w^2\mu_s^2}{w^2\mu_s^2 + V_{opt}(\bar{y}_{sp})} \tag{30}$$

Therefore, the optimum estimator, say $\bar{y}_{skopt}$ is given by:

$$\bar{y}_{skopt} = K_{opt}\,\bar{y}_{sp} + (1-K_{opt})\mu_{s0} \tag{31}$$

It Means Square Error (MSE) is given by:

$$MSE(\bar{y}_{skopt}) = \frac{w^2\mu_s^2V_{opt}(\bar{y}_{sp})}{w^2\mu_s^2 + V_{opt}(\bar{y}_{sp})} \tag{32}$$

The Relative Efficiency (RE) of the optimum estimator $\bar{y}_{skopt}$ with respect to the proposed estimator $\bar{y}_{sp}$ is given by $RE = \frac{V_{opt}(\bar{y}_{sp})}{MSE(\bar{y}_{skopt})}$:

$$RE = 1 + \frac{V_{opt}(\bar{y}_{sp})}{w^2\mu_s^2} \tag{33}$$

Since, RE >1, it is clear that the optimum estimator, $\bar{y}_{skopt}$ is always more efficient than the proposed estimator $\bar{y}_{sp}$.

RESULTS AND DISCUSSION

Some alternative family of estimators using auxiliary variable: The use of auxiliary information has been proven to improve the process of estimation of study variable in the literature. The procedures of ratio, product and regression methods of Estimation are well known for this. Cochran (1940) first introduced ratio estimator of the population mean to show the contribution of supplementary information in the estimation process. The use of auxiliary information can also be utilized in sensitive surveys (as in usual surveys) of which our interest is to estimate the parameter (say mean or proportion) of the population bearing stigmatizing attribute. Auxiliary information can either be utilized at the design or estimation stage. Yan (2005), Diana and Perri (2009, 2010) and Hussain *et al* (2015) are some of the studies that utilize auxiliary information at estimation stage. This research also proposes a class of estimators which utilizes the known supplementary information. From Eq. 7 we have that the population mean of sensitive variable is estimated by:

$$\hat{\mu}_s = \alpha_1\hat{\mu}_{s1} + \alpha_2\hat{\mu}_{s2}$$

Where:

$$\alpha_1 + \alpha_2 = 1,\quad \hat{\mu}_{s1} = \bar{y}_1 - \bar{y}_2 \quad \text{and} \quad \hat{\mu}_{s2} = \left(\bar{y} + \sum_{i=1}^{g}\mu_{ti}\right)\Big/ g$$

We can respectively, replace:

$$\bar{y}_1, \bar{y}_2 \text{ and } \bar{y} \text{ by } \bar{y}_{1R}, \bar{y}_{2R} \text{ and } \bar{y}_R;$$
$$\bar{y}_{1P}, \bar{y}_{2P} \text{ and } \bar{y}_P; \text{ and } \bar{y}_{1lr}, \bar{y}_{2lr} \text{ and } \bar{y}_{lr}$$

Where:

$$\bar{y}_{iR} = \frac{\bar{y}_i\bar{X}}{\bar{x}}, \bar{y}_R = \frac{\bar{y}\bar{X}}{\bar{x}}; \quad \bar{y}_{iP} = \frac{\bar{y}_i\bar{x}}{\bar{X}}, \bar{y}_P = \frac{\bar{y}\bar{x}}{\bar{X}};$$
$$\bar{y}_{ilr} = \bar{y}_i + b\left(\bar{X}-\bar{x}\right), \bar{y}_{lr} = \bar{y} + b\left(\bar{X}-\bar{x}\right)$$

(i = 1, 2) and $\bar{X}$ and $\bar{x}$ are the population mean and sample mean of auxiliary variable, respectively in order to estimate μ_s. We will now look at these methods of estimation one after the other.

Ratio method of estimation: By replacing $\bar{y}_1$, $\bar{y}_2$ and $\bar{y}$ by $\bar{y}_{1R}$, $\bar{y}_{2R}$, respectively, in Eq. 7, we propose a new class of estimators given by:

$$\bar{y}_{sp1} = \alpha_1\hat{\mu}_{s11} + \alpha_2\hat{\mu}_{s21}$$

Where:

$$\hat{\mu}_{s11} = \bar{y}_{1R} - \bar{y}_{2R} \text{ and } \hat{\mu}_{s21} = \left(\bar{y}_R - \sum_{i=1}^{g}\mu_{ti}\right) \Big/ g$$

Then the new class of regression estimators is given by:

$$\bar{y}_{sp1} = \alpha_1\left(\bar{y}_{1R} - \bar{y}_{2R}\right) + \alpha_2\left(\frac{\bar{y}_R - \sum_{i=1}^{g}\mu_{ti}}{g}\right) \tag{34}$$

The biases of $\bar{y}_{iR}$ and $\bar{y}_R$ (i = 1, 2) are respectively, given by:

$$\text{Bias}\left(\bar{y}_{iR}\right) = \frac{1-f_i}{n_i}\bar{Y}_i\left(C_x^2 - \rho_{y_ix}C_{y_i}C_x\right)$$

And:

$$\text{Bias}\left(\bar{y}_R\right) = \frac{1-f}{n}\bar{Y}\left(C_x^2 - \rho_{yx}C_yC_x\right)$$

And the bias $\bar{y}_R$ of is given by:

$$\text{Bias}\left(\bar{y}_R\right) = \frac{1-f}{n}\bar{Y}\left(C_x^2 - \rho_{yx}C_yC_x\right)$$

The Mean Square Errors (MSEs) of $\bar{y}_{iR}$ and $\bar{y}_R$ are respectively, given by:

$$\text{MSE}\left(\bar{y}_{iR}\right) = \frac{1-f_i}{n_i}\left(\sigma_{y_i}^2 + R_i^2\sigma_x^2 - 2R_i\rho_{y_ix}\sigma_{y_i}\sigma_x\right)$$

And:

$$\text{MSE}\left(\bar{y}_R\right) = \frac{1-f}{n}\left(\sigma_y^2 + R^2\sigma_x^2 - 2R\rho_{yx}\sigma_y\sigma_x\right)$$

Where:

$$R_i = \frac{\bar{Y}_i}{\bar{X}}, R = \frac{\bar{Y}}{\bar{X}}, f_i = \frac{n_i}{N}, f = \frac{n}{N}, \bar{X}$$

The population mean of the auxiliary variable and N the population size. Now, the bias of $\bar{y}_{sp1}$ is given by:

$$\text{Bias}\left(\bar{y}_{sp1}\right) = \alpha_1\text{Bias}\left(\hat{\mu}_{s11}\right) + \alpha_2\text{Bias}\left(\hat{\mu}_{s21}\right)$$

$$\text{Bias}\left(\bar{y}_{sp1}\right) = \alpha_1\left[\frac{1-f_1}{n_1}\left(\mu_s + \sum_{i=1}^{g}\mu_{ti}\right)\left(C_x^2 - \frac{\rho_{sx}C_sC_x}{1+\sum_{i=1}^{g}\mu_{ti}\Big/\mu_s}\right) - \frac{1-f_2}{n_2}\left(\sum_{i=1}^{g}\mu_{ti}\right)C_x^2\right] +$$
$$\alpha_2\left[\frac{1-f}{n}\left(g\mu_s + \sum_{i=1}^{g}\mu_{ti}\right)\left(C_x^2 - \frac{\rho_{sx}C_sC_x}{1+\sum_{i=1}^{g}\mu_{ti}\Big/g\mu_s}\right)\right] \tag{35}$$

Mean square error of $\bar{y}_{sp1}$: The mean square error MSE of the suggested estimator, $\bar{y}_{sp1}$ is given by:

$$\text{MSE}\left(\bar{y}_{sp1}\right) = \alpha_1^2\,\text{MSE}\left(\hat{\mu}_{s11}\right) + \alpha_2^2\,\text{MSE}\left(\hat{\mu}_{s21}\right)$$

Where $\hat{\mu}_{s11}$ and $\hat{\mu}_{s21}$ as defined above we have that:

$$\text{MSE}\left(\bar{y}_{sp1}\right) = \alpha_1^2\left[\text{MSE}\left(\hat{\mu}_{s11}\right) + \text{MSE}\left(\hat{\mu}_{s21}\right)\right]$$
$$-2\alpha_1\text{MSE}\left(\hat{\mu}_{s21}\right) + \text{MSE}\left(\hat{\mu}_{s21}\right)$$

And at optimal point:

$$\text{MSE}_{opt}(\bar{y}_{sp1}) = \text{MSE}(\hat{\mu}_{s21}) - \frac{\left(\text{MSE}(\hat{\mu}_{s21})\right)^2}{\text{MSE}(\hat{\mu}_{s11}) + \text{MSE}(\hat{\mu}_{s21})} \tag{36}$$

The $\text{MSE}(\hat{\mu}_{s11}) = \text{MSE}(\bar{y}_{1R}) + \text{MSE}(\bar{y}_{2R})$ and $\text{MSE}(\hat{\mu}_{s21})$ are given by:

$$\text{MSE}\left(\hat{\mu}_{s11}\right) = \frac{1-f_1}{n_1}\left(\sigma_s^2 + R_1^2\sigma_x^2 - 2R_1\rho_{sx}\sigma_s\sigma_x + \frac{2R_1\sigma_x^2\sum_{i=1}^{g}\mu_{ti}}{\bar{X}} - \frac{2\rho_{sx}\sigma_s\sigma_x\sum_{i=1}^{g}\mu_{ti}}{\bar{X}}\right) +$$
$$\frac{n - n_1f_2 - n_2f_1}{n_1n_2}\left(\sum_{i=1}^{g}\sigma_{ti}^2 + \frac{(\sum_{i=1}^{g}\mu_{ti})^2\sigma_x^3}{\bar{X}^2}\right)$$

$$\mathrm{MSE}(\hat{\mu}_{s21}) = \frac{1-f}{ng^2}\left[g^2\sigma_s^2 + g^2R_1^2\sigma_x^2 - 2g^2R_1\sigma_s\sigma_x + \sum_{i=1}^{g}\sigma_{ti}^2 + \frac{2gR_1\sigma_x^2\sum_{i=1}^{g}\mu_{ti}}{\bar{X}} + \frac{\sigma_x^2(\sum_{i=1}^{g}\mu_{ti})^2}{\bar{X}^2} - \frac{2g\rho_{sx}\sigma_s\sigma_x\sum_{i=1}^{g}\mu_{ti}}{\bar{X}}\right]$$

Product method of estimation: By replacing $\bar{y}_1$, $\bar{y}_2$ and $\bar{y}$ by $\bar{y}_{1p}$, $\bar{y}_{2p}$, respectively in Eq. 7, we propose a new class of estimators given by:

$$\bar{y}_{sp2} = \alpha_1\hat{\mu}_{s12} + \alpha_2\hat{\mu}_{s22}$$

Where:

$$\hat{\mu}_{s12} = \bar{y}_{1P} - \bar{y}_{2P} \text{ and } \hat{\mu}_{s22} = \frac{\bar{y}_P - \sum_{i=1}^{g}\mu_{ti}}{g}$$

Then the new class of regression estimators is given by:

$$\bar{y}_{sp2} = \alpha_1(\bar{y}_{1P} - \bar{y}_{2P}) + \alpha_2\left(\frac{\bar{y}_P - \sum_{i=1}^{g}\mu_{ti}}{g}\right) \quad (37)$$

The biases of $\bar{y}_{ip}$ and $\bar{y}_p$ (i = 1, 2) are respectively, given by:

$$\mathrm{Bias}(\bar{y}_{iP}) = \frac{1-f_i}{n_i}\bar{Y}_i\left(C_x^2 + \rho_{y_ix}C_{y_i}C_x\right)$$

And:

$$\mathrm{Bias}(\bar{y}_p) = \frac{1-f}{n}\bar{Y}\left(C_x^2 + \rho_{yx}C_yC_x\right)$$

And their Mean Square Errors (MSEs) are respectively, given by:

$$\mathrm{MSE}(\bar{y}_{iP}) = \frac{1-f_i}{n_i}\left(\sigma_{y_i}^2 + R_i^2\sigma_x^2 + 2R_i\rho_{y_ix}\sigma_{y_i}\sigma_x\right)$$

And:

$$\mathrm{MSE}(\bar{y}_P) = \frac{1-f}{n}(\sigma_y^2 + R^2\sigma_x^2 + 2R\rho_{yx}\sigma_y\sigma_x)$$

where:

$$R_i = \frac{\bar{Y}_i}{\bar{X}}, R = \frac{\bar{Y}}{\bar{X}}, f_i = \frac{n_i}{N}, f = \frac{n}{N}, \bar{X}$$

The population mean of the auxiliary variable and N the population size. Now, the bias of $\bar{y}_{sp2}$ is given by:

$$\mathrm{Bias}(\bar{y}_{sp2}) = \alpha_1\left\{\frac{1-f_1}{n_1}\left(\mu_s + \sum_{i=1}^{g}\mu_{ti}\right)\left(C_x^2 + \frac{\rho_{sx}C_sC_x}{1+\sum_{i=1}^{g}\mu_{ti}/\mu_s}\right) - \frac{1-f_2}{n_2}C_x^2\sum_{i=1}^{g}\mu_{ti}\right\} - \alpha_2\left[\frac{1-f}{n}\left(g\mu_s + \sum_{i=1}^{g}\mu_{ti}\right) + \frac{\rho_{sx}C_sC_x}{1+\sum_{i=1}^{g}\mu_{ti}/g\mu_s}\right] \quad (38)$$

Mean square error of $\bar{y}_{sp}2$: The mean square error of $\bar{y}_{sp2}$ is given by:

$$\mathrm{MSE}(\bar{y}_{s22}) = \alpha_1^2\mathrm{MSE}(\hat{\mu}_{s12}) + \alpha_2^2\mathrm{MSE}(\hat{\mu}_{s22})$$

It also follows that at optimal point, MSE of $\bar{y}_{sp2}$ will be given by:

$$\mathrm{MSE}_{opt}(\bar{y}_{sp2}) = \mathrm{MSE}(\hat{\mu}_{s22}) - \frac{(\mathrm{MSE}(\hat{\mu}_{s22}))^2}{\mathrm{MSE}(\hat{\mu}_{s12}) + \mathrm{MSE}(\hat{\mu}_{s22})} \quad (39)$$

The $\mathrm{MSE}(\hat{\mu}_{s12}) = \mathrm{MSE}(\bar{y}_{1P}) + \mathrm{MSE}(\bar{y}_{2P})$ and $\mathrm{MSE}(\hat{\mu}_{s22})$ are given by:

$$\mathrm{MSE}(\hat{\mu}_{s12}) = \frac{1-f_1}{n_1}\left(\sigma_s^2 + R_1^2\sigma_x^2 + 2R_1\rho_{sx}\sigma_s\sigma_x + \frac{2R_1\sigma_x^2\sum_{i=1}^{g}\mu_{ti}}{\bar{X}} + \frac{2\rho_{sx}\sigma_s\sigma_x\sum_{i=1}^{g}\mu_{ti}}{\bar{X}}\right) + \frac{n-n_1f_2-n_2f_1}{n_1n_2}\left(\sum_{i=1}^{g}\sigma_{ti}^2 + \frac{\left(\sum_{i=1}^{g}\mu_{ti}\right)^2\sigma_x^2}{\bar{X}^2}\right) \quad (40)$$

$$\mathrm{MSE}(\hat{\mu}_{s22}) = \frac{1-f}{ng^2}\left(g^2\sigma_s^2 + g^2R_1^2\sigma_x^2 + 2g^2R_1\rho_{sx}\sigma_s\sigma_x + \sum_{i=1}^{g}\sigma_{ti}^2 + \frac{2gR_1\sigma_x^2\sum_{i=1}^{g}\mu_{ti}}{\bar{X}} + \frac{\sigma_x^2(\sum_{i=1}^{g}\mu_{ti})^2}{\bar{X}^2} + \frac{2g\rho_{sx}\sigma_s\sigma_x\sum_{i=1}^{g}\mu_{ti}}{\bar{X}}\right) \quad (41)$$

Where:

$$R_1 = \frac{\mu_s}{\bar{X}} \text{ and } \rho_{yx} = \rho_{sx}g\sigma_s/\sigma_y$$

Regression method of estimation: Equally, by replacing $\bar{y}_1$, $\bar{y}_2$ and $\bar{y}$ in our original model by $\bar{y}_{1lr}$, $\bar{y}_{2lr}$ and $\bar{y}_{lr}$, respectively in Eq. 7, we propose a new family of regression estimators:

$$\bar{y}_{sp3} = \alpha_1(\hat{\mu}_{s13})+\alpha_2(\hat{\mu}_{s23})$$

Where:

$$\hat{\mu}_{s13} = \bar{y}_{1lr}-\bar{y}_{2lr} \text{ and } \hat{\mu}_{s23} = \frac{\bar{y}_{lr}-\sum_{i=1}^{g}\mu_{ti}}{g}$$

Therefore:

$$\bar{y}_{sp3} = \alpha_1\left(\bar{y}_{1lr}-\bar{y}_{2lr}\right)+\alpha_2\left(\frac{\bar{y}_{lr}-\sum_{i=1}^{g}\mu_{ti}}{g}\right)$$

Where:

$\alpha_1+\alpha_2 = 1$

$\bar{y}_{ilr} = \bar{y}_i+b(\bar{X}-\bar{x})$

$\bar{y}_{lr} = \bar{y}+b(\bar{X}-\bar{x})$

$i = 1, 2$

Since, the regression coefficient may or may not be known, we consider both cases.

When coefficient b is known: With the estimators $\bar{y}_{ilr} = \bar{y}_i+b(\bar{X}-\bar{x})$ and $\bar{y}_{lr} = \bar{y}+b(\bar{X}-\bar{x})$ (i = 1, 2) we have that:

$$\text{Bias}\left(\bar{y}_{ilr}\right) = \text{Bias}\left(\bar{y}_{lr}\right) = 0$$

And their variances are given by:

$$\text{Var}\left(\bar{y}_{ilr}\right) = \frac{1-f_i}{n_i}\left(\sigma_{y_i}^2+b^2\sigma_x^2-2b\rho_{yx}\sigma_y\sigma_x\right) \quad (42)$$

And

$$\text{Var}\left(\bar{y}_{lr}\right) = \frac{1-f}{n}\left(\sigma_y^2+b^2\sigma_x^2-2b\rho_{yx}\sigma_y\sigma_x\right) \quad (43)$$

Since, b is a known regression coefficient, then:

$$b = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{\rho_{yx}\sigma_y}{\sigma_x}$$

And:

$$b = \frac{\sigma_{y_ix}}{\sigma_x^2} = \frac{\rho_{y_ix}\sigma_{y_i}}{\sigma_x}$$

(i =1, 2) also minimizes the values of Var ($\bar{y}_{lr}$) and Var ($\bar{y}_{ilr}$) (i = 1, 2), respectively. The substitution of the optimal value of b in Eq. 42 and 43 gives the minimum variances of $\bar{y}_{ilr}$ and $\bar{y}_{lr}$. Thus:

$$\text{Var}_{min}\left(\bar{y}_{ilr}\right) = \frac{1-f_i}{n_i}\sigma_{y_i}^2\left(1-\rho_{y_ix}^2\right)$$

And:

$$\text{Var}_{min}\left(\bar{y}_{lr}\right) = \frac{1-f}{n}\sigma_y^2\left(1-\rho_{yx}^2\right)$$

Recall $\rho_{yx} = \rho_{sx}g\sigma_s/\sigma_y$ and it also follows that $\rho_{y1x} = \rho_{sx}\sigma_s/\sigma_{y1}$ and $\rho_{y2x} = 0$, since, $\rho_{sx} = 0$. Now:

$$\text{Var}\left(\bar{y}_{sp3}\right) = \alpha_1^2\text{Var}\left(\hat{\mu}_{s13}\right)+\alpha_2^2\text{Var}\left(\hat{\mu}_{s23}\right)$$

It follows from study 4.2 that the optimum variance of $\bar{y}_{sp3}$ is given by:

$$V_{opt}(\bar{y}_{sp3}) = \text{Var}(\hat{\mu}_{s23})-\frac{\left(\text{Var}(\hat{\mu}_{s23})\right)^2}{\text{Var}(\hat{\mu}_{s13})+\text{Var}(\hat{\mu}_{s23})} \quad (44)$$

With appropriate substitution, we can write the expression for the variance of the suggested estimator, $\bar{y}_{sp3}$ as:

$$\text{Var}\left(\bar{y}_{sp3}\right) = \beta_1\sigma_s^2\left(1+\frac{\sum_{i=1}^{g}\sigma_{ti}^2}{g^2\sigma_s^2}-\rho_{sx}^2\right)-\frac{\left\{\beta_1\sigma_s^2\left(1+\frac{\sum_{i=1}^{g}\sigma_{ti}^2}{g^2\sigma_s^2}-\rho_{sx}^2\right)\right\}^2}{\left(\beta_2\sigma_s^2\left(1-\rho_{sx}^2\right)+\beta_3\sum_{i=1}^{g}\sigma_{ti}^2\right)+\beta_1\sigma_s^2\left(1+\frac{\sum_{i=1}^{g}\sigma_{ti}^2}{g^2\sigma_s^2}-\rho_{sx}^2\right)} \quad (45)$$

Where:

$\beta_1 = 1\text{-}f/n$

$\beta_2 = 1\text{-}f_1/n_1$

$\beta_3 = n\text{-}n_1f_1\text{-}n_2f_1$

When b is unknown: Suppose in the regression of $y = b_0+bx+e$ where b_0 is a constant term and e is a random error term, the regression coefficient, b is not known. The unbiased ordinary least square estimator of b is given by $\hat{b} = \sigma_{xy}/\sigma_x^2$ which also minimizes the error sum of squares. Thus, we have new class of estimators as:

$$\bar{y}_{sp3} = \alpha_1(\bar{y}_{1lr}-\bar{y}_{2lr})+\alpha_2\left(\frac{\bar{y}_{lr}-\sum_{i=1}^{g}\mu_{ti}}{g}\right) \quad (46)$$

If the same steps in Eq. 43 are followed, we will arrive at the same variance and as such there is no need to consider bothcases or regression coefficient separately, since, the result in one case remains valid in the other case.

Efficiency comparison

The proposed $\bar{y}_{sp}$ with ratio estimator, $\bar{y}_{sp1}$: The estimator $\bar{y}_{sp1}$ will be more efficient than the estimator $\bar{y}_{sp}$, if:

$$V_{opt}(\bar{y}_{sp})\text{-}MSE_{opt}(\bar{y}_{sp1}) > 0$$

From Eq. 10 and 36, we have:

$$Var(\hat{\mu}_{s2})\text{-}\frac{(Var(\hat{\mu}_{s2}))^2}{Var(\hat{\mu}_{s1})+Var(\hat{\mu}_{s2})}\text{-}\left\{MSE(\hat{\mu}_{s21})\text{-}\frac{(MSE(\hat{\mu}_{s21}))^2}{MSE(\hat{\mu}_{s11})+MSE(\hat{\mu}_{s21})}\right\}>0 \quad (47)$$

The above inequality in Eq. 52 can only hold if $Var(\hat{\mu}_{s2})\text{-}MSE(\hat{\mu}_{s21})$ and $Var(\hat{\mu}_{s1})\text{-}MSE(\hat{\mu}_{s11})$ are all non negative. From Eq. 10 and 37, being non negative implies that:

$$\rho_{sx} > \frac{\sigma_x\left(gR_1+\bar{X}^{-1}\sum_{i=1}^{g}\mu_{ti}\right)}{2g\sigma_s} \quad (48)$$

From Eq. 5 and 48, $Var(\hat{\mu}_{s1})\text{-}MSE(\hat{\mu}_{s11})$ being non negative implies that:

$$\rho_{sx} > \frac{\sigma_x(R_1+\bar{X}^{-1}\sum_{i=1}^{g}\mu_{ti})}{2\sigma_s}+\frac{n_1(1\text{-}f_2)}{n_2(1\text{-}f_1)}\frac{\sigma_x(\sum_{i=1}^{g}\mu_{ti})^2}{2\sigma_s\bar{X}^2(R_1+\bar{X}^{-1}\sum_{i=1}^{g}\mu_{ti})} \quad (49)$$

We infer that the estimator $\bar{y}_{sp1}$ is more efficient than estimator $\bar{y}_{sp}$ if the above inequalities in Eq. 48 and 49 are satisfied.

The proposed estimator, $\bar{y}_{sp}$ with the product estimator, $\bar{y}_{sp2}$: The estimator $\bar{y}_{sp2}$ will be more efficient than the estimator $\bar{y}_{sp}$ if:

$$V_{opt}(\bar{y}_{sp})\text{-}MSE_{opt}(y_{sp2}) > 0$$

From Eq. 10 and 40, we have:

$$Var(\hat{\mu}_{s2})\text{-}\frac{(Var(\hat{\mu}_{s2}))^2}{Var(\hat{\mu}_{s1})+Var(\hat{\mu}_{s2})}\text{-}\left\{MSE(\hat{\mu}_{s22})\text{-}\frac{(MSE(\hat{\mu}_{s22}))^2}{MSE(\hat{\mu}_{s12})+MSE(\hat{\mu}_{s22})}\right\}>0 \quad (50)$$

The above inequality in Eq. 50 can only hold if $Var(\hat{\mu}_{s2})\text{-}MSE(\hat{\mu}_{s22})$ and $Var(\hat{\mu}_{s1})\text{-}MSE(\hat{\mu}_{12})$ are all non negative. From Eq. 10 and 42, $Var(\hat{\mu}_{s2})\text{-}MSE(\hat{\mu}_{s22})$ being non negative implies that:

$$\rho_{sx} < \text{-}\frac{\sigma_x\left(gR_1+\bar{X}^{-1}\sum_{i=1}^{g}\mu_{ti}\right)}{2g\sigma_s} \quad (51)$$

From Eq. 5 and 41, $Var(\hat{\mu}_{s1})\text{-}MSE(\hat{\mu}_{12})$ being non negative implies that:

$$\rho_{sx} < \text{-}\left\{\frac{\sigma_x\left(R_1+\bar{X}^{-1}\sum_{i=1}^{g}\mu_{ti}\right)}{2\sigma_s}+\frac{n_1(1\text{-}f_2)}{n_2(1\text{-}f_1)}\frac{\sigma_x(\sum_{i=1}^{g}\mu_{ti})^2}{2\sigma_s\bar{X}^2\left(R_1+\bar{X}^{-1}\sum_{i=1}^{g}\mu_{ti}\right)}\right\} \quad (52)$$

We infer that the estimator $\bar{y}_{sp2}$ is more efficient than estimator $\bar{y}_{sp}$ if the above inequalities in Eq. 51 and 52 are satisfied.

The proposed estimator, $\bar{y}_{sp}$ with regression estimator, $\bar{y}_{sp3}$: The estimator $\bar{y}_{sp3}$ will be more efficient than the proposed estimator $\bar{y}_{sp}$ if:

$$V_{opt}(\bar{y}_{sp})\text{-}V_{optt}(\bar{y}_{sp3}) > 0$$

From Eq. 10 and 40, we have:

$$Var(\hat{\mu}_{s2})\text{-}\frac{(Var(\hat{\mu}_{s2}))^2}{Var(\hat{\mu}_{s1})+Var(\hat{\mu}_{s2})}\text{-}\left\{Var(\hat{\mu}_{s23})\text{-}\frac{(Var(\hat{\mu}_{s23}))^2}{Var(\hat{\mu}_{s13})+Var(\hat{\mu}_{s23})}\right\}>0 \quad (53)$$

The above inequality in Eq. 58 can only hold if $Var(\hat{\mu}_{s2})\text{-}Var(\hat{\mu}_{s23})$ and $Var(\hat{\mu}_{s1})\text{-}Var(\hat{\mu}_{13})$ areall non negative. From Eq. 10 and 45, $Var(\hat{\mu}_{s2})\text{-}Var(\hat{\mu}_{s23})$ being non negative implies that:

$$\rho_{sx} > 0 \quad (54)$$

From Eq. 5 and 45, $Var(\hat{\mu}_{s1})\text{-}Var(\hat{\mu}_{13})$ being non negative implies that:

$$\rho_{sx} > 0 \quad (55)$$

With the results in Eq. 54 and 55 we can infer that the estimator $\bar{y}_{sp}$ is more efficient than the proposed estimator,

Ratio estimator, $\bar{y}_{sp1}$ with regression estimator, $\bar{y}_{sp3}$: The estimator $\bar{y}_{sp3}$ will be more efficient than the estimato $\bar{y}_{sp1}$ if:

$$MSE_{opt}(\bar{y}_{sp1})-V_{opt}(\bar{y}_{sp3})>0$$

From Eq. 36 and 45, this becomes:

$$MSE(\hat{\mu}_{s21})-\frac{(MSE(\hat{\mu}_{s21}))^2}{MSE(\hat{\mu}_{s11})+MSE(\hat{\mu}_{s21})}-\left\{Var(\hat{\mu}_{s23})-\frac{(Var(\hat{\mu}_{s23}))^2}{Var(\hat{\mu}_{s13})+Var(\hat{\mu}_{s23})}\right\}>0 \quad (56)$$

The above inequality in Eq. 61 can only hold if $MSE(\hat{\mu}_{s21})-Var(\hat{\mu}_{s23})$ and $MSE(\hat{\mu}_{s11})-Var(\hat{\mu}_{s13})$ are all non negative. From Eq. 37 and 45, $MSE(\hat{\mu}_{s21})-Var(\hat{\mu}_{s23})$ being non negative implies that:

$$\left(g\sigma_s\rho_{sx}-\sigma_x\left(gR_1+\bar{X}^{-1}\sum_{i=1}^{g}\mu_{ti}\right)\right)^2\geq 0 \quad (57)$$

From Eq. 37 and 45, $MSE(\hat{\mu}_{s11})-Var(\hat{\mu}_{s13})$ being non negative implies that:

$$\frac{1-f_1}{n_1}\left(\sigma_s\rho_{sx}-\sigma_x(R_1+\bar{X}^{-1}\sum_{i=1}^{g}\mu_{ti})\right)^2+\frac{1-f_2}{n_2}\left(\sigma_x\bar{X}^{-1}\sum_{i=1}^{g}\mu_{ti}\right)^2\geq 0 \quad (58)$$

Since, the inequalities in Eq. 55 and 58 always hold, it implies that Eq. 56 is satisfied and we can infer that the estimator $\bar{y}_{sp3}$ is more efficient than the estimator, $\bar{y}_{sp1}$.

Product estimator, $\bar{y}_{sp2}$with regression estimator, $\bar{y}_{sp3}$: The regression estimator $\bar{y}_{sp3}$will be more efficient than the estimator $\bar{y}_{sp2}$ if:

$$MSE_{opt}(\bar{y}_{sp2})-V_{opt}(\bar{y}_{sp3})>0$$

From Eq. 40 and 45, we have:

$$MSE(\hat{\mu}_{s22})-\frac{(MSE(\hat{\mu}_{s22}))^2}{MSE(\hat{\mu}_{s12})+MSE(\hat{\mu}_{s22})}-\left\{Var(\hat{\mu}_{s23})-\frac{(Var(\hat{\mu}_{s23}))^2}{Var(\hat{\mu}_{s13})+Var(\hat{\mu}_{s23})}\right\}>0 \quad (59)$$

The above inequality in Eq. 59 can only hold if $MSE(\hat{\mu}_{s22})-Var(\hat{\mu}_{s23})$ and $MSE(\hat{\mu}_{s12})-Var(\hat{\mu}_{s13})$ are all non negative. From Eq. 42 and 45, $MSE(\hat{\mu}_{s22})-Var(\hat{\mu}_{s23})$ being non negative implies:

$$\left(g\sigma_s\rho_{sx}+\sigma_x(R_1g+\bar{X}^{-1}\sum_{i=1}^{g}\mu_{ti})\right)^2\geq 0 \quad (60)$$

From Eq. 41 and 45, $MSE(\hat{\mu}_{s12})-Var(\hat{\mu}_{s13})$ being non negative implies that:

$$\frac{1-f_1}{n_1}\left(\sigma_s\rho_{sx}+\sigma_x(R_1+\bar{X}^{-1}\sum_{i=1}^{g}\mu_{ti})\right)^2+\frac{1-f_2}{n_2}\left(\sigma_x\bar{X}^{-1}\sum_{i=1}^{g}\mu_{ti}\right)^2\geq 0 \quad (61)$$

Since, the inequalities in Eq. 60 and 61 always hold, it therefore, means that Eq. 60 is satisfied and we can infer that the estimator $\bar{y}_{sp3}$ is always more efficient than the estimator, $\bar{y}_{sp2}$.

CONCLUSION

In this research, we proposed an estimator which is a linear combination of two existing estimators (Trappmann *et al.*, 2013; Hussain *et al.*, 2015) estimators in the literature. The proposed estimator, $\bar{y}_{sp}$ has been shown to be better when its efficiency was compared to those of theexisting estimators. This provides a good alternative in measurement of sensitive items in the population especially where accuracy and efficiency are of primary concern. Furthermore, we presented some families of improved estimators of the population mean of sensitive variable, μ_s using different procedures. It has also been shown that these classes of estimators under certain conditions are more efficient than the proposed estimator, $\bar{y}_{sp}$. Moreover, the results obtained after comparing the efficiencies among these classes of estimators were consistent with those of the existing literature.

RECOMMENDATIONS

We recommend the application of this technique in the on going debate of sensitive studies as an alternative technique especially where cost is not an issue and efficiency and precision are of interest.

REFERENCES

Chaudhuri, A. and T.C. Christofides, 2007. Item count technique in estimating the proportion of people with a sensitive feature. J. Stat. Plann. Inf., 137: 589-593.

Cochran, W.G., 1940. The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce. J. Agric. Sci., 30: 262-275.

Diana, G. and P.F. Perri, 2009. Estimating a sensitive proportion through randomized response procedures based on auxiliary information. Stat. Pap., 50: 661-672.

Diana, G. and P.F. Perri, 2010. New scrambled response models for estimating the mean of a sensitive quantitative character. J. Appl. Stat., 37: 1875-1890.

Droitcour, J.A., E.M. Larson and U.S. General, 2001. The three card method: Estimating sensitive survey items with permanent anonymity of response. Proceedings of the International Conference on Social Statistics Section Vol. 63, August 5-9, 2001, American Statistical Association, USA., pp: 432-434.

Geurts, M.D., 1980. Using a randomized response research design to eliminate non-response and response biases in business research. J. Acad. Marketing Sci., 8: 83-91.

Gjestvang, C.R. and S. Singh, 2006. A new randomized response model. J. R. Stat. Soc.: Ser. B (Stat. Methodol.), 68: 523-530.

Holbrook, A.L. and J.A. Krosnick, 2009. Social desirability bias in voter turnout reports: Tests using the item count technique. Pub. Opin. Quart., 74: 37-67.

Horvitz, D.G., B.V. Shah and W.R. Simmons, 1967. The unrelated questions RR model. Proc. Soc. Stat. Sect. Am. Stat. Assoc., 1967: 65-72.

Hubbard, M.L., R.A. Casper and J.T. Lessler, 1989. Respondents reactions to item count lists and randomized response. Proc. Surv. Res. Sect. Am. Stat. Assoc., 1989: 544-448.

Hussain, Z., N. Shabbir and J. Shabbir, 2015. An alternative item sum technique for improved estimators of population mean in sensitive surveys. Hacettepe Univ. Bull. Nat. Sci. Eng. Ser. Math. Stat., 46: 1-10.

Kim, J.M. and W.D. Warde, 2004. A stratified Warner's randomized response model. J. Statist. Plann. Inference, 120: 155-165.

Kuk, A.Y.C., 1990. Asking sensitive questions indirectly. Biometrika, 77: 436-438.

Mathur, N. and H.P. Singh, 2008. Estimation of population mean with prior information using scrambled response technique. Brazil. J. Probab. Stat., 22: 165-181.

Miller, J.D., 1985. The nominative technique: A new method of estimating heroin prevalence. NIDA. Res. Monogr., 54: 104-124.

Searls, D.T., 1964. The utilization of a known coefficient of variation in the estimation procedure. J. Am. Statist. Assoc., 59: 1225-1226.

Smith, L.L., W.T. Federer and D. Raghavarao, 1974. A comparison of three techniques for eliciting truthful answers to sensitive questions. Proc. Soc. Stat. Sect. Am. Stat. Assoc., 1974: 447-452.

Thompson, J.R., 1968. Some shrinkage techniques for estimating the mean. J. Am. Stat. Assoc., 63: 113-122.

Trappmann, M., I. Krumpal, A. Kirchner and B. Jann, 2013. Item sum: A new technique for asking quantitative sensitive questions. J. Surv. Stat. Meth., 2: 58-77.

Warner, S.L., 1965. Randomized response: A survey technique for eliminating evasive answer bias. J. Am. Stat. Assoc., 60: 63-69.

Yan, Z., 2005. Ratio method of estimation of population proportion using randomized response technique. Mod. Assisted Stat. Appl., 1: 125-130.