Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/050217

Hansen and Hurwitz Estimator with Scrambled Response on Second Call in Stratified Random Sampling

Shakeel Ahmed[*], *Javid Shabbir and Waqar Hafeez*

Department of Statistics Quaid-i-Azam University Islamabad ,44000, Pakistan.

Received: 13 Dec. 2015, Revised: 15 Apr. 2016, Accepted: 29 Apr. 2016
Published online: 1 Jul. 2016

Abstract: We propose an estimator for finite population mean of the study character in presence of non-response utilizing known coefficient of variation under stratified sampling by using the scrambled response to sub-sample the non-respondent on second call. It is assumed that non-response occurs due to sensitivity of the character under study. The main purpose is to increase response rate by protecting confidentiality of the respondents by using scrambled response model for sub-sampling non-respondents on second call in Hansen and Hurwitz [1946, Journal of The American Statistical Association, vol. 41: 517-529] technique. We also propose a generalized ratio and regression type estimators under two phase stratified sampling. Proportional allocation method is used to allocate sample size at both phases. Expressions for mean squared errors (*MSE*) are derived up to first degree of approximation. An empirical study is carried out to observe performances of the estimators.

Keywords: Non-response, Scrambled response, Coefficient of variation, Stratified sampling, Privacy protection, Efficiency, Sensitivity, Generalized.

1 Introduction

Non-response occurs when respondents don't provide the required information asked by the investigator through mailed questionnaires as well as through direct interviews when people show lack of interest, not at home, unable to understand the questions etc. In some cases they refuse to answer when questions asked to them are related to such characters which are socially undesirable to them e.g questions related to drug addiction, illegal income and cheating in the examination etc. In case of non-response [12] proposed the method of sub-sampling non-respondents to provide an estimator for population mean of the study character. Later on using [12], many authors suggested different types of estimators for estimating the unknown population parameters under different sampling schemes using the auxiliary variable which may suffer non-response (see [10],[11],[23], [9] and [8]).
When non-response is due to sensitivity of the study variable then it is difficult to get a direct response on second call that results in violation of the assumption of estimator proposed by [12]. In this situation some statistical techniques called randomized response techniques (RRTs) are used to collect information on second call. The randomized response technique is primarily used to get trustworthy data for estimating the proportion of individuals with a sensitive character. Several researchers have extended the technique introduced by [7]. [6] has introduced the method to the situation when response to sensitive question is quantitative rather than qualitative. After that some other authors have introduced the scrambled response model (SRM) (see [5], [4], [2]) to get truthful response. In SRM the respondent multiply his/her sensitive answer by a random number generated from a known distribution and give the scrambled response to the interviewer who does not know the number generated by the respondent. [1] proposed an estimator of the population mean of quantitative sensitive variable in presence of non-response assuming that the people who refuse to answer on first attempt give scrambled response on second call. This estimator reduces non-response by protecting confidentiality of respondents but on the other hand variance increases due to use of scrambled response to non-respondents.
One possible way to estimate the population mean with maximum precision is to divide the whole population into certain internally homogeneous and externally heterogeneous groups, called strata and then selecting independent samples of different sizes from each stratum. [15] have suggested a method of optimum stratification after that many authors have

[*] Corresponding author e-mail: shakeelatish05@gmail.com

suggested different types of estimators using the auxiliary information in stratified random sampling (see [16], [19],[20] etc). In case of heterogeneous population, when non-response occurs in each stratum, [17] has proposed an estimator of population mean and also obtain the allocation of sample size in different strata for a fixed cost. [18] has suggested an estimator for population mean using post stratification after that [21] has proposed some estimators for population mean by using post stratification using the auxiliary information in presence of non-response. [22] has also constructed some separate generalized ratio type estimators for population mean in presence of non-response in stratified random sampling.
Taking motivation from [13] and [14], we propose an estimator for population mean of a sensitive quantitative variable using known coefficient of variation of the study variable under stratified random sampling. We also develop ratio-cum-product and ratio-in-regression estimators in stratified random sampling for improving efficiency of [1] estimator. In Section 4, we derive the expression for *MSE* and in Section 5 conditions, under which the proposed estimators are more efficient than the relevant estimators, are obtained. A numerical study is carried out in Section 6.

2 Notations

Consider a finite population $U = (U_1, U_2, U_3, \ldots.., U_N)$ of size N and is divided into L strata, each of size N_h for $(h = 1, 2, \ldots.L)$, such that $\sum_{h=1}^{L} N_h = N$. Let (y_{hi}, x_{hi}) be the observed values of (y, x) on the i^{th} unit of the h^{th} stratum, where $(h = 1, 2, \ldots.., L)$ and $(i = 1, 2, 3, \ldots., N_h)$. We select a sample of size n_h from the h^{th} stratum by using SRSWOR. When stratum mean of the auxiliary variable $\bar{X}_h$ is unknown then we use two-phase stratified random sampling scheme. In first phase, select a sample of size $\acute{n}_h (\acute{n}_h < N_h)$ from the h^{th} stratum by using SRSWOR to estimate $\bar{X}_h$ and in second phase, take a sub-sample of size $n_h (n_h < \acute{n}_h)$ from $\acute{n}_h$ selected units. Proportional allocation is used to allocate the sample size in different strata at both phases. Suppose that from n_h sampling units only n_{h1} units respond on first call and (n_{h2}) units don't respond. So we select a sub-sample of size $r_h = \frac{n_{h2}}{k_h} (k_h > 1)$ from n_{h2} non-responding units by making an extra effort. Consequently whole population is divided into two groups U_1 (respondents) and U_2 (non-respondents). Some more symbols are given below:
N_{h1} : Number of units in response group of the h^{th} stratum.
N_{h2} : Number of units in non-response group of the h^{th} stratum.
$P_h = \frac{N_h}{N}$: Stratum weight of the h^{th} stratum .
$f_h = \frac{n_h}{N_h}$: Sampling fraction of the h^{th} stratum.
$\bar{Y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi}$: Population mean of the study variable for the h^{th} stratum
$\bar{X}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} x_{hi}$: Population mean of the auxiliary variable for the h^{th} stratum
$S_{yh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$: Population variance of the study variable for the h^{th} stratum.
$S_{xh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2$: Population variance of the auxiliary variable for the h^{th} stratum.
$S_{yxh} = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)(x_{hi} - \bar{X}_h)$: Population covariance between the study variable and the auxiliary variable for the h^{th} stratum.
$S_{yh(2)}^2 = \frac{1}{N_{h2} - 1} \sum_{i=1}^{N_{h2}} (y_{hi} - \bar{Y}_{h2})^2$: Population variance of the study variable for non-response group in the h^{th} stratum.
$S_{xh(2)}^2 = \frac{1}{N_{h2} - 1} \sum_{i=1}^{N_{h2}} (x_{hi} - \bar{X}_{h2})^2$: Population variance of the auxiliary variable for non-response group in the h^{th} stratum.
$S_{yxh(2)} = \frac{1}{N_{h2} - 1} \sum_{i=1}^{N_{h2}} (y_{hi} - \bar{Y}_{h2})(x_{hi} - \bar{X}_{h2})$: Population covariance between the study variable and the auxiliary variable for non-response group in the h^{th} stratum.
$\bar{y}_{h1} = \frac{1}{n_{h1}} \sum_{i=1}^{n_{h1}} y_{hi}$: Sample mean of the study variable of units respond on first call in the h^{th} stratum.
$\bar{x}_{h1} = \frac{1}{n_{h1}} \sum_{i=1}^{n_{h1}} x_{hi}$: Sample mean of the auxiliary variable of units respond on first call in the h^{th} stratum.
$\acute{\bar{y}}_{h2} = \frac{1}{r_h} \sum_{i=1}^{r_h} y_{hi}$: Sample mean of the study variable of units respond on second call in the h^{th} stratum.
$\acute{\bar{x}}_{h2} = \frac{1}{r_h} \sum_{i=1}^{r_h} x_{hi}$: Sample mean of the auxiliary variable of units respond on second call in the h^{th} stratum.

3 The Estimators

Using [12] technique, the estimator for population mean in stratified sampling, is given by:

$$\bar{y}_{st}^{*} = \sum_{h=1}^{L} P_h \left(w_{h1} \bar{y}_{h1} + w_{h2} \acute{\bar{y}}_{h2} \right), \tag{1}$$

where $w_{h1} = \frac{n_{h1}}{n_h}$ and $w_{h2} = \frac{n_{h2}}{n_h}$. The variance of $\bar{y}^*_{st}$, after ignoring the correction factor $(1-f_h)$ for ease of computation,is given by

$$V(\bar{y}^*_{st}) = \sum_{h=1}^{L} P_h^2 \left[\frac{1}{n_h} S^2_{yh} + \frac{W_{h2}(k_h-1)}{n_h} S^2_{yh(2)}\right]. \tag{2}$$

Assuming that the coefficient of variation in each stratum is known, so Equation (1) becomes

$$\bar{y}^{**}_{st} = \sum_{h=1}^{L} P_h \bar{y}^{**}_h, \tag{3}$$

where $\bar{y}^{**}_h = a_h \bar{y}^*_h$ and the value of constant a_h for which MSE of $\bar{y}^{**}_{st}$ is minimum, is given by

$$a_{h(opt)} = \left[1 + \frac{1-f_h}{n_h} C^2_{yh} + \frac{W_{h2}(k_h-1)}{n_h} C^2_{yh(2)}\right]^{-1}.$$

Since $\frac{S^2_{yh}}{Y_h}$ and $\frac{S^2_{yh(2)}}{Y_h}$ don't differ significantly, so we may approximate $\frac{S^2_{yh(2)}}{Y_h} \cong \frac{S^2_{yh}}{Y_h} = C^2_{yh}$. The estimated value of a_h after ignoring the correction factor $(1-f_h)$ is given by :

$$\hat{a}_{h(opt)} = \left[1 + \frac{C^2_{yh}}{n_h}\left\{1 + \frac{n_{h2}}{n_h}(k_h-1)\right\}\right]^{-1}.$$

Now improved estimator becomes:

$$\bar{y}^{**}_{st} = \sum_{h=1}^{L} P_h \left[1 + \frac{C^2_{yh}}{n_h}\left\{1 + \frac{n_{h2}}{n_h}(k_h-1)\right\}\right]^{-1} \bar{y_h}^*.$$

The MSE of $\bar{y}^{**}_{st}$, is given by

$$MSE(\bar{y}^{**}_{st}) = \sum_{h=1}^{L} P_h^2 \left[(1-B_{1h})\frac{S^2_{yh}}{n_h} + (1-2B_{2h})\frac{W_{h2}(k_h-1)}{n_h} S^2_{yh(2)}\right], \tag{4}$$

where

$$B_{1h} = \frac{C^2_{yh}}{n_h}\left[1 - W^2_{h2}(k_h-1)^2\right] \text{ and } B_{2h} = \frac{C^2_{yh}}{n_h}\left[1 + W_{h2}(k_h-1)\right].$$

By Equations (2) and (4), we see that $MSE(\bar{y}^{**}_{st}) < V(\bar{y}^*_{st})$, if

$$\sum_{h=1}^{L} P_h^2 \frac{C^2_{yh} S^2_{yh}}{n_h^2}\left[1 + W_{h2}(k_h-1)\right]^2 > 0.$$

This indicates that $\bar{y}^{**}_{st}$ is always more efficient than $\bar{y}^*_{st}$.
Assuming (Y) as a quantitative sensitive variable, [1] have made some modifications by using RRT in [12] technique to sub-sample non-respondents, which is given below.
Let Z_h be the scrambled response in stratum h based on two independent scrambled random variables A_h and B_h which are unrelated to Y_h with known means (μ_{Ah}, μ_{Bh}) and variances $(\sigma^2_{Ah}, \sigma^2_{Bh})$ in the h^{th} stratum such that:

$$Z_h = A_h Y_h + B_h, \tag{5}$$

where $E_R(Z_h) = \mu_{Ah} Y_h + \mu_{Bh}$ with variance $V_R(Z_h) = \sigma^2_{Ah} Y_h^2 + \sigma^2_{Bh}$ for $(h = 1, 2, \ldots.., L)$. Here (E_R, V_R) are expectation and variance with respect to randomization device. Let $\hat{y}_{hi}$ be the transformed randomized response of the i^{th} unit in the h^{th} stratum, i.e

$$\hat{y}_{hi} = \frac{z_{hi} - \mu_{Bh}}{\mu_{Ah}},$$

where $E_R(\hat{y}_{hi}) = y_{hi}$ and the variance of $\hat{y}_{hi}$, is given by

$$V_R(\hat{y}_{hi}) = \frac{\sigma_{Ah}^2 Y_{hi}^2 + \sigma_{Bh}^2}{\mu_{Ah}^2} = \phi_{hi}. \tag{6}$$

We propose an estimator in stratified random sampling as:

$$\hat{\bar{y}}_{st}^{*} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^{*}, \tag{7}$$

where $\hat{\bar{y}}_h^{*} = w_{h1}\bar{y}_{h1} + w_{h2}\hat{\bar{y}}_{h2}$, $E(\hat{\bar{y}}_{st}^{*}) = \bar{Y}$ with variance

$$V(\hat{\bar{y}}_{st}^{*}) = \sum_{h=1}^{L} P_h^2 \left[\frac{1-f_h}{n_h} S_{yh}^2 + \frac{W_{h2}(k_h-1)}{n_h} S_{yh2}^2 + \frac{k_h}{n_h N_h} \sum_{i=1}^{N_{h2}} \phi_{hi} \right],$$

where

$$\frac{1}{N_{h2}} \sum_{i=1}^{N_{h2}} \phi_{hi} = \frac{\sigma_{Ah}^2 \mu_{yh(2)} + \sigma_{Ah}^2}{\mu_{Ah}^2}, \quad \mu_{yh(2)} = S_{yh(2)}^2 + \bar{Y}_{h2}^2.$$

There are two possible ways to obtain unknown $\mu_{yh(2)}$. One is to use a guess from previous work or pilot survey, otherwise sample estimate has to supply information about second moment keeping in mind its sensitive nature. After ignoring correction factor $(1-f_h)$, we have

$$V(\hat{\bar{y}}_{st}^{*}) = \sum_{h=1}^{L} P_h^2 \left[\frac{1}{n_h} S_{yh}^2 + \frac{W_{h2}(k_h-1)}{n_h} S_{yh(2)}^2 + \frac{k_h}{n_h N_h} S_{hr}^2 \right]. \tag{8}$$

where $S_{hr}^2 = \sum_{i=1}^{N_{h2}} \phi_{hi}$.
From (2) and (8), it is obvious that $\hat{\bar{y}}_{st}^{*}$ is less efficient than $\bar{y}_{st}^{*}$ but former gives greater privacy protection than later. In this paper our concern is to obtain efficiency of the estimators. For this purpose we utilize known coefficient of variation of study character to propose an estimator of finite population mean under stratified random sampling scheme, which gives more efficient results than the estimator proposed by [1].
The proposed estimator is:

$$\hat{\bar{y}}_{st}^{**} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^{**}, \tag{9}$$

where $\hat{\bar{y}}_h^{**} = k_{h1}\hat{\bar{y}}_h^{*}$. The optimum value of k_{h1} which minimize MSE of $\hat{\bar{y}}_{st}^{**}$, is given by

$$k_{h1(opt)} = \left[1 + \frac{1}{n_h} C_{yh}^2 + \frac{W_{h2}(k_h-1)}{n_h} \frac{S_{yh(2)}^2}{\bar{Y}_h^2} + \frac{k_h}{n_h N_h} \frac{S_{hr}^2}{\bar{Y}_h^2} \right]^{-1}.$$

As we discussed earlier $\frac{S_{yh}^2}{\bar{Y}_h^2}$ and $\frac{S_{yh(2)}^2}{\bar{Y}_h^2}$ don't differ significantly, so we may approximate
$\frac{S_{yh}^2}{\bar{Y}_h^2} \cong \frac{S_{yh(2)}^2}{\bar{Y}_h^2} \cong C_{yh}^2$. So estimated value of k_{h1} becomes

$$\hat{k}_{h1(opt)} = \left[1 + \frac{C_{yh}^2}{n_h} \left\{ 1 + \frac{n_{h2}}{n_h}(k_h-1) \right\} + \frac{k_h}{n_h N_h} \frac{S_{hr}^2}{\bar{Y}_h^2} \right]^{-1}.$$

Now proposed estmator for optimum value of k_{h1} becomes

$$\hat{\bar{y}}_{st}^{**} = \sum_{h=1}^{L} P_h \left[1 + \left\{ 1 + \frac{n_{h2}}{n_h}(k_h-1) \right\} \frac{C_{yh}^2}{n_h} + \frac{k_h}{n_h N_h} \frac{S_{hr}^2}{\bar{Y}_h^2} \right]^{-1} \hat{\bar{y}}_h^{*}. \tag{10}$$

The Bias and MSE of $\hat{\bar{y}}_{st}^{**}$ to first order approximation, are given by

$$Bias(\hat{\bar{y}}_{st}^{**}) \cong -\sum_{h=1}^{L} P_h \left[\left\{ 1 + W_{h2}(k_h - 1) \right\} \frac{C_{yh}^2}{n_h} + \frac{k_h}{n_h N_h} \frac{S_{hr}^2}{\bar{Y}_h^2} \right] \bar{Y}_h \tag{11}$$

and

$$MSE(\hat{\bar{y}}_{st}^{**}) \cong \sum_{h=1}^{L} P_h^2 \left[(1 - A_h^*) \frac{S_{yh}^2}{n_h} + (1 - 2B_h^*) \frac{W_{h2}(k_h - 1)}{n_h} S_{yh2}^2 + \frac{k_h}{n_h N_h} S_{hr}^2 \left\{ 1 - \frac{k_h}{n_h N_h} \frac{S_{hr}^2}{\bar{Y}_h^2} \right\} \right], \tag{12}$$

where

$$A_h^* = \frac{C_{yh}^2}{n_h} \{ 1 - W_{h2}^2 (k_h - 1)^2 \} + \frac{2k_h}{n_h N_h} \frac{S_{hr}^2}{\bar{Y}_h^2}$$

and $$B_h^* = \frac{C_{yh}^2}{n_h} \{ 1 + W_{h2}(k_h - 1) \} + \frac{k_h}{n_h N_h} \frac{S_{hr}^2}{\bar{Y}_h^2}.$$

By (8) and (12), assuming that $\frac{S_{yh}^2}{n_h} \cong \frac{S_{yh(2)}^2}{n_h}$, we see $MSE(\hat{\bar{y}}_{st}^{**}) < MSE(\hat{\bar{y}}_{st}^{*})$, if

$$\sum_{h=1}^{L} P_h^2 \frac{S_{yh}^2}{n_h} \left[\frac{C_h^2}{n_h} \left\{ 1 + W_{h2}(k_h - 1) \right\}^2 + \frac{2k_h}{n_h N_h} \frac{S_{hr}^2}{\bar{Y}_h^2} \left\{ 1 + W_{h2}(k_h - 1) \right\} + \left(\frac{k_h}{n_h N_h} \frac{S_{hr}^2}{\bar{Y}_h^2} \right)^2 \right] > 0.$$

This shows that $\hat{\bar{y}}_{st}^{**}$ is always more efficient than $\hat{\bar{y}}_{st}^{*}$.
The generalized ratio and regression type estimators using [14] estimator under stratified two-phase sampling scheme in case of complete and incomplete information on x_h are given by:

$$t_{st1} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^{**} \left(\frac{\bar{x}_h^*}{\overleftarrow{\bar{x}}_h} \right)^{a_{h1}}, \quad t_{st2} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^{**} \left(\frac{\bar{x}_h}{\overleftarrow{\bar{x}}_h} \right)^{a_{h2}},$$

$$t_{(lr)st1} = \sum_{h=1}^{L} P_h \left[\hat{\bar{y}}_h^{**} + b_h^{**} \left(\overleftarrow{\bar{x}}_h - \bar{x}_h^* \right) \right], \; t_{(lr)st2} = \sum_{h=1}^{L} P_h \left[\hat{\bar{y}}_h^{**} + b_h^{*} \left(\overleftarrow{\bar{x}}_h - \bar{x}_h \right) \right].$$

Here a_{h1} and a_{h2} are constants to be determined , $b_h^{**} = \frac{s_{yxh}^*}{s_{xh}^{*2}}$ and $b_h^* = \frac{s_{yxh}^*}{s_{xh}^2}$. Also s_{xh}^2 and s_{xh}^{*2} denote the estimates of S_{xh}^2 based on n_h and $(n_{h1} + r_h)$ observations respectively.
By putting $a_{h1} = a_{h2} = -1$ and $a_{h1} = a_{h2} = 1$ in t_{h1} and t_{h2} the estimators reduce to conventional and alternative stratified two phase ratio and product type estimators respectively, using scrambled response model to non-response group using coefficient of variation of the study character. The alternative stratified two-phase ratio type estimators are given by:

$$t_{st3} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^{**} \left(\frac{\overleftarrow{\bar{x}}_h}{\bar{x}_h^*} \right), \; t_{st4} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^{**} \left(\frac{\overleftarrow{\bar{x}}_h}{\bar{x}_h} \right), \; t_{st5} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^{**} \left(\frac{\bar{x}_h^*}{\overleftarrow{\bar{x}}_h} \right), \; t_{st6} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^{**} \left(\frac{\bar{x}_h}{\overleftarrow{\bar{x}}_h} \right).$$

Now putting $k_{h1} = 1, a_{h1} = a_{h2} = 1$ and $a_{h1} = a_{h2} = -1$, in t_{h1} , t_{h2}, $t_{(lr)h1}$ and $t_{(lr)h2}$, we get conventional and alternative stratified two-phase ratio, product and regression type estimators respectively, using scrambled response model to non-response group.

$$T_{st3} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^{*} \left(\frac{\overleftarrow{\bar{x}}_h}{\bar{x}_h^*} \right), \; T_{st4} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^{*} \left(\frac{\overleftarrow{\bar{x}}_h}{\bar{x}_h} \right), \; T_{st5} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^{*} \left(\frac{\bar{x}_h^*}{\overleftarrow{\bar{x}}_h} \right), \; T_{st6} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^{*} \left(\frac{\bar{x}_h}{\overleftarrow{\bar{x}}_h} \right),$$

$$T_{(lr)st1} = \sum_{h=1}^{L} P_h \left[\hat{\bar{y}}_h^{*} + b_h^{*} \left(\overleftarrow{\bar{x}}_h - \bar{x}_h^* \right) \right], \text{ and } T_{(lr)st2} = \sum_{h=1}^{L} P_h \left[\hat{\bar{y}}_h^{*} + b_h^{*} \left(\overleftarrow{\bar{x}}_h - \bar{x}_h \right) \right].$$

4 Some additional estimators

Taking motivation from [24] and [3], we introduce two additional generalized estimators as follow;

$$T_{rp(1)} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^* \left(\delta_{h1} \frac{\overleftarrow{\bar{x}}_h}{\bar{x}_h^*} + (1-\delta_{h1}) \frac{\bar{x}_h^*}{\overleftarrow{\bar{x}}_h} \right) \tag{13}$$

$$T_{rp(2)} = \sum_{h=1}^{L} P_h \hat{\bar{y}}_h^* \left(\delta_{h2} \frac{\overleftarrow{\bar{x}}_h}{\bar{x}_h} + (1-\delta_{h2}) \frac{\bar{x}_h}{\overleftarrow{\bar{x}}_h} \right) \tag{14}$$

$$T_{rl(1)} = \sum_{h=1}^{L} P_h \left[\hat{\bar{y}}_h^* \left(\frac{\overleftarrow{\bar{x}}_h}{\bar{x}_h^*} \right)^{\alpha_{h1}} + b_h^{**} \left(\overleftarrow{\bar{x}}_h - \bar{x}_h^* \right) \right] \tag{15}$$

$$T_{rl(2)} = \sum_{h=1}^{L} P_h \left[\hat{\bar{y}}_h^* \left(\frac{\overleftarrow{\bar{x}}_h}{\bar{\bar{x}}_h} \right)^{\alpha_{h2}} + b_h^{*} \left(\overleftarrow{\bar{x}}_h - \bar{x}_h \right) \right] \tag{16}$$

where δ_{h1}, δ_{h2}, α_{h1} and α_{h2} are constants. Putting $\delta_{h1} = \delta_{h2} = 1$ in Equations (13) and (14), we get T_{st3} and T_{st5} respectively. For $\delta_{h1} = \delta_{h2} = 0$ the estimators in (13) and (14) reduce to T_{st4} and T_{st6} respectively. Similarly for $b_h^* = 0$ putting $\alpha_{h1} = \alpha_{h2} = 1$ in Equation (15) and (16), we get T_{st3} and T_{st4} respectively. For $b_h^* = 0$, putting $\alpha_{h1} = \alpha_{h2} = -1$ in Equation (15) and (16), we get T_{st5} and T_{st6} respectively. For $\alpha_{h1} = \alpha_{h2} = 0$ the estimators in Equations (15) and (16) reduce to $T_{(lr)st1}$ and $T_{(lr)st2}$ respectively.

5 The Mean Squared Errors

In order to obtain the expressions for mean squared errors, we define:

$$\hat{e}_{0h}^* = \frac{\hat{\bar{y}}_h^* - \bar{Y}_h}{Y_h}, \quad e_{1h}^* = \frac{\bar{x}_h^* - \bar{X}_h}{X_h} \text{ and } \acute{e}_{1h} = \frac{\overleftarrow{\bar{x}}_h - \bar{X}_h}{X_h}$$

such that $E(\hat{e}_{0h}^*) = E(e_{1h}^*) = E(\acute{e}_{1h}) = 0$ and

$$E(\hat{e}_{0h}^{*2}) = \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \frac{S_{yh}^2}{\bar{Y}_h^2} + \frac{W_{h2}(k_h-1)}{n_h} \frac{S_{yh(2)}^2}{\bar{Y}_h^2} + \frac{k_h}{n_h N_h} \frac{\sum_{i=1}^{N_{h2}} \phi_{hi}}{\bar{Y}_h^2}$$

$$E(e_{1h}^{*2}) = \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \frac{S_{xh}^2}{\bar{X}_h^2} + \frac{W_{h2}(k_h-1)}{n_h} \frac{S_{xh(2)}^2}{\bar{X}_h^2}, \; E(\acute{e}_{1h}^2) = \left(\frac{1}{\acute{n}_h} - \frac{1}{N_h} \right) \frac{S_{xh}^2}{\bar{X}_h^2},$$

$$E(e_{1h}^* \acute{e}_{1h}) = \left(\frac{1}{\acute{n}_h} - \frac{1}{N_h} \right) \frac{S_{xh}^2}{\bar{X}_h^2}, \; E(\hat{e}_{0h}^* \acute{e}_{1h}) = \left(\frac{1}{\acute{n}_h} - \frac{1}{N_h} \right) \frac{S_{yxh}^2}{\bar{X}_h \bar{Y}_h},$$

$$E(\hat{e}_{0h}^* e_{1h}^*) = \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \frac{S_{yxh}^2}{\bar{X}_h \bar{Y}_h} + \frac{W_{h2}(k_h-1)}{n_h} \frac{S_{yxh(2)}^2}{\bar{X}_h \bar{Y}_h}.$$

Consider the estimator t_{h1} in term of errors:

$$t_{h1} = \hat{\bar{y}}_h^{**} \left(\frac{\bar{x}_h^*}{\overleftarrow{\bar{x}}_h} \right)^{a_{h1}} = k_{h1} \hat{\bar{y}}_h^* \left(\frac{\bar{x}_h^*}{\overleftarrow{\bar{x}}_h} \right)^{a_{h1}}$$

$$= k_{h1} \bar{Y}_h (1 + \hat{e}_{0h}^*) \left[\frac{(1+e_{1h}^*)}{(1+\acute{e}_{1h})} \right]^{a_{h1}}$$

$$t_{h1} - \bar{Y}_h = (k_{h1}-1)\bar{Y}_h + k_{h1}\bar{Y}_h \left[\hat{e}_{0h}^* + a_{h1} e_{1h}^* - a_{h1} \acute{e}_{1h} + \frac{a_{h1}(a_{h1}-1)}{2} e_{1h}^{*2} \right.$$

$$\left. + \frac{a_{h1}(a_{h1}+1)}{2} \acute{e}_{1h}^2 + a_{h1} e_{1h}^* \hat{e}_{0h}^* - a_{h1} \acute{e}_{1h} \hat{e}_{0h}^* - a_{h1}^2 \acute{e}_{1h} \hat{e}_{1h}^* \right].$$

Squaring, neglecting higher order terms and taking expectation, we get MSE of t_{h1}

$$MSE(t_{h1}) \cong (k_{h1}-1)^2\bar{Y}_h^2 + k_{h1}^2 V(\hat{\bar{y}}_h^*) + \bar{Y}_h^2\Bigg[k_{h1}a_{h1}(2k_{h1}a_{h1} - k_{h1} - a_{h1} + 1)\left\{\left(\frac{1}{n_h}-\frac{1}{\acute{n}_h}\right)C_{xh}^2 + \frac{W_{h2}(k_h-1)}{n_h}C_{xh(2)}^2\right\} + 2k_{h1}a_{h1}(2k_{h1}-1)\left\{\left(\frac{1}{n_h}-\frac{1}{\acute{n}_h}\right)C_{yxh} + \frac{W_{h2}(k_h-1)}{n_h}C_{yxh(2)}\right\}\Bigg].$$

Expanding k_{h1} and neglecting higher order terms, the optimum value of k_{h1}, is given by

$$k_{h1(opt)} = 1 - \left[\frac{C_{yh}^2}{n_h}\left\{1 + W_{h2}(k_h-1)\right\} + \frac{k_h}{n_h N_h}\frac{S_{hr}^2}{\bar{Y}_h^2}\right],$$

where $k_{h1} \cong 1 - B_h^*$ and $B_h^* = \frac{C_{yh}^2}{n_h}\{1+W_{h2}(k_h-1)\} + \frac{k_h}{n_h N_h}\frac{S_{hr}^2}{\bar{Y}_h^2}$.
Similarly $k_{h1}^2 \cong 1 - 2\left[\frac{C_{yh}^2}{n_h}\left\{1+W_{h2}(k_h-1)\right\} + \frac{k_h}{n_h N_h}\frac{S_{hr}^2}{\bar{Y}_h^2}\right] \cong (1-2B_h^*)$ and so on.
Substituting these results in $MSE(t_{h1})$, we get MSE of t_{st1}:

$$MSE(t_{st1}) \cong \sum_{h=1}^{L} P_h^2\Bigg[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + a_{h1}\left\{a_{h1} - (3a_{h1}-1)B_h^*\right\}R_h^2\left\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\right\} + 2a_{h1}(1-3B_h^*)R_h\left\{A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\right\}\Bigg], \tag{17}$$

where

$$A_{h1} = \left(\frac{1}{n_h} - \frac{1}{\acute{n}_h}\right),\ A_{h3} = \frac{W_{h2}(k_h-1)}{n_h} \text{ and } R_h = \frac{\bar{Y}_h}{\bar{X}_h}.$$

Similarly we obtain $MSE(t_{st2})$ as ;

$$MSE(t_{st2}) \cong \sum_{h=1}^{L} P_h^2\Bigg[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + A_{h1}a_{h2}\left\{(a_{h2} - (3a_{h2}-1)B_h^*)R_h^2 S_{xh}^2 + 2R_h(1-3B_h^*)S_{yxh}\right\}\Bigg]. \tag{18}$$

The optimum values of a_{h1} and a_{h2} are $a_{h1(opt)} = -\left[\frac{B_h^*}{2(1-3B_h^*)} + \frac{A_{h1}S_{yxh}+A_{h3}S_{yxh(2)}}{R_h\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\}}\right]$
and $a_{h2(opt)} = -\left[\frac{B_h^*}{2(1-3B_h^*)} + \frac{S_{yxh}}{R_h S_{xh}^2}\right]$ respectively. The MSE of ratio-cum-product type estimator T_{rp} and generalized ratio in regression estimator are given by

$$MSE(T_{rp(1)}) \cong \sum_{h=1}^{L} P_h^2\Bigg[V(\hat{\bar{y}}_h^*) + (1-2\delta_h)^2 R_h^2\left\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\right\} + 2(1-2\delta_h)R_h\left\{A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\right\}\Bigg], \tag{19}$$

$$MSE(T_{rp(2)}) \cong \sum_{h=1}^{L} P_h^2\Bigg[V(\hat{\bar{y}}_h^*) + A_{h1}\left\{(1-2\delta_h)^2 R_h^2 S_{xh}^2 + 2(1-2\delta_h)R_h S_{yxh}\right\}\Bigg], \tag{20}$$

where $\delta_{h1(opt)} = \frac{1}{2} + \frac{A_{h1}S_{yxh}+A_{h3}S_{yxh(2)}}{R_h\{A_{h1}S_{xh}^2+A_{h3}S_{xh(2)}^2\}}$ and $\delta_{h2(opt)} = \frac{1}{2} + \frac{S_{yxh}}{R_h\{S_{xh}^2\}}$. Also

$$MSE(T_{rl(1)}) \cong \sum_{h=1}^{L} P_h^2 \left[V(\hat{\bar{y}}_h^*) + \left(\alpha_{h1}\bar{Y} + b_h^{**}\bar{X}\right)^2 \left\{ A_{h1}C_{xh}^2 + A_{h3}C_{xh(2)}^2 \right\} - 2\left(\alpha_{h1}\bar{Y} + b_h^{**}\bar{X}\right)\bar{Y} \left\{ A_{h1}C_{yxh} + A_{h3}C_{yxh(2)} \right\} \right], \tag{21}$$

$$MSE(T_{rl(2)}) \cong \sum_{h=1}^{L} P_h^2 \left[V(\hat{\bar{y}}_h^*) + A_{h1} \left\{ \left(\alpha_{h2}\bar{Y} + b_h^{*}\bar{X}\right)^2 C_{xh}^2 - 2\left(\alpha_{h2}\bar{Y} + b_h^{*}\bar{X}\right)\bar{Y}C_{yxh} \right\} \right], \tag{22}$$

with $\alpha_{h1(opt)} = \frac{A_{h1}C_{yxh}+A_{h3}C_{yxh(2)}}{A_{h1}C_{xh}^2+A_{h3}C_{xh(2)}^2} - \frac{b_h^{**}}{R_h}$ and $\alpha_{h2(opt)} = \frac{C_{yxh}}{C_{xh}^2} - \frac{b_h^*}{R_h}$. Also putting optimum values of δ_{h1} and α_{h1} in MSE of $T_{rp(1)}$ and $T_{rl(1)}$, we get

$$MSE(T_{rp(1)(min)}) = MSE(T_{rl(1)(min)}) \cong \sum_{h=1}^{L} P_h^2 \left[V(\hat{\bar{y}}_h^*) - \frac{\left\{A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\right\}^2}{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2} \right] \tag{23}$$

and putting optimum values of δ_{h2} and α_{h2} in MSE of $T_{rp(2)}$ and $T_{rl(2)}$, we get

$$MSE(T_{rp(2)(min)}) = MSE(T_{rl(2)(min)}) \cong \sum_{h=1}^{L} P_h^2 \left[V(\hat{\bar{y}}_h^*) - \frac{A_{h1}S_{yxh}^2}{S_{xh}^2} \right] \tag{24}$$

The mean square errors of regression type estimator and different members of generalized estimators for both cases are given in table 1. The MSE of t_{st3} and t_{st4} are obtained by putting $a_{h1} = a_{h2} = -1$ in (17) and (18) respectively, the MSE of t_{st5} and t_{st6} are obtained by putting $a_{h1} = a_{h2} = 1$ in (17) and (18) and the MSE of T_{st3}, T_{st4}, T_{st5} and T_{st6}, are obtained by putting $a_{h1} = a_{h2} = -1$ and $a_{h1} = a_{h2} = 1$ with $k_{h1} = 1$ in Equations (17) and (18).

Table 1: The Mean Squared Errors of proposed Estimators.

Table 1: The Mean Squared Errors of proposed Estimators

Estimators	Mean Squared Errors
t_{st3}	$\sum_{h=1}^{L} P_h^2\Big[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + (1-4B_h^*)R_h^2\Big\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\Big\}$ $-2(1-3B_h^*)R_h\Big\{A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\Big\}\Big]$
t_{st4}	$\sum_{h=1}^{L} P_h^2\Big[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + A_{h1}\Big\{(1-4B_h^*)R_h^2S_{xh}^2 - 2(1-3B_h^*)R_hS_{yxh}\Big\}\Big].$
t_{st5}	$\sum_{h=1}^{L} P_h^2\Big[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + (1-2B_h^*)R_h^2\Big\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\Big\}$ $+2(1-3B_h^*)R_h\Big\{A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\Big\}\Big]$
t_{st6}	$\sum_{h=1}^{L} P_h^2\Big[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + A_{h1}\Big\{(1-2B_h^*)R_h^2S_{xh}^2 + 2(1-3B_h^*)R_hS_{yxh}\Big\}\Big].$
$t_{(lr)st1}$	$\sum_{h=1}^{L} P_h^2\Big[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + \beta_h^2\Big\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\Big\}$ $-2\beta_h(1-B_h^*)\Big\{(A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\Big\}\Big]$
$t_{(lr)st2}$	$\sum_{h=1}^{L} P_h^2\Big[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + A_{h1}\Big\{\beta_h^2S_{xh}^2 - 2\beta_h(1-B_h^*)S_{yxh}\Big\}\Big].$
T_{st3}	$\sum_{h=1}^{L} P_h^2\Big[V(\hat{\bar{y}}_h^*) + R_h^2\Big\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\Big\}$ $-2R_h\Big\{A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\Big\}\Big],$
T_{st4}	$\sum_{h=1}^{L} P_h^2\Big[(1-2B_h^*)V(\hat{\bar{y}}_h) + A_{h1}\Big\{R_h^2S_{xh}^2 - 2R_hS_{yxh}\Big\}\Big],$
T_{st5}	$\sum_{h=1}^{L} P_h^2\Big[V(\hat{\bar{y}}_h^*) + R_h^2\Big\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\Big\}$
	$+2R_h\Big\{A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\Big\}\Big],$
T_{st6}	$\sum_{h=1}^{L} P_h^2\big[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + A_{h1}\{R_h^2S_{xh}^2 + 2R_hS_{yxh}\}\big],$
$T_{(lr)st1}$	$\sum_{h=1}^{L} P_h^2\Big[V(\hat{\bar{y}}_h^*) + \beta_h^2\Big\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\Big\}$ $-2\beta_h\Big\{(A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\Big\}\Big]$
$T_{(lr)st2}$	$\sum_{h=1}^{L} P_h^2\Big[V(\hat{\bar{y}}_h^*) + A_{h1}\Big\{\beta_h^2S_{xh}^2 - 2\beta_h(1-B_h^*)S_{yxh}\Big\}\Big]$

6 Efficiency comparison

The conditions under which the proposed estimators are more efficient than the existing estimators are given below

Table 2: Conditions.

	Estimators	Conditions
i	$MSE(t_{st1}) < V(\hat{\bar{y}}_h^*)$ if	$\sum_{h=1}^{L} P_h^2 \Big[-2B_h^* V(\hat{\bar{y}}_h^*) + a_{h1} \Big\{ a_{h1} - (3a_{h1} - 1)B_h^* \Big\} R_h^2 \Big\{ A_{h1} S_{xh}^2 + A_{h3} S_{xh(2)}^2 \Big\} + 2a_{h1}(1 - 3B_h^*) R_h \Big\{ A_{h1} S_{yxh} + A_{h3} S_{yxh(2)} \Big\} \Big] < 0.$
ii	$MSE(t_{st2}) < V(\hat{\bar{y}}_h^*)$ if	$\sum_{h=1}^{L} P_h^2 \Big[-2B_h^* V(\hat{\bar{y}}_h^*) + A_{h1} a_{h2} \Big[\Big\{ a_{h2} - (3a_{h2} - 1)B_h^* \Big\} R_h^2 S_{xh}^2 + 2a_{h2}(1 - 3B_h^*) R_h S_{yxh} \Big\} < 0.$
iii	$MSE(t_{st3}) < MSE(T_{st3})$ if	$\sum_{h=1}^{L} P_h^2 \Big[-B_h^* V(\hat{\bar{y}}_h^*) - 2B_h^* R_h^2 \Big\{ A_{h1} S_{xh}^2 + A_{h3} S_{xh(2)}^2 \Big\} + 3\mathrm{B}_h^* R_h \Big\{ A_{h1} S_{yxh} + A_{h3} S_{yxh(2)} \Big\} < 0.$
iv	$MSE(t_{st4}) < MSE(T_{st4})$ if	$\sum_{h=1}^{L} P_h^2 \Big[-B_h^* V(\hat{\bar{y}}_h^*) - A_{h1} B_h^* \Big\{ 2R_h^2 S_{xh}^2 - 3\mathrm{B}_h^* R_h S_{yxh} \Big\} < 0.$
v	$MSE(t_{st5}) < MSE(T_{st5})$ if	$-\sum_{h=1}^{L} P_h^2 B_h^* \Big[V(\hat{\bar{y}}_h^*) + R_h^2 \Big\{ A_{h1} S_{xh}^2 + A_{h3} S_{xh(2)}^2 \Big\} + 3\mathrm{R}_h \Big\{ A_{h1} S_{yxh} + A_{h3} S_{yxh(2)} \Big\} < 0.$
vi	$MSE(t_{st6}) < MSE(T_{st6})$ if	$-\sum_{h=1}^{L} P_h^2 B_h^* \Big[V(\hat{\bar{y}}_h^*) + A_{h1} \Big\{ R_h^2 S_{xh}^2 + 3R_h S_{yxh} \Big\} \Big] < 0.$
vii	$MSE(t_{st1}) < MSE(t_{st2})$ if	$\sum_{h=1}^{L} P_h^2 \Big[\Big\{ a_{h1} \Big(a_{h1} - (3a_{h1} - 1)B_h^* \Big) - a_{h2} \Big(a_{h2} - (3a_{h2} - 1)B_h^* \Big) \Big\} R_h^2 \Big\{ A_{h1} S_{xh}^2 + A_{h3} S_{xh(2)}^2 \Big\} + 2(a_{h1} - a_{h2})(1 - 3B_h^*) R_h \Big\{ A_{h1} S_{yxh} + A_{h3} S_{yxh(2)} \Big\} \Big] < 0.$
viii	$MSE(t_{st1}) < MSE(t_{st3})$ if	$\sum_{h=1}^{L} P_h^2 \Big[\Big\{ a_{h1} \Big(a_{h1} - (3a_{h1} - 1)B_h^* \Big) - (1 - 4B_h^*) \Big\} R_h^2 \Big\{ A_{h1} S_{xh}^2 + A_{h3} S_{xh(2)}^2 \Big\} + 2(a_{h1} + 1)(1 - 3B_h^*) R_h \Big\{ A_{h1} S_{yxh} + A_{h3} S_{yxh(2)} \Big\} \Big] < 0.$
ix	$MSE(t_{st2}) < MSE(t_{st4})$ if	$\sum_{h=1}^{L} P_h^2 A_{h1} \Big[\Big\{ a_{h2} \Big(a_{h2} - (3a_{h2} - 1)B_h^* \Big) - (1 - 4B_h^*) \Big\} R_h^2 S_{xh}^2 + 2(a_{h2} + 1)(1 - 3B_h^*) R_h S_{yxh} \Big] < 0.$
x	$MSE(t_{st1}) < MSE(t_{st5})$ if	$\sum_{h=1}^{L} P_h^2 \Big[\Big\{ a_{h1} \Big(a_{h1} - (3a_{h1} - 1)B_h^* \Big) - (1 - 2B_h^*) \Big\} R_h^2 \Big\{ A_{h1} S_{xh}^2 + A_{h3} S_{xh(2)}^2 \Big\} + 2(a_{h1} - 1)(1 - 3B_h^*) R_h \Big\{ A_{h1} S_{yxh} + A_{h3} S_{yxh(2)} \Big\} \Big] < 0.$
xi	$MSE(t_{st2}) < MSE(t_{st6})$ if	$\sum_{h=1}^{L} P_h^2 A_{h1} \Big[\Big\{ a_{h1} \Big(a_{h2} - (3a_{h2} - 1)B_h^* \Big) - (1 - 2B_h^*) \Big\} R_h^2 S_{xh}^2 + 2(a_{h2} - 1)(1 - 3B_h^*) R_h S_{yxh} < 0.$
xii	$MSE(t_{(lr)st1}) < MSE(T_{(lr)st1})$ if	$\sum_{h=1}^{L} P_h^2 \Big[-V(\hat{\bar{y}}_h^*) + \beta_h \Big\{ A_{h1} S_{yxh} + A_{h3} S_{yxh(2)} \Big\} \Big] < 0.$
xiii	$MSE(t_{(lr)st2}) < MSE(T_{(lr)st2})$ if	$\sum_{h=1}^{L} P_h^2 \Big[-V(\hat{\bar{y}}_h^*) + \beta_h A_{h1} S_{yxh} \Big] < 0.$

7 Numerical Study

We use the following data sets for efficiency comparison.
Population 1 ([25])
Y: County-wise number of Non-employer establishment .
X: County-wise number of Non-farm establishment.
Data for private non-farm establishments with non employees for the U.S. Puerto Rico and the Island Areas are published in County Business Patterns (CBP). Basic data items are extracted from the Business Register, a file of all known single and multi-establishment companies maintained and updated by the Bureau of the Census from various Census Bureau programs, such as Bureau of Labor Statistics, and the Social Security Administration. We take five states of USA (Kansas, Iowa, Kentucky, Indiana, Illnois), having different number of counties [counties are taken as population units], as strata. Assuming different non-response rate in different strata. Information for all strata are given in Table 1.

Table 3: Summary Statistics for Data 1:

h	N_h	W_{2h}	$\acute{n}_h$	n_h	$\bar{Y}_h$	$\bar{X}_h$	$\bar{Y}_{2h}$	$\bar{X}_{2h}$	S_{yh}	S_{xh}
1	105	0.14	61	20	1043.5	407.80	441.20	174.13	1394.4	566.86
2	98	0.18	57	19	1767.7	700.30	2293.1	925.33	2027.7	860.60
3	120	0.24	70	23	2293.6	745.73	1727.6	520.17	4992.3	1934.7
4	93	0.20	54	18	3585.1	1313.7	2545.2	1059.5	4772.7	1782.4
5	98	0.22	57	19	3141.6	1205.4	3910.3	1508.7	4772.2	1862.0
h	S_{y2h}	S_{x2h}	S_{yxh}	$\sigma_{yx(2)h}$	ρ_h	ρ_{2h}	μ_{Ah}	μ_{Bh}	σ^2_{Ah}	σ^2_{Bh}
1	328.38	119.07	760247.6	37017.4	0.962	0.9467	0.520	0.48	0.0744	0.076
2	2435.3	1106.1	1722564	2648316	0.987	0.983	0.540	0.490	0.0882	0.085
3	1706.6	582.21	9620257	966310.2	0.996	0.972	0.489	0.480	0.0876	0.083
4	2523.6	1267.0	8395327	3124659	0.9868	0.977	0.480	0.476	0.0761	0.074
5	4885.5	1895.4	2823171	9192888	0.993	0.993	0.510	0.550	0.071	0.084

We obtain PRE's of different estimators in Table 4.

Table 4: PRE of diferrent estimators using Data 1:

h	N_h	W_{2h}	$\acute{n}_h$	n_h	$\bar{Y}_h$	$\bar{X}_h$	$\bar{Y}_{2h}$	$\bar{X}_{2h}$	S_{yh}	S_{xh}
1	105	0.14	61	20	1043.5	407.80	441.20	174.13	1394.4	566.86
2	98	0.18	57	19	1767.7	700.30	2293.1	925.33	2027.7	860.60
3	120	0.24	70	23	2293.6	745.73	1727.6	520.17	4992.3	1934.7
4	93	0.20	54	18	3585.1	1313.7	2545.2	1059.5	4772.7	1782.4
5	98	0.22	57	19	3141.6	1205.4	3910.3	1508.7	4772.2	1862.0
h	S_{y2h}	S_{x2h}	S_{yxh}	$\sigma_{yx(2)h}$	ρ_h	ρ_{2h}	μ_{Ah}	μ_{Bh}	σ^2_{Ah}	σ^2_{Bh}
1	328.38	119.07	760247.6	37017.4	0.962	0.9467	0.520	0.48	0.0744	0.076
2	2435.3	1106.1	1722564	2648316	0.987	0.983	0.540	0.490	0.0882	0.085
3	1706.6	582.21	9620257	966310.2	0.996	0.972	0.489	0.480	0.0876	0.083
4	2523.6	1267.0	8395327	3124659	0.9868	0.977	0.480	0.476	0.0761	0.074
5	4885.5	1895.4	2823171	9192888	0.993	0.993	0.510	0.550	0.071	0.084

In second row of Table 4 PRE of [14] estimator is calculated w.r.t $\bar{y}^*_{st}$ for different values of k_h, where (h=1,2,3,4,5). The rest of table 4 contains PRE of different generalized-type estimators w.r.t $\hat{\bar{y}}^*_{st}$ in stratified sampling. Table 4 shows that the generalized ratio and regression estimators perform better for all combination of k. It can be inferred from Table 4 the PRE tend to increase as we move from left to right of the Table 4, this shows that for larger values k_h PRE tends increase. The numerical values of conditions using Data 1 are obtain in Table 5.

Table 5: Condition values using Data 1:

Conditions	$k_h(h = 1, ..., 4.)$ *2,2,1.5,1.5,1.5*	$k_h(h = 1, ..., 4.)$ *1.5,1.5,1.5,2,2*	$k_h(h = 1, ..., 4.)$ *2.5,2.5,2,2,2*	$k_h(h = 1, ..., 4.)$ *3,3,2,2,2*
(i)	$-69271.7 < 0$	$-73207.2 < 0$	$-76769.7 < 0$	$-77448.48 < 0$
(ii)	$-66094.5 < 0$	$-68484.4 < 0$	$-71048.6 < 0$	$-71366.89 < 0$
(iii)	$-14493.5 < 0$	$-15858.5 < 0$	$-17291.7 < 0$	$-17486.06 < 0$
(iv)	$-15621.7 < 0$	$-17919.3 < 0$	$-19748.9 < 0$	$-20066.33 < 0$
(v)	$-8.93646 < 0$	$-23.9570 < 0$	$-27.6459 < 0$	$-29.76850 < 0$
(viii)	$-1254.48 < 0$	$-1511.20 < 0$	$-3062.41 < 0$	$-3061.005 < 0$
(ix)	$-1205.59 < 0$	$-1390.06 < 0$	$-2859.56 < 0$	$-2861.156 < 0$
(xii)	$-7201.84 < 0$	$-7931.53 < 0$	$-8516.81 < 0$	$-8617.698 < 0$
(xiii)	$-4538.35 < 0$	$-3114.98 < 0$	$-2766.72 < 0$	$-2529.706 < 0$

Population 2([26])

Y: Price of Diamond

X: Depth of diamond.

We use data listed in [26] pertaining to 235 stones which educates us on the relative pricing of diamonds and the different grades of its clarity and weight. The Data consist of diamond stones [population units], divided into four strata according to its clarity i.e. IF (internally flawless), VVS1 (very very slightly imperfect), VVS2(very very slightly imperfect type 2) and VS1(very slightly imperfects). Numbers of diamond stones in each group are different. Assuming a fix non-response rate in each strata information for all strata are given in Table 6.

Table 6: Summary Statistics for Data 2:

h	N_h	W_{2h}	$\acute{n}_h$	n_h	$\bar{Y}_h$	$\bar{X}_h$	$\bar{Y}_{2h}$	$\bar{X}_{2h}$	S_{yh}	S_{xh}
1	100	0.30	64	21	36.23	66.44	36.01	62.84	45.16	23.06
2	44	0.20	28	9	42	65.25	39.03	61.48	152.94	41.24
3	71	0.25	45	15	40.57	65.55	49.64	61.04	204.6	21.6
4	20	0.20	13	5	47.63	62.87	47.48	62.82	452.82	9.31
h	$S_{yh(2)}$	$\S_{x2h}$	S_{yxh}	$S_{yxh(2)}$	ρ_h	ρ_{2h}	μ_{Ah}	μ_{Bh}	σ^2_{Ah}	σ^2_{Bh}
1	84.02	33.02	-9.15	-15.84	-0.283	-0.174	0.508	0.510	0.0857	0.091
2	116.3	10.53	-20.23	-1.42	-0.255	-0.578	0.457	0.490	0.071	0.093
3	488.82	5.84	-17.99	-12.77	-0.270	-0.337	0.420	0.557	0.081	0.088
4	732.2	1.77	-15.61	-0.171	-0.240	-0.433	0.530	0.520	0.045	0.071

Table 7: PRE of different estimators using Data 2:

Estimators	2.5,2.5,2,2	1.5,1.5,2,2	2.5,2.5,3,3	2.5,3,3,3.5	k_h 4, 4, 3.5 ,3.5
$\bar{y}^*_{st}$	100	100.00	100.00	100.00	100.00
$\bar{y}^{**}_{st}$	102.26	102.36	102.85	103.15	103.01
$\hat{\bar{y}}^*_{st}$	100.00	100.00	100.00	100.00	100.00
$\hat{\bar{y}}^{**}_{st}$	103.73	104.62	104.96	107.96	108.23
t_{1st}	108.83	110.77	110.92	115.60	117.81
t_{2st}	108.38	110.45	110.37	115.08	117.16
t_{5st}	108.75	110.67	110.85	115.56	117.81
t_{6st}	108.33	110.36	110.32	115.06	117.16
T_{5st}	101.35	101.42	101.31	101.15	101.09
T_{6st}	101.00	101.13	100.83	100.75	100.58
$t_{lr1(st)}$	108.97	110.94	111.10	115.79	118.02
$t_{lr2(st)}$	108.49	110.56	110.47	115.2	117.28
$T_{lr1(st)}$	101.54	101.64	101.50	101.30	101.25
$T_{lr2(st)}$	101.12	101.31	101.00	100.85	100.67
T_{rp1} & T_{lr1}	101.5	101.69	101.38	101.33	101.41
T_{rp2} & T_{lr2}	101.07	101.34	100.86	100.83	100.11

The PRE in first two rows are calculate w.r.t $\bar{y}^*_{st}$ and the remaining are calculated w.r.t $\hat{\bar{y}}^*_{st}$. Table 7 confirm that the generalized ratio and regression type estimators perform better in case of negative correlation between the study and auxiliary variable. The numerical values of conditions are obtain in Table 8.

Table 8: Condition values using Data 2:

Conditions	k_h 2.5,2.5,2,2	k_h 1.5,1.5,2,2	k_h 2.5,2.5,3,3	k_h 2.5,3,3,3.5	k_h 4,4,3.5,3.5
(i)	$-0.48633 < 0$	$-0.42449 < 0$	$-0.76548 < 0$	$-0.85737 < 0$	$-1.36060 < 0$
(ii)	$-0.46155 < 0$	$-0.40821 < 0$	$-0.72852 < 0$	$-0.82065 < 0$	$-1.30678 < 0$
(v)	$-0.39399 < 0$	$-0.34143 < 0$	$-0.66004 < 0$	$-0.75248 < 0$	$-1.23515 < 0$
(vi)	$-0.39584 < 0$	$-0.34252 < 0$	$-0.66346 < 0$	$-0.75585 < 0$	$-1.24241 < 0$
(vii)	$-0.00197 < 0$	$-0.00150 < 0$	$-0.00556 < 0$	$-0.00573 < 0$	$-0.00861 < 0$
(x)	$-0.02148 < 0$	$-0.02010 < 0$	$-0.02745 < 0$	$-0.02743 < 0$	$-0.03162 < 0$
(xi)	$-0.01422 < 0$	$-0.01419 < 0$	$-0.01291 < 0$	$-0.01239 < 0$	$-0.01154 < 0$
(xii)	$-0.39120 < 0$	$-0.33847 < 0$	$-0.65618 < 0$	$-0.74815 < 0$	$-1.23046 < 0$
(xiii)	$-0.39294 < 0$	$-0.33956 < 0$	$-0.65955 < 0$	$-0.75155 < 0$	$-1.23747 < 0$

8 Conclusion

In this paper, we proposed an estimator in stratified random sampling under proportional allocation using [1] estimator. To improve efficiency of the estimators, we use co-efficient of variation of the study character assuming that it is known in all strata. On the basis of proposed estimators, we developed some generalized ratio, regression, ratio-cum-product and ratio in regression estimator. Table 4 and Table 7 show that for fixed sample sizes $(\acute{n}_h, n_h)$ in all strata, the estimator $\hat{\bar{y}}_{st}^{**}$ and all the other generalized estimators, constructed using proposed estimator, are more efficient than [1] estimator $\hat{\bar{y}}_{st}^{*}$ for all combination of k_h considered in data set 1 and 2. Further generalized ratio type estimators t_{1st}, t_{2st} are more efficient than other estimators suggested in this study. It can be concluded by comparing Table 4 and 7 that efficiency of generalized estimators are higher when there is a positive correlation between the study variable and the auxiliary variable. In this paper we have discussed only linear scrambled response model used by [2] because our concern is only in improvement of efficiency of the estimators for a fixed level of privacy protection. We can also use different models to increase privacy protection as well. Future work can be done to improve privacy protection by using different scrambled response models using known coefficient of variation of the sensitive character in stratified random sampling.

Acknowledgement

The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

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

[2] G. Diana, P.F Perri , New scrambled response model for estimating the mean of a sensitive quantitative character, Journal of Applied Statistics **37**, 1875-1890 (2010).

[3] B.B Khare,, H. U Rehman, A generalized ratio in regression type estimator for population mean using auxiliary variables in the presence of non-response, International Journal of Statistics Economics **15(3)**, 64-68 (2014).

[4] B. Eichhorn, L.S Hayre, Scrambled randomized response method for obtaining sensitive quantitative data, Journal of Statistical Planning and Inference **7**, 307-316 (1983).

[5] K.H Pollock, Y. Bek, A comparison of three randomized response models for quantitative data, Journal of The American Statistical Association **71**, 884-886 (1976).

[6] B.G Greenberg, R. R Kuebler, Damuth, J. R Abernathy, D. G Horvitz, Application of Randomized response technique in obtaining quantitative data, Jounal of The Ameracan Statistical Association **99**, 243-250 (1971).

[7] S.L Warner, Randomized response: A survey technique for eliminating evasive answer bias, Journal of the American Statistical Association **60**, 63-69 (1965).

[8] H.P Singh, S. Kumar, Improved estimation of population mean under two phase sampling with sub-sampling the non-respondents, Journal of Statistical Planning and Inference **140**, 2536-2550 (2010).

[9] B.B Khare, S. Srivastava, Generalized two phase sampling estimators for the population mean in presence of non-response, Aligarh Journal of Statistics **30**, 39-54 (2010).

[10] B.B Khare, R.R Sinha, On class of estimator for population mean using Multi-auxilary characters in presence of non-response, Statistics in Transition-New Series. **10(1)**, 3-14 (2009).

[11] B.B Khare, S. Srivastava, Estimation of population mean using auxiliary charactersin presence of non response, National Academy of Science letters-India, **16(3)**, 111-114. (1993).
[12] M. H Hansen, W.N Hurwitz, The problem of non response in sample surveys, Journal of The American Statistical Association **41**, 517-529 (1946).
[13] D.T Searls, The utilization of a known coefficient of variation in the estimation procedure, Journal of The American Statistical Association, **59**, 1225-1226. (2002).
[14] B.B Khare, S. Kumar, Estimation of population mean using known coefficient of variation of the study character in presence of non response, Communication in Statistics-Theory and method **40**, 2044-2058 (2011).
[15] R. Singh, and B.V Sukhatme, Optimum Stratification, Annals of the Institute of Statistical Mathematics (AISM) **21**, 515-528 (1969).
[16] R. Singh, and B.V Sukhatme, Optimum stratification with ratio and regression method of estimation, Annals of the Institute of Statistical Mathematics **25**, 627-633 (1973).
[17] B.B Khare, Allocation in stratified sampling In presence of non-response, Metron **41(I/II)**, 213-221 (1987).
[18] B.B Khare, Estimation of population mean in hetrogenous population in the presence of non-response, Presented in-National conference on recent devolopment in sampling technique and related inference, at BHU , India, April 14-16 (1995).
[19] C. Kadilar, H. Cingi, Ratio Estimator in stratified sampling, Biometrical Journal**45(2)**, 218-225 (2003).
[20] C. Kadilar, H. Cingi, A new ratio Estimator in stratified sampling, Communication in Statistics-Theory and Method **34**, 597-602 (2005).
[21] F.C Okafor, On double sampling for stratification with subsampling the non-respondents, Journal of Indian Social and Agriculture Statistics **48**, 105-113 (1996).
[22] B. B Khare, Seperate Generalized estimators for population mean in the presence of non-response in stratified random sampling, Advance in Statistics and Optimization , 150-158 (2013).
[23] B.B Khare, S. Srivastava, Study of conventional and alternative two phase sampling ratio, product and regression estimators in presence of non response, Proceedings of the National Academy of Science letters-India **65(a)II**, 195-203 (1995).
[24] H.P Singh, M.R Espejo, Double Sampling Ratio-product Estimator of a Finite Population Mean in Sample Surveys, Journal of Applied Statistics **34(1)**, 71-85 (2007).
[25] US Census Bureau, State and County Quick-Facts Data Source for Labour Statistics, **99**, (2012):URL:http://quickfacts.census.gov/qfd/states/48/48497.html [Online; accessed 10-May-2015].
[26] Statistics Education, Diamond Pricing Data from Journal of Statistics Education Volume **9** (2001), URL:http://www.amstat.org/publications/jse/contents_2001.html [Online; accessed 19-May-2015].

Shakeel Ahmad Lecturer, Department of Mathematics and Statics, The University of Lahore, Sargodha Campus, Pakistan

Javid Shabbir Chairman, Department of Statistics, Quaid-i-Azam University Islamabad, Pakistan

Waqar Hafeez Lecturer Department of Statistics University of Central Punjab, Sargodha Campus, Pakistan