
بسم الله الرحمن الرحيم

QUAID-I-AZAM UNIVERSITY ISLAMABAD

M.PHIL THESIS

Estimation of finite Population mean in Non-response using Scrambled Response Model on second call

Author:
Shakeel Ahmed

Supervisor:
Prof.Dr. Javed Shabbir

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Philosophy in Statistics

in the

Department of statistics
Quaid-i-Azam Unversity Islamabad

June 2015

Declaration of Authorship

I, Shakeel Ahmed, declare that this thesis titled, 'Estimation of finite Population Mean' and the work presented in it are my own. I confirm that:

- This work was done wholly in candidature for a degree of M.Phil Statistics at this University.
- Where I got help from the published work of others, this is always clearly stated.
- Where I have quoted from the work of others, the source is always mentioned. Except of such quotations, this thesis is entirely my own research work.
- Where the thesis is based on work done by myself jointly with my supervisor, I have made clear exactly what was done by others and what I have suggested.

Signed:

Date:

Abstract

This study is conducted to improve efficiency of the estimators of finite population mean in presence of non-response using Randomized Response Technique (RRT) to sub-sample non-respondents, assuming that non-response is due to sensitivity of the variable the variable under study. We suggest an estimator for the finite population mean incorporating known coefficient of variation of the study variable in case of quantitative sensitive variable considering a randomization mechanism on the second call that provides privacy protection to the respondents to get truthful information. We also propose generalized ratio and regression type estimators under two-phase sampling. Secondly, we use stratified random sampling to improve efficiency of the propose estimators are derived. Expressions for mean square error of estimators are derived and conditions for which the proposed estimators are more efficient than the relevant estimators under scrambled response model have been obtained in case of simple random sampling and Stratified sampling. Numerical studies are carried out to evaluate performances of the estimators in both sampling schemes. Thirdly, we propose estimators in case of non-response using Ranked set sampling, Extreme Ranked set sampling and Median ranked set sampling to sub-sample non-respondents under randomized response model. Expressions for variances are obtain and conditions under which the proposed estimators are preferred on their counterpart in Simple random sampling with replacement. We conduct a monte carle experiment to see efficiency of the proposed estimators.

Acknowledgements

First and foremost thanks to Allah Almighty, the most compassionate and merciful, who granted me the spirit and enthusiasm required for this task.
Heartiest gratitude to my honorable supervisor Prof.Dr.Javid Shabbir, Chairman of the department. Without his Scholastic guidance and knowledge, this study would not have been completed. His valuable help with constructive comments and suggestions throughout the thesis work have contributed to the success of this study. I would like pay my sincere thanks to those who has generously shared with his view point of my work. I would like to express my heartiest gratitude to my other respected teachers Dr. Zahid Asghar, Dr. Muhammad Yousuf Shad, Dr. Zawar Hussain, Dr. Ijaz Hussain and all my teachers who teach, help and guide me. I am also thankful to my college teachers Mr. Ghullam Abbas and Mr.Shabir Hussain for building my basic concepts.
Specially, i would like to thank my whole family for the support through my entire life and in particular, my parents, without their prayers, love, encouragement and support, I would not have complete my thesis. Very special thanks go out to my father Haji Hassan and my brother Sajid Ali, who motivated and supported me toward higher education. I am heartily grateful and thankful to my friends Zakir Qamar, Sajjad Hussain, Shahenshah Bangesh, Talha Umer, Said Arab, Irfanullah Jan, and Imran Shahid, who always helped and encouraged me a lot at many critical moments of my career and hence ensured my survival.
No acknowledge would ever adequately express our obligations to our teachers who have always wished to see us ying high up at the skies of success. Without their prayers and encouragements our success would have been merely dream.

Shakeel Ahmed

Contents

List of Tables

Abbreviations

SRS	**S**imple **R**andom **S**ampling
SRSWR	**S**imple **R**andom **S**ampling **W**ith **R**eplacement
SRSWOR	**S**imple **R**andom **S**ampling **W**ith **O**ut **R**eplacement
RSS	**R**anked **S**et **S**ampling
ERSS	**E**xtreme **R**anked **S**et **S**ampling
MRSS	**M**edian **R**anked **S**et **S**ampling
RRT	**R**andomized **R**esponse **T**echnique
MSE	**M**ean **S**quare **E**rror
PRE	**P**ercentage **R**elative **E**fficiency

Dedicated to

My Parents,
without whom none of my success would be possible . . .

Chapter 1

Introduction and Review of Literature

1.1 Introduction

The purpose of survey sampling is to get a reliable information about the characteristics of population under study by selecting a sample of certain size from that population. Researcher's interest lies in estimating unknown population parameters by using information contain in sample. Whatever sampling technique we use, unfortunately, in some cases we cannot obtain complete information about the population from selected sample i.e. some units in sample don't give response and this phenomena is named as "Non-response". In almost every field of research related to human beings non-response problems occur when people are contacted through telephonic, mail or direct interview. It depends on the nature of the required information whether survey is conducted on general or sensitive issues of the society. Often in surveys related to general issues like age, income and education etc., the non-response occurrence may be in the form of unavailability of people, respondent not at home, unable to understand the questionnaire. When sensitive information on topics like drug addiction, gambling, illegal income etc is required, then usually peoples hesitate to give true response and either they refuse to answer or give false response. Consequently it estimates population parameters significantly too high or too low. This study covers the topic of estimation of finite population mean with an improvement in precision under non-response. The first attempt to improve efficiency is utilization of known coefficient of variation of study character at estimation stage in SRSWOR, second is use of stratified random sampling to improve efficiency of estimators constructed in SRSWOR and third one is use of RSS, ERSS and MRSS.

1.2 Objectives of the study

This study is conducted in light of following objectives;

1. To enhance response rate using scrambled response model on second call to protect confidentiality in surveys related to sensitive social issues.
2. To improve efficiency in estimation of finite population mean under scrambled response model, using different sampling scheme, when there is non-response on the study variable.
3. To utilize known coefficient of variation of the study variable in case of non-response.
4. To compare efficiency of different estimators.

1.3 Review of Literature

In order to reduce non-response bias and to estimate the unknown characteristic of interest in a population, Hansen and Hurwitz (1946) introduced a procedure of sub-sampling the non-respondents in which it is assumed that all respondents give full response on second call. Later many authors suggested different estimators for the unknown population parameters using some auxiliary information which may suffer from non-response, see Khare and Srivastava (1993), Khare and Srivastava (1995), Khare and Sinha (2009) and Singh and Kumar (2010). They have considered both cases; auxilary variable with non-response and without non-response. In case of sensitive characteristics, it is hard to get a direct truthful response even on second call and results are in violation of the Hansen and Hurwitz (1946) assumption. When survey is concerned with sensitive characteristics of a population then it is imperative to reduce non-response bias and to get reliable information from respondents. Some statistical techniques exist to protect the confidentiality and privacy of respondents and to get the truthful information. These techniques are known as Randomized Response Techniques (RRTs). Warner (1965) introduced the RRT to estimate the proportion of population possessing sensitive attribute which required choosing a yes or no response from a set of nominal categories. After that many authors have contributed towards improving efficiency of the estimators by using this technique. These include Mangat and Singh, Shabbir and Gupta (2005) and Diana and Perri (2009). An RRT method provides quantitative response which depends on a random number from a known distribution. For quantitative sensitive response models see, Pollock and Bek (1976), Eichhorn and Hayre (1983) and Diana and Perri (2011) . Diana et al. (2014) proposed an unbiased estimator of population mean of a quantitative sensitive variable assuming that the people who refuse to respond on first call give scrambled response on second call. This estimator reduces non-response bias by increasing response rate but its variance goes up due to the use of scrambled response model for non-response group.

Several research works exist in literature for reducing the variance of the finite population mean estimator. Searls (1964) used the known coefficient of variation of the study variable in estimating the population mean to improve efficiency. Using Hansen and Hurwitz (1946) technique, Khare and Srivastava (1993), Khare and Srivastava (1995) proposed an estimator for population mean in presence of non-response under two-phase sampling scheme. Khare and Kumar (2009) have proposed an estimator for population mean utilizing known coefficient of variation of the study character using the auxiliary information in estimation under non-response. Khare and Kumar (2011) also proposed generalized ratio and regression type estimators for population mean in two-phase sampling in presence of non-response.

The precision of an estimator depends on the variability among units in the population. One possible way to estimate the population mean with maximum precision is to divide the whole population into certain groups, called strata, which are internally homogeneous and externally heterogeneous and then selecting independent samples of different sizes from each stratum using SRSWOR. Singh and Sukhatme (1969) suggested method of optimum stratification. After that many authors have suggested different types of estimators using the auxiliary information in stratified random sampling (see Singh and Sukhatme (1973), Kadilar (2003), Kadilar (2005) etc). In case of heterogeneous population, when non-response occurs in each stratum. Khare (1987) has proposed an estimator of population mean and also obtain the method of allocation of sample size in different strata for a fixed cost. Khare (1995) proposed an estimator for population mean using post stratification. Later on Okafor (1996) has proposed some estimators for population mean by using post stratification using the auxiliary information in presence of non-response. Khare (2013) also proposed separate generalized ratio type estimators for population mean in presence of non-response in stratified random sampling.

Another way to improve efficiency is the use of Ranked Set Sampling (RSS). RSS is a better alternative to simple random sampling that can sometimes offers large improvement in precision. It was originally developed for estimating herbage yield in agricultural researches by McIntyre (1952). In recent years it has been applied particularly to problems in environmental science. RSS is preferred when actual measurement of a unit is either expensive or time consuming and ranking of a small set of experimental units is cheap and easy. Dell and Clutter (1972) proved that even if ranking is not perfect, the ranked set sampling is still unbiased. Patil (2002) gave a review of the theme of RSS. Many authors including Muttlak (1996) and Samawi and Ahmed (1996) showed that RSS is better than SRSWR in term of accuracy. Bouza (2009) used RSS sampling to Randomized response procedure for estimating population mean of sensitive quantitative character to protect confidentiality of respondents.

Bouza (2002a) proposed an unbiased estimator of population mean using RSS in presence of non-response. Bouza (2010) introduced an estimator for population mean using RSS to sub-sample the non-respondents on second call by claiming that the first visit allows information on Y for ranking the units in sub-sample $\acute{s}_2$ from non-response group s_2 and use different RSS methods for selecting sub-sample on second call.

Taking inspiration from all these works, firstly we propose an estimator of finite population mean under non-response utilizing known coefficient of variation of the study variable, using scrambled response model to sub-sample non-respondents, in SRSWOR. It is assumed that non-response is due to sensitivity of the variable under study. Moreover we suggest two-phase generalized ratio and regression type estimators to improve efficiency using Khare and Kumar (2011) estimators. Secondly, we used stratified random sampling to obtain the proposed estimators assuming that the population of interest is heterogeneous. Finally, we apply RSS, ERSS and MRSS with randomized response technique to sub-sample non-respondents on second call for estimating finite population mean more precisely. We conduct analytical and numerical comparisons between the proposed and existing estimators.

Chapter 2

Estimation of population mean in Simple Random Sampling

2.1 Introduction

Simple random sampling consist of selecting units randomly (with or without replacement) from the whole population without imposing any restriction on population that is why it is easy to handle and less expensive method of selecting a sample. As we select units from whole population without any restriction it tends to results a high variation in estimation of characteristics of population under study.Therefore researchers are primarily interested in reducing variability in estimating population parameters. Keeping this point under consideration, the study proposes an estimator for population mean of a sensitive quantitative variable utilizing known coefficient of variation of the study variable under two-phase sampling scheme using the RRT for sub-sampling non-respondents on the second call; utilization of known constants in estimation stage is a good practice for improving efficiency of estimators. Furthermore, using the proposed estimator, some generalized ratio-type and regression-type estimators are constructed. As a special case, for different values of constants involve, the members of the proposed estimators are identified with their properties. An empirical study is given to evaluate the performances of the mean estimators in SRSWOR.

2.2 Estimators

Let $U = (U_1, U_2, ..., U_N)$ be a finite population of size N. We draw a sample of size n from the population by using SRSWOR. Let y_i and x_i be characteristics of the study variable (y) and the auxilary variable (x) respectively. Searls (1964) proposed an estimator $\bar{y}_s = a\bar{y}$ for estimating the finite population mean, by using optimum value of a i.e. $a_{opt} = \left(1 + \dfrac{1-f}{n} C_y^2\right)^{-1}$, where$f = \dfrac{n}{N}$ and C_y^2 is the coefficient of variation of y. When the population mean $(\bar{X})$ of the auxiliary variable is unknown, we use two-phase sampling scheme. In first phase, we select a sample of size $\acute{n}$ where $\acute{n} < N$ by using SRSWOR to estimate $(\bar{X})$ and then in second phase, we take a smaller sample of size n from the initial $\acute{n}$ units to obtain sample means of the study variable (y) and the auxiliary variable (x). Suppose that from n sampled units in phase 2 only n_1 units respond on first call and n_2 units do not respond. Subsequently, a sample of size $r = \dfrac{n_2}{k}$, where $k > 1$, is drawn from the n_2 non-responding units. Consequently the whole population U is divided into two groups U_1 (respondents) and U_2(non-respondents) of size N_1 and N_2 respectively. When N_1 and N_2 are unknown in advance, Hansen and Hurwitz (1946) proposed an estimator for population mean $\bar{Y}$ which is given by

$$\bar{y}^*_{srs} = w_1\bar{y}_1 + w_2\acute{\bar{y}}_2, \tag{2.1}$$

where

$$w_1 = \frac{n_1}{n},\ w_2 = \frac{n_2}{n},\ \bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i \ \text{ and } \acute{\bar{y}}_2 = \frac{1}{r}\sum_{i=1}^{r} y_i.$$

Also $E(\bar{y}^*_{srs}) = \bar{Y}$ and variance of $\bar{y}^*$ is:

$$V(\bar{y}^*) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 + \frac{W_2(k-1)}{n} S^2_{y(2)},$$

where

$$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y}) \quad \text{and} \quad S^2_{y(2)} = \frac{1}{N_2-1}\sum_{i=1}^{N_2}(y_i - \bar{Y}_2).$$

By ignoring correction factor $1-f$ for easy of computation, we have

$$V(\bar{y}^*) = \frac{1}{n} S_y^2 + \frac{W_2(k-1)}{n} S^2_{y(2)}, \tag{2.2}$$

Using idea of Searls (1964) and Khare and Kumar (2011), the improved Hansen and Hurwitz (1946) estimator is $\bar{y}^{**} = a\bar{y}^*$, where a is Searls constant. The value of a for which $MSE(\bar{y}^{**})$ is minimum, given as

$$a_{opt} = \left[1 + \frac{1}{n} C_y^2 + \frac{W_2(k-1)}{n} \frac{S^2_{y(2)}}{\bar{Y}^2}\right]^{-1}$$

Since $\frac{S_y^2}{\bar{Y}^2}$ and $\frac{S^2_{y(2)}}{\bar{Y}^2}$ don't differ significantly, so we may approximate $\frac{S^2_{y(2)}}{\bar{Y}^2} \cong \frac{S_y^2}{\bar{Y}^2} = C_y^2$. The estimator of a_{opt} for known C_y^2, is given by

$$\hat{a}_{opt} = \left[1 + \frac{C_y^2}{n}\left\{1 + \frac{n_2}{n}(k-1)\right\}\right]^{-1}. \tag{2.3}$$

Now improved estimator for population mean becomes

$$\bar{y}^{**} = \left[1 + \frac{C_y^2}{n}\left\{1 + \frac{n_2}{n}(k-1)\right\}\right]^{-1} \bar{y}^*. \tag{2.4}$$

The $Bias$ and MSE of $\bar{y}^{**}$, is given by

$$Bias(\bar{y}^{**}) = \left[\left\{1 + W_2(k-1)\right\}\frac{C_y^2}{n}\right]\bar{Y}$$

and

$$MSE(\bar{y}^{**}) = (1 - B_1)\frac{S_y^2}{n} + (1 - 2B_2)\frac{W_2(k-1)}{n}S_{y(2)}^2, \tag{2.5}$$

where $B_1 = \frac{C_y^2}{n}\left[1 - W_2^2(k-1)^2\right]$ and $B_2 = \frac{C_y^2}{n}\left[1 + W_2(k-1)\right]$.
By (2.2) and (2.5), we see that $MSE(\bar{y}^{**}) < V(\bar{y}^{*})$, if

$$B_1\frac{S_y^2}{n} + 2B_2\frac{W_2(k-1)}{n}S_{y(2)}^2 > 0.$$

Using the value of B_1 and B_2 and using assumption that $S_y^2 \cong S_{y(2)}^2$, we get

$$\left[1 + W_2(k-1)\right]^2 > 0,$$

which is true for all values of k, this shows that $\bar{y}^{**}$ is always more efficient than $\bar{y}^{*}$. In case of non-response on y and incomplete or complete information on x in a given sample of size n the conventional and alternative ratio and product estimators under two-phase sampling (T_1, T_2) and (T_3, T_4) (Khare and Srivastava (1993)) and regression type estimators (T_{lr1}, T_{lr2}) (Khare and Srivastava (1995)) for $\bar{Y}$ are given as below:

$$T_1 = \bar{y}^{*}\frac{\acute{\bar{x}}}{\bar{x}^{*}}, \quad T_2 = \bar{y}^{*}\frac{\acute{\bar{x}}}{\bar{x}}, \quad T_3 = \bar{y}^{*}\frac{\bar{x}^{*}}{\acute{\bar{x}}}, \quad T_4 = \bar{y}^{*}\frac{\bar{x}}{\acute{\bar{x}}}, \quad T_{lr1} = \bar{y}^{*} + b^{**}(\acute{\bar{x}} - \bar{x}^{*})$$
$$\text{and} \quad T_{lr2} = \bar{y}^{*} + b^{*}(\acute{\bar{x}} - \bar{x}),$$

where

$$\bar{x}^* = w_1\bar{x}_1 + w_2\acute{\bar{x}}_2,\ b^{**} = \frac{s^*_{yx}}{s^{*2}_x},\ b^{**} = \frac{s^*_{yx}}{s^2_x},\ s^{*2}_x = \frac{1}{n-1}\left(\sum_{i=1}^{n_1} x_i^2 + k\sum_{i=1}^{r} x_i^2 - n\bar{x}^{*2}\right),$$

$$s^2_x = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^{*2}\right),\quad s^*_{yx} = \frac{1}{n-1}\left(\sum_{i=1}^{n_1} y_i x_i + k\sum_{i=1}^{r} y_i x_i - n\bar{y}^*\bar{x}^*\right),$$

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i \quad \text{and} \quad \acute{\bar{x}}_2 = \frac{1}{r}\sum_{i=1}^{r} x_i.$$

Using known coefficient of variation of study variable , Khare and Kumar (2011) proposed a generalized ratio type(t_1, t_2) and regression type (t_{lr1}, t_{lr2}) estimators for $\bar{Y}$ as :

$$t_1 = \bar{y}^{**}\left(\frac{\bar{x}^*}{\acute{\bar{x}}}\right)^{\alpha_1},\quad t_2 = \bar{y}^{**}\left(\frac{\bar{x}}{\acute{\bar{x}}}\right)^{\alpha_2}\quad t_{lr1} = \bar{y}^{**} + b^{**}(\acute{\bar{x}} - \bar{x}^*)$$
$$\text{and}\quad t_{lr2} = \bar{y}^{**} + b^{*}(\acute{\bar{x}} - \bar{x}),$$

where α_1 and α_2 are constants. Assuming Y as a quantitative sensitive variable, Diana et al. (2014) have made some modifications to the Hansen and Hurwitz (1946) estimator. They assumed that one group of people give direct truthful response on first call and the other group don't respond on first call later on they give scrambled response. The aim of doing so was to encourage people to respond truthfully by ensuring them that their privacy is protected. Several scrambled response models are available in literature including Additive, Multiplicative and Subtractive scrambled response models etc. Diana et al. (2014) considered the linear combination scrambled response model that was earlier defined by Diana and Perri (2010). A slightly modified version of the Diana et al. (2014) model is as given below.

Let Z be the scrambled response and A and B be two independent random variables unrelated to Y with known means (μ_A, μ_B) and variances (σ^2_A, σ^2_B), such that:

$$Z = AY + B, \tag{2.6}$$

where $E_R(Z) = \mu_A Y + \mu_B$ and variance of Z is $V_R(Z) = \sigma_A^2 Y^2 + \sigma_B^2$, here E_R, V_R are expectation and variance with respect to randomization device .

Let $\hat{y}_i$ be transformed scrambled response of the i^{th} unit whose expectation under randomization mechanism equals to true response y_i.

$$\hat{y}_i = \frac{z_i - \mu_B}{\mu_A}, \quad E_R(\hat{y}_i) = y_i$$
$$V_R(\hat{y}_i) = \frac{\sigma_A^2 y_i^2 + \sigma_B^2}{\mu_A^2} = \phi_i \tag{2.7}$$

Diana et al. (2014) proposed following estimator

$$\hat{\bar{y}}^*_{srs} = w_1 \bar{y}_1 + w_2 \acute{\hat{\bar{y}}}_2, \tag{2.8}$$

where $\acute{\hat{\bar{y}}}_2 = \frac{1}{\acute{n}_2} \sum_{i=1}^{\acute{n}_2} \hat{y}_i$. It is easy to show that $E(\hat{\bar{y}}^*) = \mu$ using the fact that $E_R(\acute{\hat{y}}_2) = \acute{\bar{y}}_2$. The variance of $\hat{\bar{y}}^*$ after ignoring correction factor, is given by

$$V(\hat{\bar{y}}^*) = \frac{1}{n} S_y^2 + \frac{W_2(k-1)}{n} S_{y(2)}^2 + \frac{W_2 k}{n} \left[\frac{\sigma_A^2}{\mu_A^2} \left\{ S_{y(2)}^2 + \mu_{y(2)}^2 \right\} + \frac{\sigma_A^2}{\mu_A^2} \right].$$

It can also be written as :

$$V(\hat{\bar{y}}^*) = \frac{1}{n} S_y^2 + \frac{W_2(k-1)}{n} S_{y(2)}^2 + \frac{k}{n_h N} \sum_{i=1}^{N_2} \phi_i. \tag{2.9}$$

Diana and Perri (2010) suggested two possible ways to estimate unknown $\mu_{y(2)}$, One is to use a good guess from previous works, and the other is to use a pilot survey so that sample estimate can supply information about the second moment keeping in mind its sensitive nature. Comparing Equations (2.2) and (2.9), we see that $\hat{\bar{y}}^*$ is less efficient than $\bar{y}^*$ but on the other hand it gives greater privacy protection. Diana et al. (2014) have made a trade off between efficiency and privacy by choosing a suitable scrambled response model from among several models because efficiency and privacy move in opposite directions. So it is impossible to simultaneously keep both of these at a desired level for a fixed sample size. Keeping this point in mind we make an attempt to improve efficiency at a fixed level of privacy. For this purpose we introduce an estimator for population mean of a sensitive quantitative variable by utilizing known coefficient of variation of the study variable which is more efficient than the Diana et al. (2014) estimator under certain assumption. The proposed estimator is given by

$$\bar{y}^{**} = k_1 \bar{y}^*, \tag{2.10}$$

where k_1 is a constant. The optimum value of k_1 which minimize MSE of $\hat{\bar{y}}^{**}$, is given by

$$k_{1(opt)} = \left[1 + \frac{1}{n}C_y^2 + \frac{W_2(k-1)}{n}\frac{S_{y(2)}^2}{\bar{Y}^2} + \frac{k}{n_h N_h}\frac{\sigma_r^2}{\bar{Y}^2}\right]^{-1}.$$

As we discussed earlier $\frac{S_y^2}{\bar{Y}^2}$ and $\frac{S_{y(2)}^2}{\bar{Y}^2}$ don't differ significantly, so we may approximate

$\frac{S_y^2}{\bar{Y}^2} \cong \frac{S_{y(2)}^2}{\bar{Y}^2} \cong C_y^2$. So estimated value of k_1 becomes

$$\hat{k}_{1(opt)} = \left[1 + \frac{C_y^2}{n}\left\{1 + \frac{n_2}{n}(k-1)\right\} + \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right]^{-1}.$$

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

$$Bias(\hat{\bar{y}}^{**}) = -\left[\left\{1 + W_2(k-1)\right\}\frac{C_y^2}{n} + \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right]\bar{Y} \qquad \text{(see: Appendix A)} \tag{2.11}$$

and

$$MSE(\hat{\bar{y}}^{**}) = (1-A^*)\frac{S_y^2}{n} + (1-2B^*)\frac{W_2(k-1)}{n}S_{y(2)}^2 + \frac{k}{nN}\sigma_r^2\left\{1-\frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right\}, \quad (2.12)$$

(see:Appendix A)

where

$$A^* = \frac{C_y^2}{n}\left\{1 - W_2^2(k-1)^2\right\} + \frac{2k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}$$
$$\text{and} \quad B^* = \frac{C_y^2}{n}\left\{1 + W_2(k-1)\right\} + \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2} \quad (2.13)$$

By (2.9) and (2.12), we see that $MSE(\hat{\bar{y}}^{**}) < V(\hat{\bar{y}}^*)$, if

$$\left[\frac{C_y^2}{n}\left\{1 - W_2^2(k-1)^2\right\} + \frac{2k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right]\frac{S_y^2}{n} + 2\left[\frac{C_y^2}{n}\left\{1 + W_2(k_h-1)\right\} + \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right]\frac{W_2(k-1)S_{y(2)}^2}{n} + \left(\frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right)^2 > 0,$$

as the last term is positive and assuming that $\frac{S_y^2}{n} \cong \frac{S_{y(2)}^2}{n}$, we get

$$\frac{C^2}{n}\{1 - W_2^2(k-1)^2\} + 2\frac{C^2}{n}\{1 + W_2(k-1)\}W_2(k-1) + \frac{2k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\{1 + W_2(k-1)\} > 0,$$

$=> \left\{1 + W_2(k-1)\right\}^2 > 0$ and $\left\{1 + W_2(k-1)\right\} > 0.$

Both conditions satisfy for all $k > 1$. This shows that $\hat{\bar{y}}^{**}$ is always more efficient than $\hat{\bar{y}}^*$ To increase efficiency in estimation of finite population mean we propose generalized ratio type estimators using Khare and Kumar (2011) estimator for the case of non-response on x and no non-response on x as follow:

$$\hat{t}_1 = \hat{\bar{y}}^{**}\left(\frac{\bar{x}^*}{\acute{\bar{x}}}\right)^{a_1} \quad (2.14)$$

and

$$\hat{t}_2 = \hat{\bar{y}}^{**} \left(\frac{\bar{x}}{\acute{\bar{x}}} \right)^{a_2}, \tag{2.15}$$

where a_1 and a_2 are constants to be determine. The generalized ratio type estimators for the case of non-response on x and no non-response on x are

$$\hat{t}_{(lr)1} = \hat{\bar{y}}^{**} + b^{**} \left(\acute{\bar{x}} - \bar{x}^* \right) \tag{2.16}$$

and

$$\hat{t}_{(lr)1} = \hat{\bar{y}}^{**} + b^{*} \left(\acute{\bar{x}} - \bar{x} \right), \tag{2.17}$$

where b^{**} and b^* are defined earlier.

Now different members of these generalized ratio and regression type estimators for certain values of constants involved are obtain. By putting $a_1 = a_2 = -1$ and $a_1 = a_2 = 1$ in (2.14) and (2.15) the estimators reduce to conventional and alternative to two phase stratified ratio estimator and Product type estimators respectively , using scrambled response model to non-response group using coefficient of variation of the study character.

The alternative two-phase ratio type estimators are given by:

$$\hat{t}_3 = \hat{\bar{y}}^{**} \left(\frac{\acute{\bar{x}}}{\bar{x}^*} \right) \quad \text{and} \quad \hat{t}_4 = \hat{\bar{y}}^{**} \left(\frac{\acute{\bar{x}}}{\bar{x}} \right)$$

The alternative two-phase product type estimators are given by:

$$\hat{t}_5 = \hat{\bar{y}}^{**} \left(\frac{\bar{x}^*}{\acute{\bar{x}}} \right) \quad \text{and} \quad \hat{t}_6 = \hat{\bar{y}}^{**} \left(\frac{\bar{x}}{\acute{\bar{x}}} \right)$$

Now for $k = 1, a_1 = a_2 = 1$ and $a_1 = a_2 = -1$, the estimator in (2.14), (2.15), (2.16) and (2.17) reduce to conventional and alternative to two-phase stratified ratio, product and regression type estimators respectively, using scrambled response model to non-response group.

$$\hat{T}_3 = \hat{\bar{y}}^* \left(\frac{\acute{\bar{x}}}{\bar{x}^*}\right), \quad \hat{T}_4 = \hat{\bar{y}}^* \left(\frac{\acute{\bar{x}}}{\bar{x}}\right), \quad \hat{T}_5 = \hat{\bar{y}}^* \left(\frac{\bar{x}^*}{\acute{\bar{x}}}\right), \quad \hat{T}_6 = \hat{\bar{y}}^* \left(\frac{\bar{x}}{\acute{\bar{x}}}\right),$$

$$\hat{T}_{(lr)1} = \hat{\bar{y}}^* + b^{**}(\acute{\bar{x}} - \bar{x}^*) \text{ and } \hat{T}_{(lr)2} = \hat{\bar{y}}^* + b^{*}(\acute{\bar{x}} - \bar{x})$$

2.3 Mean Square Errors of different estimators

In this section, we derive expressions for MSE of different estimators. For this purpose, we define following error terms.

Let

$$\hat{e}_0^* = \frac{\hat{\bar{y}}^* - \bar{Y}}{\bar{Y}}, \quad e_1^* = \frac{\bar{x}^* - \bar{X}}{\bar{X}} \text{ and } \acute{e}_1 = \frac{\acute{\bar{x}} - \bar{X}}{\bar{X}}$$

such that $E(\hat{e}_0^*) = E(e_1^*) = E(\acute{e}_1) = 0$ and

$$E(\hat{e}_0^{*2}) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_y^2}{\bar{Y}^2} + \frac{W_2(k-1)}{n}\frac{S_{y(2)}^2}{\bar{Y}^2} + \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}$$

$$E(e_1^{*2}) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_x^2}{\bar{X}^2} + \frac{W_2(k-1)}{n}\frac{S_{x(2)}^2}{\bar{X}^2}, \; E(\acute{e}_1^2) = \left(\frac{1}{\acute{n}} - \frac{1}{N}\right)\frac{S_x^2}{\bar{X}^2},$$

$$E(e_1^*\acute{e}_1) = \left(\frac{1}{\acute{n}} - \frac{1}{N}\right)\frac{S_x^2}{\bar{X}^2}, \; E(\hat{e}_0^*\acute{e}_1) = \left(\frac{1}{\acute{n}} - \frac{1}{N}\right)\frac{S_{yx}}{\bar{X}\bar{Y}},$$

$$E(\hat{e}_0^* e_1^*) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_{yx}}{\bar{X}_h\bar{Y}} + \frac{W_2(k-1)}{n}\frac{S_{yx(2)}}{\bar{X}\bar{Y}}.$$

Consider the generalized estimator $\hat{t}$ in term of errors:

$$\begin{aligned}
\hat{t}_1 &= \hat{\bar{y}}^{**}\left(\frac{\bar{x}^*}{\acute{\bar{x}}}\right)^{a_1} \\
&= k_1\hat{\bar{y}}^*\left(\frac{\bar{x}^*}{\acute{\bar{x}}}\right)^{a_1} \\
&= k_1\bar{Y}(1+\hat{e}_0^*)\left[\frac{(1+e_1^*)}{(1+\acute{e}_1)}\right]^{a_1} \\
\hat{t}_1 - \bar{Y} &= (k_1-1)\bar{Y} + k_1\bar{Y}\Bigg[\hat{e}_0^* + a_1e_1^* - a_1\acute{e}_1 + \frac{a_1(a_1-1)}{2}e_1^{*2} \\
&\quad + \frac{a_1(a_1+1)}{2}\acute{e}_1^2 + a_1e_1^*\hat{e}_0^* - a_1\acute{e}_1\hat{e}_0^* - a_1^2\acute{e}_1\hat{e}_1^*\Bigg]
\end{aligned}$$

Taking square on both side and neglecting higher term from the right hand side.

$$\begin{aligned}
(\hat{t}_1-\bar{Y})^2 \cong (k_1-1)^2\bar{Y}^2 + \bar{Y}^2\Bigg[&k_1^2\Big\{\hat{e}_0^{*2} + a_1^2e_1^{*2} + a_1^2\acute{e}_1^2 - 2a_1e_1^*\acute{e}_1 \\
&- 2a_1\hat{e}_0^*\acute{e}_1 + 2a_1\hat{e}_0^*e_1^*\Big\} + k_1(k_1-1)\Big\{a_1(a_1-1)e_1^{*2} - 2a_1^2e_1^*\acute{e}_1 \\
&+ a_1(a_1+1)\acute{e}_1^2 + 2a_1\hat{e}_0^*e_1^* - 2a_1\hat{e}_0^*\acute{e}_1\Big\}\Bigg].
\end{aligned}$$

Taking expectation, we have

$$\begin{aligned}
MSE(\hat{t}_1) \cong (k_1-1)^2\bar{Y}^2 + k_1^2V(\hat{\bar{y}}^*) + \bar{Y}^2\Bigg[&k_1a_{h1}(2k_1a_1 - k_1 - a_1 + 1)\Bigg\{\left(\frac{1}{n}-\frac{1}{\acute{n}}\right)C_x^2 \\
&+\frac{W_2(k-1)}{n}C_{x(2)}^2\Bigg\} + 2k_1a_1(2k_1-1)\Bigg\{\left(\frac{1}{n}-\frac{1}{\acute{n}}\right)C_{yx} + \frac{W_2(k-1)}{n}C_{yx(2)}\Bigg\}\Bigg]
\end{aligned}$$

We have

$$B^* = \frac{C_y^2}{n}\Big\{1 + W_2(k-1)\Big\} + \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}$$

After expanding k_1 and neglecting higher order terms, the optimum value of k_1, is given by

$$k_{1(opt)} = 1 - \left[\frac{C_y^2}{n}\left\{1 + W_2(k-1)\right\} + \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right].$$

So we can write $k_1 \cong 1 - B^*$. Similarly

$$k_1^2 \cong 1 - 2\left[\frac{C_y^2}{n}\left\{1 + W_2(k-1)\right\} + \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right] \cong (1 - 2B^*) \text{ and so on.}$$

Substituting this result in $MSE(t_1)$ we get

$$MSE(\hat{t}_1) \cong (1 - 2B^*)V(\hat{\bar{y}}^*) + \bar{Y}^2\left[a_1\left\{a_1 - (3a_1 - 1)B^*\right\}\left\{A_1 C_x^2 + A_3 C_{x(2)}^2\right\} + 2a_1(1 - 3B^*)\left\{A_1 C_{yx} + A_3 C_{yx(2)}\right\}\right]$$

or

$$MSE(\hat{t}_1) \cong (1 - 2B^*)V(\hat{\bar{y}}^*) + a_1\left\{a_1 - (3a_1 - 1)B^*\right\}R^2\left\{A_1 S_x^2 + A_3 S_{x(2)}^2\right\} + 2a_1(1 - 3B^*)R\left\{A_1 S_{yx} + A_3 S_{yx(2)}\right\}, \tag{2.18}$$

where

$$A_1 = \frac{1}{n} - \frac{1}{\acute{n}}, \quad A_3 = \frac{W_2(k-1)}{n} \text{ and } R = \frac{\bar{Y}}{\bar{X}}.$$

The $MSE$of $\hat{t}_2$ can be obtain easily as:

$$MSE(\hat{t}_2) \cong (1 - 2B^*)V(\hat{\bar{y}}^*) + A_1 a_2\left\{(a_2 - (3a_2 - 1)B^*)R^2 S_x^2 + 2R(1 - 3B^*)S_{yx}\right\} \tag{2.19}$$

We obtain optimum values of a_1 and a_2 , after differentiating (2.18) and (2.19) w.r.t a_1 and a_2 respectively:

$$a_{1(opt)} = -\left[\frac{B^*}{2(1-3B^*)} + \frac{A_1 S_{yx} + A_3 S_{yx(2)}}{R\{A_1 S_x^2 + A_3 S_{x(2)}^2\}}\right]$$

and

$$a_{2(opt)} = -\left[\frac{B^*}{2(1-3B^*)} + \frac{S_{yx}}{RS_x^2}\right]$$

Using $a_{1(opt)}$ and $a_{2(opt)}$ in Equation (2.18) and (2.19) respectively, we get minimum MSE of t_1 and t_2

$$\begin{aligned} MSE(\hat{t}_1)_{min} \cong{}& (1-2B^*)V(\hat{\bar{y}}^*) - (1-3B^*)\frac{\{A_1 S_{yx} + A_3 S_{yx(2)}\}^2}{A_1 S_x^2 + A_3 S_{x(2)}^2} \\ &- \frac{B^{*2}R^2\{A_1 S_x^2 + A_3 S_{x(2)}^2\}}{4(1-3B^*)} - B^* R\left\{A_1 S_{yx} + A_3 S_{yx(2)}\right\} \end{aligned} \quad (2.20)$$

and

$$\begin{aligned} MSE(\hat{t}_2)_{min} \cong{}& (1-2B^*)V(\hat{\bar{y}}^*) - (1-3B^*)\frac{A_1 S_{yx}^2}{S_x^2} \\ &- \frac{B^{*2}R^2 A_1 S_x^2}{4(1-3B^*)} - B^* R A_1 S_{yx}. \end{aligned} \quad (2.21)$$

MSE of t_3 and t_4, we put $a_1 = a_2 = -1$ in Equations (2.18) and (2.19) we get :

$$\begin{aligned} MSE(\hat{t}_3) \cong{}& (1-2B^*)V(\hat{\bar{y}}^*) + (1-4B^*)R^2\left\{A_1 S_x^2 + A_3 S_{x(2)}^2\right\} \\ &- 2(1-3B^*)R\left\{A_1 S_{yx} + A_3 S_{yx(2)}\right\} \end{aligned} \quad (2.22)$$

and

$$MSE(\hat{t}_4) \cong (1-2B^*)V(\hat{\bar{y}}^*) + A_1\left\{(1-4B^*)R^2 S_x^2 - 2(1-3B^*)R S_{yx}\right\}. \quad (2.23)$$

To find the MSE of t_5 and t_6 we put $a_1 = a_2 = 1$ in Equations (2.18) and (2.19) , we get:

$$MSE(\hat{t}_5) \cong (1-2B_h^*)V(\hat{\bar{y}}_h^*) + (1-2B^*)R^2\left\{A_{h1}S_x^2 + A_3 S_{x(2)}^2\right\} + 2(1-3B^*)R_h\left\{A_1 S_{yx} + A_{h3}S_{yx(2)}\right\} \tag{2.24}$$

and

$$MSE(\hat{t}_6) \cong (1-2B^*)V(\hat{\bar{y}}^*) + A_1\left\{(1-2B^*)R^2 S_x^2 + 2(1-3B^*)RS_{yx}\right\}. \tag{2.25}$$

The MSE of regression type estimator in both cases are given by:

$$MSE(\hat{t}_{(lr)1}) \cong (1-2B^*)V(\hat{\bar{y}}^*) + \beta^2\left\{A_1 S_x^2 + A_3 S_{x(2)}^2\right\} - 2\beta(1-B^*)\left\{(A_1 S_{yx} + A_3 S_{yx(2)}\right\} \tag{2.26}$$

and

$$MSE(\hat{t}_{(lr)2}) = (1-2B^*)V(\hat{\bar{y}}^*) + A_1\left\{\beta^2 S_x^2 - 2\beta(1-B^*)S_{yx}\right\}. \tag{2.27}$$

To obtain MSE of $\hat{T}_3$, $\hat{T}_4$, $\hat{T}_5$, $\hat{T}_6$, $\hat{T}_{(lr)1}$ and $\hat{T}_{(lr)2}$, put $k_1 = 1$ in Equations (2.21), (2.22), (2.23),(2.24),(2.25) and (2.26) as follows;

$$MSE(\hat{T}_3) \cong V(\hat{\bar{y}}^*) + R^2\left\{A_1 S_x^2 + A_3 S_{x(2)}^2\right\} - 2R\left\{A_1 S_{yx} + A_3 S_{yx(2)}\right\}, \tag{2.28}$$

$$MSE(\hat{T}_4) \cong (1-2B^*)V(\hat{\bar{y}}) + A_1\left\{R^2 S_x^2 - 2RS_{yx}\right\}, \tag{2.29}$$

$$MSE(\hat{T}_5) \cong V(\hat{\bar{y}}^*) + R^2\left\{A_1 S_x^2 + A_3 S_{x(2)}^2\right\} + 2R\left\{A_1 S_{yx} + A_3 S_{yx(2)}\right\}, \quad (2.30)$$

$$MSE(\hat{T}_6) \cong (1-2B^*)V(\hat{\bar{y}}^*) + A_1\left\{R^2 S_x^2 + 2RS_{yx}\right\}, \quad (2.31)$$

$$MSE(\hat{T}_{(lr)1}) \cong V(\hat{\bar{y}}^*) + \beta^2\left\{A_1 C_x^2 + A_3 C_{x(2)}^2\right\} - 2\beta\left\{A_1 C_{yx} + A_3 C_{yx(2)}\right\} \quad (2.32)$$

and

$$MSE(\hat{T}_{(lr)2}) \cong V(\hat{\bar{y}}^*) + A_1\left\{\beta^2 C_x^2 - 2\beta(1-B^*)C_{yx}\right\}. \quad (2.33)$$

2.4 Comparison of different Estimators

Conditions under which proposed estimators are better than relevant existing estimator in term of efficiency are given in this section.

Condition (i)

By (2.9) and (2.18), $MSE(\hat{t}_1) < V(\hat{\bar{y}}^*)$

$$\rho < \frac{2B^*\frac{C_y^2}{n} - A_1 a_1\{a_1 - (3a_1 - 1)B^*\}C_x^2}{2a_1(1-3B^*)A_1 C_y C_x} \quad \text{and}$$
$$\rho_2 < \frac{2B^* C_{y(2)}^2 - \frac{R^2}{R_2^2} a_1\{a_1 - (3a_1 - 1)B^*\}C_{x(2)}^2}{2\frac{R}{R_2} a_1(1-3B^*)C_{y(2)}C_{x(2)}}.$$

Condition (ii)

By (2.9) and (2.19), $MSE(\hat{t}_2) < V(\hat{\bar{y}}^*)$, if

$$\rho < \frac{2B^*\frac{C_y^2}{n} - A_1 a_2\{a_2 - (3a_2 - 1)B^*\}C_x^2}{2A_1 a_2(1-3B^*)C_y C_x}.$$

Condition (iii)

By (2.22) and (2.28), $MSE(\hat{t}_3) < MSE(\hat{T}_3)$ if

$$\rho < \frac{2C_x}{3C_y} + \frac{C_y}{3nA_1C_x} \text{ and } \rho_2 < \frac{2C_{x(2)}}{3\frac{R_2}{R}C_{y(2)}} + \frac{C_{y(2)}}{3\frac{R}{R_2}C_{x(2)}}.$$

Condition (iv)

By (2.23) and (2.29), $MSE(\hat{t}_4) < MSE(\hat{T}_4)$ if

$$\rho < \frac{2C_x}{3C_y} + \frac{C_y}{3nA_1C_x}.$$

Condition (v)

By (2.24) and (2.30), $MSE(\hat{t}_5) < MSE(\hat{T}_5)$ if

$$\rho > -\left[\frac{C_x}{3C_y} + \frac{C_y}{3nA_1C_x}\right] \text{ and } \rho_2 > -\left[\frac{C_{x(2)}}{3\frac{R_2}{R}C_{y(2)}} + \frac{C_{y(2)}}{3\frac{R}{R_2}C_{x2}}\right].$$

Condition (vi)

By (2.25) and (2.31), $MSE(\hat{t}_6) < MSE(\hat{T}_6)$ if

$$\rho > -\left[\frac{C_x}{3C_y} + \frac{C_y}{3nA_1C_x}\right] \text{ and}$$

Condition (vii)

By (2.18) and (2.19), $MSE(\hat{t}_1) < MSE(\hat{t}_2)$ if

$$\rho < -\frac{[a_1\{a_1 - (3a_1 - 1)B^*\} - a_2\{a_2 - (3a_2 - 1)B^*\}]\frac{C_x}{C_y}}{2(a_1 - a_2)(1 - 3B^*)} \text{ and}$$
$$\rho_2 < -\frac{\{a_1 - (3a_1 - 1)B^*\}\frac{R}{R_2}C_{x(2)}}{2(1 - 3B^*)C_{y(2)}}$$

Condition (viii)

By (2.18) and (2.22), $MSE(\hat{t}_1) < MSE(\hat{t}_3)$ if

$$\rho < -\frac{\left[a_1\{a_1-(3a_1-1)B^*\}-(1-4B^*)\right]\frac{C_x}{C_y}}{2(a_1+1)(1-3B^*)} \quad \text{and}$$

$$\rho_2 < -\frac{\{a_1-(3a_1-1)B^*\}\frac{R}{R_2}\frac{C_{x(2)}}{C_{y(2)}}}{2(a_1+1)(1-3B^*)}$$

Condition (ix)

By (2.19) and (2.29), $MSE(\hat{t}_2) < MSE(\hat{t}_4)$ if

$$\rho < -\frac{\left[a_2\{a_2-(3a_2-1)B^*\}-(1-4B^*)\right]\frac{C_x}{C_y}}{2(a_2+1)(1-3B^*)}$$

Condition (x)

By (2.26) and (2.32), $MSE(\hat{t}_{(lr)1}) < MSE(\hat{T}_{(lr)1})$ if

$$\beta < \frac{V(\hat{\bar{y}}^*)}{A_1S_{yx}+A_3S_{yx(2)}}$$

Condition (xi)

By (2.27) and (2.33), $MSE(\hat{t}_{(lr)2}) < MSE(\hat{T}_{(lr)2})$ if

$$\beta < \frac{V(\hat{\bar{y}}^*)}{A_1S_{yx}}$$

Condition (i)-(xi) can easily be verified by using empirical data.

2.5 Empirical Study

We use the following data sets, for an empirical study.

2.5.1 Data 1 [Sourc: Khare and Kumar (2011)]

Y: Number of cultivators

X: Population of villages

The proportion of non-respondents in the population is 25%, so they considered last 24 units of population as non-respondents. It is also assumed that A and B are two independent scrambled variables, each distributed uniformly in the interval $[0, 1]$. The summary statistics are:

$N = 96$, $n = 25$, $\bar{Y} = 185.22$, $\bar{X} = 1807.23$, $\bar{Y}_2 = 128.46$, $\bar{X}_2 = 1571.71$, $S_y = 195.03$, $S_x = 1921.77$, $S_{y(2)} = 97.82$, $S_{x(2)} = 1068.44$, $S_{yx} = 338835.88$, $S_{yx(2)} = 93560.01$, $\rho = 0.904$, $\rho_2 = 0.895$, $\mu_A = 0.51$, $\mu_B = 0.509$, $\sigma_A^2 = 0.08037$ and $\sigma_B^2 = 0.08597$.

The results are given in Table 2.1.

TABLE 2.1: MSE and PRE of different estimators w.r.t $\bar{y}^*$ and $\hat{\bar{y}}^*$ for different values of k :

Estimators	$k=2$		$k=3$		$k=4$	
	MSE	**PRE**	MSE	**PRE**	MSE	**PRE**
$\bar{y}^*$	1617.15	100.00*	1712.84	100.00*	1808.53	100.00*
$\bar{y}^{**}$	1543.28	104.78*	1636.77	104.64*	1734.45	104.27*
$\hat{\bar{y}}^*$	1789.62	100.00	1971.52	100.00	2153.44	100.00
$\hat{\bar{y}}^{**}$	1698.61	105.42	1867.67	105.56	2039.53	105.80
$\hat{t}_1$	1074.33	166.57	1114.98	176.82	1146.04	187.90
$\hat{t}_2$	1368.35	130.78	1537.46	128.73	1698.71	126.77
$\hat{t}_3$	1095.44	163.36	1149.10	171.57	1193.47	180.10
$\hat{t}_4$	1161.52	154.07	1278.60	154.19	1385.91	155.38
$\hat{t}_{lr1}$	1098.07	162.97	1156.51	170.47	1209.72	178.01
$\hat{t}_{lr2}$	1158.60	154.46	1275.22	154.60	1381.95	155.82
$\hat{T}_3$	1257.83	142.27	1367.88	144.13	1477.93	145.71
$\hat{T}_4$	1329.70	134.58	1511.61	130.42	1693.53	127.15
$\hat{T}_{lr1}$	1247.72	143.43	1354.05	145.60	1460.38	147.45
$\hat{T}_{lr2}$	1323.30	135.33	1505.22	130.97	1687.13	127.67

Here "*" stands for the percentage relative efficiency (PRE) of the estimators without using RRT method. And $PRE(\bullet) = \dfrac{V(\hat{\bar{y}}^*)}{MSE(\bullet)} \times 100$, where $MSE(\bullet)$ are the MSE of different estimators.

Having a view on Table 2.1 , we see from first and third row that some efficiency has been lost due to use of scrambled response model to sub-sample non-respondents. But it shows that for fixed confidentiality level and fixed sample size $(\acute{n}, n)$ the estimator $\hat{\bar{y}}^{**}$ is more efficient than Diana et al. (2014) estimator ($\hat{\bar{y}}^*$) for all values of $k > 1$ used here. Further more generalized estimators $(\hat{t}_1, \hat{t}_2)$ are more efficient than the proposed estimator as well as conventional and alternative two-phase ratio type estimators $(\hat{t}_3, \hat{t}_4)$ using known coefficient of variation. Also two-phase ratio type estimators $(\hat{t}_3, \hat{t}_4)$ and regression type estimators $(\hat{t}_{lr1}, \hat{t}_{lr2})$ respectively using known coefficient of variation of the study variable are more efficient than the corresponding estimators $(\hat{T}_3, \hat{T}_4)$ and $(\hat{T}_{lr1}, \hat{T}_{lr2})$ without using coefficients of variation under scrambled response model respectively. The PRE of different estimators reduces as k increases but on the other hand PRE of the proposed estimators increases.

2.5.2 Data 2 [Sourc: Khare and Kumar (2011)]

Y : Average value of product sold (Dollar thousands)

X : Average size of farms (hundreds of acres)

The proportion of non-respondents in the population is 20%. We consider last 11 units of population as non-respondents. It is also assumed that A and B are two independent scrambled variables each distributed uniformly in the interval $[0, 1]$. The summary statistics are:

$N = 56$, $n = 15$, $\bar{Y} = 61.59$, $\bar{X} = 75.79$, $\bar{Y}_2 = 51.02$, $\bar{X}_2 = 57.60$

$S_y = 24.03$, $S_x = 12.47$, $S_{y(2)} = 13.91$, $S_{x(2)} = 10.50$, $S_{yx} = -152.14$,

$S_{yx(2)} = -55.32$, $\rho = -0.508$, $\rho_2 = -0.379$, $\mu_A = 0.506$, $\mu_B = 0.477$, $\sigma_A^2 = 0.0655$ and $\sigma_B^2 = 0.0877$.

The results are given in Table 2.2.

Table 2.2: MSE and PRE of different estimators w.r.t $\bar{y}^*$ and $\hat{\bar{y}}^*$ for different values of k :

Estimators	$k = 2.5$		$k = 3$		$k = 4$	
	MSE	**PRE**	MSE	**PRE**	MSE	**PRE**
$\bar{y}^*$	42.36	100.00	43.65	100.00	46.23	100.00
$\bar{y}^{**}$	41.90	101.09*	43.18	101.09*	45.73	101.09*
$\hat{\bar{y}}^*$	73.94	100.00	78.87	100.00	93.33	100.00
$\hat{\bar{y}}^{**}$	72.51	101.96	77.25	102.08	91.10	102.45
$\hat{t}_1$	66.71	110.83	71.01	111.05	83.59	111.65
$\hat{t}_2$	67.11	110.16	71.55	110.22	84.45	110.51
$\hat{t}_5$	64.59	114.46	64.59	122.09	64.59	144.48
$\hat{t}_6$	64.98	113.78	64.98	121.36	64.98	143.62
$\hat{t}_{lr1}$	66.95	110.43	71.37	110.49	84.25	110.77
$\hat{t}_{lr2}$	66.95	110.42	71.38	110.49	84.24	110.79
$\hat{T}_5$	69.74	106.01	74.55	105.78	88.79	105.11
$\hat{T}_6$	70.08	105.50	75.01	105.13	89.47	104.30
$\hat{T}_{lr1}$	69.91	105.75	74.82	105.40	89.25	104.56
$\hat{T}_{lr2}$	69.97	105.67	74.90	105.29	89.36	104.44

Here "*" stands for the percentage relative efficiency (PRE) of the estimators without using RRT method. And $PRE(\bullet) = \dfrac{V(\hat{\bar{y}}^*)}{MSE(\bullet)} \times 100$, where $MSE(\bullet)$ are the MSE of different estimators.

From Table 2.2, it is obvious that in case of negative correlation between the auxiliary variable and the study variable, for fixed sample size $(\acute{n}, n)$, the estimator $\hat{\bar{y}}^{**}$ is more efficient than Diana et al. (2014) estimator ($\hat{\bar{y}}^*$) for all values of $k > 1$ used here, but PRE of $\hat{\bar{y}}^{**}$ is very small. Further, generalized estimators $(\hat{t}_1, \hat{t}_2)$ are more efficient than conventional and alternative two phase product type estimators $(\hat{t}_5, \hat{t}_6)$ using known coefficients of variation. Also under two-phase sampling product type estimators $(\hat{t}_5, \hat{t}_6)$ and regression type estimators $(\hat{t}_{lr1}, \hat{t}_{lr2})$ using known coefficients of variation are more efficient than the respective estimators $(\hat{T}_3, \hat{T}_4)$ and $(\hat{T}_3, \hat{T}_4)$ without using coefficient of variation under scrambled response model respectively. In case of negative correlation the $PRE's$ of proposed estimators tends to increase as k increases which can be notice from left to right of the Table 2.2.

2.6 Conclusion

This chapter covers a modified Diana et al. (2014) mean estimator using known coefficient of variation of study variable to improve efficiency of the estimator for a fixed level of privacy. Generalized ratio and regression type estimators, using scrambled response model to non-respondent at second call, are also obtain that are more efficient than proposed mean per unit estimator. These proposed estimators perform better in case of positive correlation between the auxiliary and the study variables. This chapter concludes that use of known coefficient of variation increase efficiency of the estimator in presence of non-response.

Chapter 3

Estimation in Stratified Random Sampling

3.1 Introduction

Stratified random sampling consist of dividing the population into certain groups called "strata" and then selecting SRS's of different sizes from different strata using some methods of allocation. The propose of doing so is to convert a heterogeneous population to small homogeneous groups so that one can select an SRS from each stratum with less variability.

In this chapter we propose an estimator for population mean of a sensitive quantitative character using known coefficient of variation under stratified random sampling in presence of non-response. We also propose separate generalized type and regression type estimators using proposed estimator. The members of these estimators for different values of constant involved are also obtained. 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. An empirical study is carried out in Section 6.

3.2 Notations

Consider a finite population $U = (U_1, U_2, U_3, ..., U_N)$ of size N and is divided into L strata, each of size N_h for $(h = 1, 2, ...L)$, 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 $(h = 1, 2, ..., L)$ and $(i = 1, 2, 3, ..., N_h)$. 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 on both phases. Now 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

$S^2_{yh} = \frac{1}{N_h-1}\sum_{i=1}^{N_h}(y_{hi} - \bar{Y}_h)^2$: Population variance of the study variable for the h^{th} stratum.

$S^2_{xh} = \frac{1}{N_h-1}\sum_{i=1}^{N_h}(x_{hi} - \bar{X}_h)^2$: 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.

$S^2_{yh(2)} = \frac{1}{N_{h2}-1}\sum_{i=1}^{N_{h2}}(y_{hi} - \bar{Y}_{h2})^2$: Population variance of the study variable for non-response group in the h^{th} stratum.

$S^2_{xh(2)} = \frac{1}{N_{h2}-1}\sum_{i=1}^{N_{h2}}(x_{hi} - \bar{X}_{h2})^2$: 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.

3.3 The Estimators

Using Hansen and Hurwitz (1946) technique, the estimator for population mean of the h^{th} stratum, is given by:

$$\bar{y}_h^* = w_{h1}\bar{y}_{h1} + w_{h2}\acute{\bar{y}}_{h2},$$

where $w_{h1} = \frac{n_{h1}}{n_h}$ and $w_{h2} = \frac{n_{h2}}{n_h}$. The sample mean estimator in stratified sampling, is given by:

$$\bar{y}_{st}^* = \sum_{h=1}^{L} P_h \bar{y}_h^*. \tag{3.1}$$

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

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

After ignoring the correction factor $(1 - f_h)$ for ease of computation, we have

$$V(\bar{y}_{st}^{*}) = \sum_{h=1}^{L} P_h^2 \left[\frac{1}{n_h} S_{yh}^2 + \frac{W_{h2}(k_h - 1)}{n_h} S_{yh(2)}^2 \right]. \tag{3.2}$$

Assuming that the coefficient of variation in each stratum is known, so Equation (3.1) becomes

$$\bar{y}_{st}^{**} = \sum_{h=1}^{L} P_h \bar{y}_h^{**}, \tag{3.3}$$

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

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

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

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

Now improved estimator becomes:

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

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

$$MSE(\bar{y}_{st}^{**}) = \sum_{h=1}^{L} P_h^2 \left[(1 - B_{1h}) \frac{S_{yh}^2}{n_h} + (1 - 2B_{2h}) \frac{W_{h2}(k_h - 1)}{n_h} S_{yh(2)}^2 \right], \tag{3.4}$$

where

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

By Equations (3.2) and (3.4), we see that $MSE(\bar{y}_{st}^{**}) < V(\bar{y}_{st}^{*})$, if

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

Using the value of B_{1h} and B_{2h} and using the assumption $S_{yh}^2 \cong S_{yh(2)}^2$, we get

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

This indicates that $\bar{y}_{st}^{**}$ is always more efficient than $\bar{y}_{st}^{*}$.
Assuming (Y) as a quantitative sensitive variable, Diana et al. (2014) have made some modifications in Hansen and Hurwitz (1946) estimator. They assumed that one group of people give direct truthful response on first call and the non-respondent group gives scrambled response on second call. They have considered the linear combination scrambled response model that was earlier defined by Diana and Perri (2010). Now we assume that (Y_h) is a sensitive quantitative variable for all h. We use a slightly modified version of Diana et al. (2014) mean estimator in stratified random sampling.
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 $(\sigma_{Ah}^2, \sigma_{Bh}^2)$ in the h^{th} stratum such that:

$$Z_h = A_h Y_h + B_h, \tag{3.5}$$

where $E_R(Z_h) = \mu_{Ah}Y_h + \mu_{Bh}$ with variance

$$V_R(Z_h) = \sigma_{Ah}^2 Y_h^2 + \sigma_{Bh}^2 \quad \text{for } (h = 1, 2, ..., L). \tag{3.6}$$

Here (E_R, V_R) are expectation and variance with respect to randomization device. It is assumed that the interviewer is completely unaware of the number generated by respondents from the scrambling distribution. This assumption increases confidentiality of respondents. Let $\hat{y}_{hi}$ be the transformed randomized response of the i^{th} unit in the h^{th} stratum, whose expectation under randomization mechanism coincides with response y_{hi} 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

$$V_R(\hat{y}_{hi}) = \frac{\sigma^2_{Ah} Y^2_{hi} + \sigma^2_{Bh}}{\mu^2_{Ah}} = \phi_{hi}. \tag{3.7}$$

We propose an estimator in stratified random sampling as:

$$\hat{\bar{y}}^*_{st} = \sum_{h=1}^{L} P_h \hat{\bar{y}}^*_h, \tag{3.8}$$

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

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

where

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

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

$$V(\hat{\bar{y}}^*_{st}) = \sum_{h=1}^{L} P_h^2 \left[\frac{1}{n_h} S^2_{yh} + \frac{W_{h2}(k_h - 1)}{n_h} S^2_{yh(2)} + \frac{k_h}{n_h N_h} S^2_{hr} \right]. \tag{3.9}$$

where $S^2_{hr} = \sum_{i=1}^{N_{h2}} \phi_{hi}$.
From (3.2) and (3.9), 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. Therefore by keeping confidentiality at fixed level we try to improve efficiency of the estimator. 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 Diana et al. (2014).
The proposed estimator is:

$$\hat{\bar{y}}^{**}_{st} = \sum_{h=1}^{L} P_h \hat{\bar{y}}^{**}_h, \tag{3.10}$$

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

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

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

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

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

$$\hat{\bar{y}}_{st}^{**} = \sum_{h=1}^{L} P_h \left[1 + \left\{1 + \frac{n_{h2}}{n_h}(k_h - 1)\right\}\frac{C_{yh}^2}{n_h} + \frac{k_h}{n_h N_h}\frac{S_{hr}^2}{\bar{Y}_h^2}\right]^{-1} \hat{\bar{y}}_h^{*}. \tag{3.11}$$

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

$$Bias(\hat{\bar{y}}_{st}^{**}) \cong -\sum_{h=1}^{L} P_h \left[\left\{1 + W_{h2}(k_h - 1)\right\}\frac{C_{yh}^2}{n_h} + \frac{k_h}{n_h N_h}\frac{S_{hr}^2}{\bar{Y}_h^2}\right]\bar{Y}_h \tag{3.12}$$

and

$$\begin{aligned} MSE(\hat{\bar{y}}_{st}^{**}) \cong \sum_{h=1}^{L} P_h^2 \Bigg[& (1 - A_h^*)\frac{S_{yh}^2}{n_h} + (1 - 2B_h^*)\frac{W_{h2}(k_h - 1)}{n_h} S_{yh2}^2 \\ & + \frac{k_h}{n_h N_h} S_{hr}^2 \left\{1 - \frac{k_h}{n_h N_h}\frac{S_{hr}^2}{\bar{Y}_h^2}\right\}\Bigg], \end{aligned} \tag{3.13}$$

where

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

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

By (3.9) and (3.13), we see that $MSE(\hat{\bar{y}}_{st}^{**}) < MSE(\hat{\bar{y}}_{st}^{*})$, if

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

where the last term is definitely positive. Now putting values of A_h^* and B_h^* and assuming that $\frac{S_{yh}^2}{n_h} \cong \frac{S_{yh(2)}^2}{n_h}$, we get

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

This shows that $\hat{\bar{y}}^{**}$ is always more efficient than $\hat{\bar{y}}^{*}$.

The generalized ratio and regression type estimators using Khare and Kumar (2011) estimator under stratified two-phase sampling scheme in case of complete and incomplete information on x_h are given by:

$$t_{st1} = \sum_{h=1}^{L} P_h t_{h1}, \quad t_{st2} = \sum_{h=1}^{L} P_h t_{h2}, \quad t_{(lr)st1} = \sum_{h=1}^{L} P_h t_{(lr)h1}, \; t_{(lr)st2} = \sum_{h=1}^{L} P_h t_{(lr)h2},$$

where

$$t_{h1} = \hat{\bar{y}}_h^{**}\left(\frac{\bar{x}_h^*}{\acute{\bar{x}}_h}\right)^{a_{h1}}, \quad t_{h2} = \hat{\bar{y}}_h^{**}\left(\frac{\bar{x}_h}{\acute{\bar{x}}_h}\right)^{a_{h2}}, \quad t_{(lr)h1} = \hat{\bar{y}}_h^{**} + b_h^{**}\left(\acute{\bar{x}}_h - \bar{x}_h^*\right),$$

$$t_{(lr)h2} = \hat{\bar{y}}_h^{**} + b_h^{*}\left(\acute{\bar{x}}_h - \bar{x}_h\right).$$

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.

Now different members of these generalized ratio and regression type estimators for different values of constants involved are obtained. 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:

$$t_{st3} = \sum_{h=1}^{L} P_h t_{h3}, \; t_{st4} = \sum_{h=1}^{L} P_h t_{h4}, \; t_{st5} = \sum_{h=1}^{L} P_h t_{h5}, \text{ and } t_{st6} = \sum_{h=1}^{L} P_h t_{h6}$$

where

$$t_{h3} = \hat{\bar{y}}_h^{**}\left(\frac{\acute{\bar{x}}_h}{\bar{x}_h^*}\right), \; t_{h4} = \hat{\bar{y}}_h^{**}\left(\frac{\acute{\bar{x}}_h}{\bar{x}_h}\right), \; t_{h5} = \hat{\bar{y}}_h^{**}\left(\frac{\bar{x}_h^*}{\acute{\bar{x}}_h}\right), \quad \text{and } t_{h6} = \hat{\bar{y}}_h^{**}\left(\frac{\bar{x}_h}{\acute{\bar{x}}_h}\right).$$

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.

$$T_{st3} = \sum_{h=1}^{L} P_h T_{h3},\ T_{st4} = \sum_{h=1}^{L} P_h T_{h4},\ T_{st5} = \sum_{h=1}^{L} P_h T_{h5},\ T_{st6} = \sum_{h=1}^{L} P_h T_{h6},$$

$$T_{(lr)st1} = \sum_{h=1}^{L} P_h T_{(lr)h1},\ \text{and}\ T_{(lr)st2} = \sum_{h=1}^{L} P_h T_{(lr)h2},$$

where

$$T_{h3} = \hat{\bar{y}}_h^* \left(\frac{\acute{\bar{x}}_h}{\bar{x}_h^*}\right),\quad T_{h4} = \hat{\bar{y}}_h^* \left(\frac{\acute{\bar{x}}_h}{\bar{x}_h}\right),\quad T_{h5} = \hat{\bar{y}}_h^* \left(\frac{\bar{x}_h^*}{\acute{\bar{x}}_h}\right),\quad T_{h6} = \hat{\bar{y}}_h^* \left(\frac{\bar{x}_h}{\acute{\bar{x}}_h}\right),$$

$$T_{(lr)h1} = t_{lrh1} = \hat{\bar{y}}_h^* + b_h^* \left(\acute{\bar{x}}_h - \bar{x}_h^*\right),\quad \text{and}\ T_{(lr)h2} = \hat{\bar{y}}_h^* + b_h^* \left(\acute{\bar{x}}_h - \bar{x}_h\ \right).$$

3.4 The Mean Squared Errors

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

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

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:

$$\begin{aligned} t_{h1} &= \hat{\bar{y}}_h^{**}\left(\frac{\bar{x}_h^*}{\acute{\bar{x}}_h}\right)^{a_{h1}} \quad = k_{h1}\hat{\bar{y}}_h^{*}\left(\frac{\bar{x}_h^*}{\acute{\bar{x}}_h}\right)^{a_{h1}} \\ &= k_{h1}\bar{Y}_h(1+\hat{e}_{0h}^*)\left[\frac{(1+e_{1h}^*)}{(1+\acute{e}_{1h})}\right]^{a_{h1}} \\ t_{h1}-\bar{Y}_h &= (k_{h1}-1)\bar{Y}_h + k_{h1}\bar{Y}_h\Bigg[\hat{e}_{0h}^* + a_{h1}e_{1h}^* - a_{h1}\acute{e}_{1h} + \frac{a_{h1}(a_{h1}-1)}{2}e_{1h}^{*2} \\ &+ \frac{a_{h1}(a_{h1}+1)}{2}\acute{e}_{1h}^2 + a_{h1}e_{1h}^*\hat{e}_{0h}^* - a_{h1}\acute{e}_{1h}\hat{e}_{0h}^* - a_{h1}^2\acute{e}_{1h}\hat{e}_{1h}^*\Bigg]. \end{aligned}$$

Squaring and neglecting higher order terms, we have

$$\begin{aligned} (t_{h1}-\bar{Y}_h)^2 \cong (k_{h1}-1)^2\bar{Y}_h^2 + \bar{Y}_h^2\Bigg[k_{h1}^2\Bigg\{\hat{e}_{0h}^{*2} + a_{h1}^2e_{1h}^{*2} + a_{h1}^2\acute{e}_{1h}^2 - 2a_{h1}e_{1h}^*\acute{e}_{1h} \\ - 2a_{h1}\hat{e}_{0h}^*\acute{e}_{1h} + 2a_{h1}\hat{e}_{0h}^*e_{1h}^*\Bigg\} + k_{h1}(k_{h1}-1)\Bigg\{a_{h1}(a_{h1}-1)e_{1h}^{*2} \\ - 2a_{h1}^2e_{1h}^*\acute{e}_{1h} + a_{h1}(a_{h1}+1)\acute{e}_{1h}^2 + 2a_{h1}\hat{e}_{0h}^*e_{1h}^* - 2a_{h1}\hat{e}_{0h}^*\acute{e}_{1h}\Bigg\}\Bigg]. \end{aligned}$$

Taking expectation, we get MSE of t_{h1}

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

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

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

where $k_{h1} \cong 1 - B_h^*$ and $B_h^* = \frac{C_{yh}^2}{n_h}\{1 + W_{h2}(k_h-1)\} + \frac{k_h}{n_hN_h}\frac{S_{hr}^2}{\bar{Y}_h^2}$.
Similarly $k_{h1}^2 \cong 1 - 2\left[\frac{C_{yh}^2}{n_h}\Bigg\{1 + W_{h2}(k_h-1)\Bigg\} + \frac{k_h}{n_hN_h}\frac{S_{hr}^2}{\bar{Y}_h^2}\right] \cong (1-2B_h^*)$ and so on.

Substituting these results in $MSE(t_{h1})$, we get:

$$MSE(t_{h1}) \cong (1-2B_h^*)V(\hat{\bar{y}}_h^*) + \bar{Y}_h^2\Bigg[a_{h1}\Bigg\{a_{h1}-(3a_{h1}-1)B_h^*\Bigg\}\Bigg\{A_{h1}C_{xh}^2 + A_{h3}C_{xh(2)}^2\Bigg\} + 2a_{h1}(1-3B_h^*)\Bigg\{(A_{h1}C_{yxh}+A_{h3}C_{yxh(2)}\Bigg\}\Bigg]$$

or

$$MSE(t_{st1}) \cong \sum_{h=1}^{L}P_h^2\Bigg[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + a_{h1}\Bigg\{a_{h1}-(3a_{h1}-1)B_h^*\Bigg\}R_h^2\Bigg\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\Bigg\} + 2a_{h1}(1-3B_h^*)R_h\Bigg\{A_{h1}S_{yxh}+A_{h3}S_{yxh(2)}\Bigg\}\Bigg], \quad (3.14)$$

where

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

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

$$MSE(t_{st2}) \cong \sum_{h=1}^{L}P_h^2\Bigg[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + A_{h1}a_{h2}\Bigg\{(a_{h2}-(3a_{h2}-1)B_h^*)R_h^2S_{xh}^2 + 2R_h(1-3B_h^*)S_{yxh}\Bigg\}\Bigg]. \quad (3.15)$$

The optimum values of a_{h1} and a_{h2} , are given by

$$a_{h1(opt)} = -\left[\frac{B_h^*}{2(1-3B_h^*)} + \frac{A_{h1}S_{yxh}+A_{h3}S_{yxh(2)}}{R_h\{A_{h1}S_{xh}^2+A_{h3}S_{xh(2)}^2\}}\right]$$

and

$$a_{h2(opt)} = -\left[\frac{B_h^*}{2(1-3B_h^*)} + \frac{S_{yxh}}{R_hS_{xh}^2}\right].$$

Using $a_{h1(opt)}$ and $a_{h2(opt)}$ in (3.14) and (3.15) respectively, we get minimum MSE of t_{st1} and t_{st2},

$$MSE(t_{st1})_{min} \cong \sum_{h=1}^{L} P_h^2 \Bigg[(1-2B_h^*)V(\hat{\bar{y}}_h^*) - (1-3B_h^*)\frac{\{A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\}^2}{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2} - \frac{B_h^2R_h^2\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\}}{4(1-3B_h^*)} - B_hR_h\Big\{A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\Big\}\Bigg] \tag{3.16}$$

and

$$MSE(t_{st2})_{min} \cong \sum_{h=1}^{L} P_h^2 \Bigg[(1-2B_h^*)V(\hat{\bar{y}}_h^*) - (1-3B_h^*)\frac{A_{h1}S_{yxh}^2}{S_{xh}^2} - \frac{B_h^2R_h^2A_{h1}S_{xh}^2}{4(1-3B_h^*)} - B_hR_hA_{h1}S_{yxh}\Bigg]. \tag{3.17}$$

The MSE of t_{st3} and t_{st4} can be obtained by putting $a_{h1} = a_{h2} = -1$ in (3.14) and (3.15) respectively

$$MSE(t_{st3}) \cong \sum_{h=1}^{L} P_h^2 \Bigg[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + (1-4B_h^*)R_h^2\Big\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\Big\} - 2(1-3B_h^*)R_h\Big\{A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\Big\}\Bigg] \tag{3.18}$$

and

$$MSE(t_{st4}) \cong \sum_{h=1}^{L} P_h^2 \Bigg[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + A_{h1}\Big\{(1-4B_h^*)R_h^2S_{xh}^2 - 2(1-3B_h^*)R_hS_{yxh}\Big\}\Bigg]. \tag{3.19}$$

To find the MSE of t_{st5} and t_{st6} we put $a_{h1} = a_{h2} = 1$ in (3.14) and (3.15), we get:

$$MSE(t_{st5}) \cong \sum_{h=1}^{L} P_h^2 \Bigg[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + (1-2B_h^*)R_h^2\Bigg\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\Bigg\} + 2(1-3B_h^*)R_h\Bigg\{A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\Bigg\}\Bigg] \quad (3.20)$$

and

$$MSE(t_{st6}) \cong \sum_{h=1}^{L} P_h^2 \Bigg[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + A_{h1}\Bigg\{(1-2B_h^*)R_h^2S_{xh}^2 + 2(1-3B_h^*)R_hS_{yxh}\Bigg\}\Bigg]. \quad (3.21)$$

The mean square errors of regression type estimator in both cases are given by:

$$MSE(t_{(lr)st1}) \cong \sum_{h=1}^{L} P_h^2 \Bigg[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + \beta_h^2\Bigg\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\Bigg\} - 2\beta_h(1-B_h^*)\Bigg\{(A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\Bigg\}\Bigg] \quad (3.22)$$

and

$$MSE(t_{(lr)st2}) = \sum_{h=1}^{L} P_h^2 \Bigg[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + A_{h1}\Bigg\{\beta_h^2S_{xh}^2 - 2\beta_h(1-B_h^*)S_{yxh}\Bigg\}\Bigg]. \quad (3.23)$$

To obtain mean square errors of T_{st3}, T_{st4}, T_{st5}, T_{st6}, $T_{(lr)st1}$ and $T_{(lr)st2}$, put $k_{h1} = 1$ in Equations (3.18), (3.19), (3.20),(3.21),(3.22) and (3.23) respectively as follows;

$$MSE(T_{st3}) \cong \sum_{h=1}^{L} P_h^2 \Bigg[V(\hat{\bar{y}}_h^*) + R_h^2\Bigg\{A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2\Bigg\} - 2R_h\Bigg\{A_{h1}S_{yxh} + A_{h3}S_{yxh(2)}\Bigg\}\Bigg], \quad (3.24)$$

$$MSE(T_{st4}) \cong \sum_{h=1}^{L} P_h^2 \left[(1-2B_h^*)V(\hat{\bar{y}}_h) + A_{h1}\left\{R_h^2 S_{xh}^2 - 2R_h S_{yxh}\right\}\right], \tag{3.25}$$

$$\begin{aligned} MSE(T_{st5}) \cong \sum_{h=1}^{L} P_h^2 \Bigg[& V(\hat{\bar{y}}_h^*) + R_h^2 \left\{A_{h1} S_{xh}^2 + A_{h3} S_{xh(2)}^2\right\} \\ & + 2R_h \left\{A_{h1} S_{yxh} + A_{h3} S_{yxh(2)}\right\}\Bigg], \end{aligned} \tag{3.26}$$

$$MSE(T_{st6}) \cong \sum_{h=1}^{L} P_h^2 \left[(1-2B_h^*)V(\hat{\bar{y}}_h^*) + A_{h1}\left\{R_h^2 S_{xh}^2 + 2R_h S_{yxh}\right\}\right], \tag{3.27}$$

$$\begin{aligned} MSE(T_{(lr)st1}) \cong \sum_{h=1}^{L} P_h^2 \Bigg[& V(\hat{\bar{y}}_h^*) + \beta_h^2 \left\{A_{h1} S_{xh}^2 + A_{h3} S_{xh(2)}^2\right\} \\ & - 2\beta_h \left\{A_{h1} S_{yxh} + A_{h3} S_{yxh(2)}\right\}\Bigg] \end{aligned} \tag{3.28}$$

and

$$MSE(T_{(lr)st2}) \cong \sum_{h=1}^{L} P_h^2 \left[V(\hat{\bar{y}}_h^*) + A_{h1}\left\{\beta_h^2 S_{xh}^2 - 2\beta_h(1-B_h^*) S_{yxh}\right\}\right]. \tag{3.29}$$

3.5 Efficiency comparison

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

Condition (i)

By (3.9) and (3.14), $MSE(t_{st1}) < V(\hat{\bar{y}}_h^*)$ if

$$\sum_{h=1}^{L} P_h^2 \Big[-2B_h^* V(\hat{\bar{y}}_h^*) + a_{h1} \Big\{ a_{h1} - (3a_{h1} - 1)B_h^* \Big\} R_h^2 \Big\{ A_{h1} S_{xh}^2 + A_{h3} S_{xh(2)}^2 \Big\} + 2a_{h1}(1 - 3B_h^*) R_h \Big\{ A_{h1} S_{yxh} + A_{h3} S_{yxh(2)} \Big\} \Big] < 0.$$

Condition (ii)

By (3.9) and (3.15), $MSE(t_{st2}) < V(\hat{\bar{y}}_h^*)$ if

$$\sum_{h=1}^{L} P_h^2 \Big[-2B_h^* V(\hat{\bar{y}}_h^*) + A_{h1} a_{h2} \Big\{ a_{h2} - (3a_{h2} - 1)B_h^* \Big\} R_h^2 \Big\{ S_{xh}^2 + 2a_{h2}(1 - 3B_h^*) R_h S_{yxh} \Big\} < 0.$$

Condition (iii)

By (3.18) and (3.24), $MSE(t_{st3}) < MSE(T_{st3})$ if

$$\sum_{h=1}^{L} P_h^2 \Big[-B_h^* V(\hat{\bar{y}}_h^*) - 2B_h^* R_h^2 \Big\{ A_{h1} S_{xh}^2 + A_{h3} S_{xh(2)}^2 \Big\} + 3B_h^* R_h \Big\{ A_{h1} S_{yxh} + A_{h3} S_{yxh(2)} \Big\} < 0.$$

Condition (iv)

By (3.19) and (3.25), $MSE(t_{st4}) < MSE(T_{st4})$ if

$$\sum_{h=1}^{L} P_h^2 \Big[-B_h^* V(\hat{\bar{y}}_h^*) - A_{h1} B_h^* \Big\{ 2R_h^2 S_{xh}^2 - 3B_h^* R_h S_{yxh} \Big\} \Big] < 0.$$

Condition (v)

By (3.20) and (3.26), $MSE(t_{st5}) < MSE(T_{st5})$ if

$$-\sum_{h=1}^{L} P_h^2 B_h^* \Big[V(\hat{\bar{y}}_h^*) + R_h^2 \Big\{ A_{h1} S_{xh}^2 + A_{h3} S_{xh(2)}^2 \Big\} + 3R_h \Big\{ A_{h1} S_{yxh} + A_{h3} S_{yxh(2)} \Big\} \Big] < 0.$$

Condition (vi)

By (3.21) and (3.27), $MSE(T_{st6}) < MSE(T_{st6})$ if

$$-\sum_{h=1}^{L} P_h^2 B_h^* \left[V(\hat{\bar{y}}_h^*) + A_{h1}\left\{ R_h^2 S_{xh}^2 + 3R_h S_{yxh} \right\}\right] < 0.$$

Condition (vii)

By (3.14) and (3.15), $MSE(t_{st1}) < MSE(t_{st2})$ if

$$\sum_{h=1}^{L} P_h^2 \left[\left\{ a_{h1}\left(a_{h1} - (3a_{h1}-1)B_h^* \right) - a_{h2}\left(a_{h2} - (3a_{h2}-1)B_h^* \right)\right\} R_h^2 \left\{ A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2 \right\} + 2\left(a_{h1}-a_{h2}\right)\left(1-3B_h^*\right)R_h \left\{ A_{h1}S_{yxh} + A_{h3}S_{yxh(2)} \right\}\right] < 0.$$

Condition (viii)

By (3.14) and (3.18), $MSE(t_{st1}) < MSE(t_{st3})$ if

$$\sum_{h=1}^{L} P_h^2 \left[\left\{ a_{h1}\left(a_{h1} - (3a_{h1}-1)B_h^* \right) - (1-4B_h^*)\right\} R_h^2 \left\{ A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2 \right\} + 2\left(a_{h1}+1\right)\left(1-3B_h^*\right)R_h \left\{ A_{h1}S_{yxh} + A_{h3}S_{yxh(2)} \right\}\right] < 0.$$

Condition (ix)

By (3.15) and (3.19), $MSE(t_{st2}) < MSE(t_{st4})$ if

$$\sum_{h=1}^{L} P_h^2 A_{h1} \left[\left\{ a_{h2}\left(a_{h2} - (3a_{h2}-1)B_h^* \right) - (1-4B_h^*)\right\} R_h^2 S_{xh}^2 + 2\left(a_{h2}+1\right)\left(1-3B_h^*\right)R_h S_{yxh}\right] < 0.$$

Condition (x)

By (3.14) and (3.20), $MSE(t_{st1}) < MSE(t_{st5})$ if

$$\sum_{h=1}^{L} P_h^2 \Bigg[\Bigg\{ a_{h1}\Bigg(a_{h1} - (3a_{h1}-1)B_h^* \Bigg) - (1-2B_h^*) \Bigg\} R_h^2 \Bigg\{ A_{h1}S_{xh}^2 + A_{h3}S_{xh(2)}^2 \Bigg\}$$
$$+\, 2\,(a_{h1}-1)\,(1-3B_h^*)R_h \Bigg\{ A_{h1}S_{yxh} + A_{h3}S_{yxh(2)} \Bigg\}\Bigg] < 0.$$

Condition (xi)

By (3.15) and (3.21), $MSE(t_{st2}) < MSE(t_{st6})$ if

$$\sum_{h=1}^{L} P_h^2 A_{h1} \Bigg[\Bigg\{ a_{h1}\Bigg(a_{h2} - (3a_{h2}-1)B_h^* \Bigg) - (1-2B_h^*) \Bigg\} R_h^2 S_{xh}^2$$
$$+\, 2\,(a_{h2}-1)\,(1-3B_h^*)R_h S_{yxh} < 0.$$

Condition (xii)

By (3.22) and (3.28), $MSE(t_{(lr)st1}) < MSE(T_{(lr)st1})$ if

$$\sum_{h=1}^{L} P_h^2 \Bigg[-V(\hat{\bar{y}}_h^*) + \beta_h \Bigg\{ A_{h1}S_{yxh} + A_{h3}S_{yxh(2)} \Bigg\}\Bigg] < 0.$$

Condition (xiii)

By (3.23) and (3.29), $MSE(t_{(lr)st2}) < MSE(T_{(lr)st2})$ if

$$\sum_{h=1}^{L} P_h^2 \Bigg[-V(\hat{\bar{y}}_h^*) + \beta_h A_{h1} S_{yxh} \Bigg] < 0.$$

3.6 Numerical Study

We use the following data sets for efficiency comparison.

Population 1(Source[1])

Y: County-wise number of Non employer establishment .

X: County-wise number of Non farm establishment.

We take five states of USA (Kansas, Iowa, Kentucky, Indiana, Illnois), having different number of counties, as strata. Assuming different non-response rate in different strata. Information for all strata are given in Table 1.

Table 3.1: 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 Table3.2 .

Table 3.2: PRE of diferrent estimators using Data 1:

Estimators	2,2,1.5,1.5,1.5	1.5,1.5,1.5,2,2	2.5,2.5,2,2,2	3,3,2,2,2
$\bar{y}^*_{st}$	100.00	100.00	100.00	100.00
$\bar{y}^{**}_{st}$	117.18	117.73	117.60	117.63
$\hat{\bar{y}}^*_{st}$	100.00	100.00	100.00	100.00
$\hat{\bar{y}}^{**}_{st}$	118.63	119.99	120.16	120.00
t_{1st}	437.58	441.89	477.11	480.34
t_{2st}	382.23	365.44	376.98	375.06
t_{3st}	430.05	431.33	450.69	458.00
t_{4st}	372.87	355.81	357.68	356.68
T_{3st}	255.49	249.00	248.46	252.40
T_{4st}	227.57	211.55	205.86	206.31
$t_{lr1(st)}$	369,78	390.32	401.36	406.32
$t_{lr2(st)}$	344.32	330.81	329.84	329.01
$T_{lr1(st)}$	261.76	254.63	254.06	256.26
$T_{lr2(st)}$	232.24	215.37	209.38	209.86

The PRE in first two rows are calculate on the w.r.t of $\bar{y}^*_{st}$ and the remaining are calculated w.r.t $\hat{\bar{y}}^*_{st}$. Table 3.2 shows that the generalized ratio and regression estimators perform better for all combination of k. The relative efficiency tend to increase as we move from left to right of the Table 3.2 so we can say for the combination of k with larger values tend to yield more efficient result. The numerical values of conditions are obtain in Table 3.3.

Table 3.3: Condition values using Data 1:

	$k_h(h = 1, ..., 4.)$	$k_h(h = 1, ..., 4.)$	$k_h(h = 1, ..., 4.)$	$k_h(h = 1, ..., 4.)$
Conditions	*2,2,1.5,1.5,1.5*	*1.5,1.5,1.5,2,2*	*2.5,2.5,2,2,2*	*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)	$-4538.35 < 0$	$-3114.98 < 0$	$-2766.72 < 0$	$-2529.706 < 0$

All conditions are satisfied for Data 1.

Population 2(Source [2])

Y: Price of Diamond

X: Depth of diamond

We divide the data into four groups according to clarity of diamond i.e. IF ("internally flawless"), VVS1 ("very very slightly imperfect), VVS2(very very slightly imperfect type 2) VS1(very slightly imperfects) having different number of stones, as strata. Assuming different non-response rate in different strata. Information for all strata are given in Table 3.4.

Table 3.4: 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	0.085	0.091
2	116.3	10.53	-20.23	-1.42	-0.255	-0.578	0.457	0.490	0.071	0.093
3	488.8	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 3.5: 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	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

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 3.5 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 3.6

Table 3.6: Condition values using Data 2

	k_h	k_h	k_h	k_h	k_h
Conditions	2.5,2.5,2,2	1.5,1.5,2,2	2.5,2.5,3,3	2.5,3,3,3.5	4,4,3.5,3.5
(i)	$-0.4863 < 0$	$-0.4245 < 0$	$-0.7655 < 0$	$-0.8574 < 0$	$-1.3606 < 0$
(ii)	$-0.4615 < 0$	$-0.4082 < 0$	$-0.7285 < 0$	$-0.8206 < 0$	$-1.3068 < 0$
(v)	$-0.3939 < 0$	$-0.3414 < 0$	$-0.6600 < 0$	$-0.7525 < 0$	$-1.2351 < 0$
(vi)	$-0.3958 < 0$	$-0.3425 < 0$	$-0.6635 < 0$	$-0.7558 < 0$	$-1.2424 < 0$
(vii)	$-0.002 < 0$	$-0.0015 < 0$	$-0.0055 < 0$	$-0.0057 < 0$	$-0.0086 < 0$
(x)	$-0.0215 < 0$	$-0.0201 < 0$	$-0.0274 < 0$	$-0.0274 < 0$	$-0.0316 < 0$
(xi)	$-0.0142 <$	$-0.0142 < 0$	$-0.0129 < 0$	$-0.0124 < 0$	$-0.0115 < 0$
(xii)	$-0.3912 < 0$	$-0.3385 < 0$	$-0.6562 < 0$	$-0.7481 < 0$	$-1.2305 < 0$
(xiii)	$-0.3929 < 0$	$-0.3396 < 0$	$-0.6595 < 0$	$-0.7515 < 0$	$-1.2375 < 0$

All conditions satisfied for Data 2.

3.7 Conclusion

In this chapter we proposed an estimator in stratified random sampling under proportional allocation using Diana et al. (2014) estimator. To improve efficiency of the estimators co-efficient of variation of the study character is used by assuming that it is known in all strata. Table 3.2 and Table 3.5 show that for fixed sample sizes $(\acute{n}_h, n_h)$ in all strata, the estimator $\hat{\hat{y}}_{st}^{**}$ is more efficient than Diana et al. (2014) estimator $\hat{\hat{y}}_{st}^{*}$ for all $k_h(h = 1, 2, 3, 4)$ for both data sets. Further generalized ratio type estimators t_{1st}, t_{2st} are more efficient than other estimators. In this paper we have discussed only linear scrambled response model used by **?** 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.

Chapter 4

Estimation of Population Mean in Ranked Set Sampling

4.1 Introduction

Ranked set sampling basically includes, selecting m samples of size m from a specified population. Ranked each sample with some expert judgment without measuring them (it may be ranked on bases of some auxiliary variable). Retain the smallest unit from 1st sample and retain the other, select the second smallest unit from the second sample and so on. Continue the process until m order units are measured to obtain one cycle. repeat the process r time to obtain rm units. In this chapter an estimator for finite population mean is proposed assuming that non-response is due to sensitivity of the study character. We use simple random SRSWR to select to select a sample of size n on first call. To obtain a sub-sample on second call, we select $\acute{n}_2$ samples of size $\acute{n}_2$ from a n_2 non-respondents. Ranked each $\acute{n}_2$ units, using information obtain on first call, without measuring them. Retain the smallest unit from 1st sample and retain the other, select the second smallest unit from the second sample and so on. Continue the process until m

order units are measured to obtain $\acute{n}_2$ units. It also uses Extreme Ranked Set Sampling(ERSS) and Median ranked set sampling (MRSS) to subsample the non-respondents. Expressions for variances of different estimators are derived. Use of RSS improves efficiency and use of RRT improves confidentiality so in this way we can obtain these twin objectives at same time therefore proposed estimators perform better than Diana and Perri (2011) estimator, in term of accuracy, and Bouza (2010) estimator in term of confidentiality. A Monte carlo experiment is carried out to see the efficiency of proposed estimators.

4.2 Estimation of mean Using SRSWR in Non-response

Let $U = (U_1, U_2, U_3, ..., U_N)$ be a finite population of size N and y_i be the observed values of the study variable y on the i^{th} unit. We select a sample of size n by using SRSWR. Now suppose that from n sampling units only n_1 units respond on first call and n_2 units don't respond. Consequently whole population divides into two groups U_1(respondents)and U_2 (non-respondent). So we select a sub-sample of size $\acute{n}_2 = \frac{n_2}{k}(k > 1)$ from n_2 non-responding units by using SRSWR. **?** estimator in SRSWR, is given by

$$\bar{y}^*_{srs} = w_1\bar{y}_1 + w_2\acute{\bar{y}}_2, \tag{4.1}$$

where $\bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i$ and $\acute{\bar{y}}_2 = \frac{1}{\acute{n}_2}\sum_{i=1}^{\acute{n}_2} y_i$. Also $E(\bar{y}^*_{srs}) = \mu$ and variance of $\bar{y}^*_{srs}$ is:

$$V(\bar{y}^*_{srs}) = \frac{1}{n}\sigma^2 + \frac{W_2(k-1)}{n}\sigma_2^2, \tag{4.2}$$

where $W_2 = \frac{N_2}{N}$ is fraction of non-respondent in population, σ^2 and σ_2^2 are population variances of the study character for whole population and population of non-respondents respectively. If non-response is due to sensitivity of the study character then it is difficult to obtain a truthful response again on second call. Diana et al. (2014) suggested an estimator for population mean using scrambled response model to overcome this deficiency.

Let Z be the scrambled response and A and B are two independent random variables unrelated to Y with known means (μ_A, μ_B) and variances (σ_A^2, σ_B^2), such that:

$$Z = AY + B \tag{4.3}$$

where $E_R(Z) = \mu_A Y + \mu_B$ and variance of Z is $V_R(Z) = \sigma_A^2 Y^2 + \sigma_B^2$, here E_R, V_R are expectation and variance with respect to randomization device .

Let $\hat{y}_i$ be transformed scrambled response of the i^{th} unit whose expectation under randomization mechanism equals to true response y_i i.e.

$$\hat{y}_i = \frac{z_i - \mu_B}{\mu_A}, \quad E_R(\hat{y}_i) = y_i$$
$$V_R(\hat{y}_i) = \frac{\sigma_A^2 y_i^2 + \sigma_B^2}{\mu_A^2}. \tag{4.4}$$

Diana et al. (2014) proposed following estimator

$$\hat{\bar{y}}^*_{srs} = w_1 \bar{y}_1 + w_2 \acute{\hat{\bar{y}}}_2, \tag{4.5}$$

where $\acute{\hat{\bar{y}}}_2 = \frac{1}{\acute{n}_2} \sum_{i=1}^{\acute{n}_2} \hat{y}_i$. It is easy to show that $E(\hat{\bar{y}}^*_{srs}) = \mu$ using the fact that $E_R(\acute{\hat{y}}_2) = \acute{\bar{y}}_2$. The variance of $\hat{\bar{y}}^*_{srs}$, is given by

$$V(\hat{\bar{y}}^*_{srs}) = \frac{1}{n}\sigma^2 + \frac{W_2(k-1)}{n}\sigma_2^2 + \frac{W_2 k}{n}\left[\frac{\sigma_A^2}{\mu_A^2}\left\{\sigma_2^2 + \mu_2^2\right\} + \frac{\sigma_B^2}{\mu_A^2}\right]. \tag{4.6}$$

The estimator in (4.5) is better than (4.1) in terms of privacy protection but it is obvious from (4.2) and (4.6) that former is less efficient than the later. So our objective is to increase efficiency of the estimator. For this purpose we use the RSS scheme which gives more efficient result than SRSWR.

4.3 Estimation of mean using RSS to sub-sample non-respondents

Bouza (2010) introduced a procedure for selecting sub-sample $\acute{S}_{2(rss)}$ of size $\acute{n}_2$ from S_2 group having size n_2 who don't respond at first call by using RSS. The procedure consist of selecting $\acute{n}_2$ sub-samples by using SRSWR. The units are ranked accordingly with the variable closely related with variable of interest Y. We have $\acute{n}_2$ independent random samples

$$Y_{11}, Y_{12}, ...Y_{1\acute{n}_2}; Y_{21}, Y_{22}.....Y_{2\acute{n}_2}; ....; Y_{\acute{n}_2 1}, Y_{\acute{n}_2 2}, ..., Y_{\acute{n}_2\acute{n}_2}.$$

After ranking, we get;

$$Y_{(1:1)}, Y_{(1:2)}, ...Y_{(1:\acute{n}_2)}; Y_{(2:1)}, Y_{(2:2)}.....Y_{(2:\acute{n}_2)}; ....; Y_{(\acute{n}_2:1)}, Y_{(\acute{n}_2:2)}, ..., Y_{(\acute{n}_2:\acute{n}_2)},$$

where $Y_{(j:t)}$ is the j^{th} order statistics (OS) of the t^{th} sample, $j = 1, 2, ...\acute{n}_2$ and $t = 1, 2, ...\acute{n}_2$. We obtain following sample:

$$Y_{(1:1)}, Y_{(2:2)}, ...Y_{(\acute{n}_2:\acute{n}_2)}.$$

The estimate of μ_2 is made by using the estimator $\acute{\bar{y}}_{2(rss)} = \frac{1}{\acute{n}_2} \sum_{j=1}^{\acute{n}_2} Y_{(j:j)}$. Also $E(Y_{(j:j)}|n_2) = \mu_{(j)}$ where $j = 1, 2, ...\acute{n}_2$. Now the estimator introduced by Bouza (2010) using (Hansen and Hurwitz, 1946) technique , is given by

$$\bar{y}^*_{rss} = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\acute{\bar{y}}_{2(rss)}, \tag{4.7}$$

with $E(\bar{y}^*_{rss}) = \mu$ and variance

$$V(\bar{y}^*_{rss}) = \frac{\sigma^2}{n} + \frac{W_2(k-1)}{n}\sigma_2^2 - \frac{W_2 k}{n}\triangle^2_{2(M)},$$

where $\quad \triangle^2_{2(M)} = E\{\frac{1}{\acute{n}_2}\sum_{j=1}^{\acute{n}_2}\triangle^2_{2(j)}\}$

$$V(\bar{y}^*_{rss}) = V(\bar{y}^*_{srs}) - \frac{W_2 k}{n}\triangle^2_{2(M)}\,. \tag{4.8}$$

Since $V(\bar{y}^*_{rss}) < V(\bar{y}^*_{srs})$ as $\triangle^2_{2(M)} > 0$. Hence $\bar{y}^*_{rss}$ is more efficient than $\bar{y}^*_{srs}$. In some cases it is difficult to rank all units, results in large error. Detecting only some units with distinct ranks may be easier and more accurate. Keeping this point in mind Samawi and Ahmed (1996) used a RSS sampling procedure called Extreme Ranked Set Sampling (ERSS). The procedure includes identification of two extreme values $Y_{(1:j)}$ and $Y_{(\acute{n}_2:j)}$ from the j^{th}sample. The extreme ranked set sampling in case of ranked set sampling works as follow. Select $Y_{2(e:j)}$ such that:

$$Y_{2(e:j)} = \begin{cases} Y_{2(1:j)} & \text{for} \quad j = 1, \ldots\frac{\acute{n}_2}{2} \\ Y_{2(\acute{n}_2:j)} & \text{for} \quad j = \frac{\acute{n}_2}{2}+1, \ldots.\acute{n}_2, \end{cases}$$

where

$$E(Y_{2(e:j)}) = \begin{cases} \mu_{2(1)} & \text{for} \quad j = 1, \ldots\frac{\acute{n}_2}{2} \\ \mu_{2(\acute{n}_2))} & \text{for} \quad j = \frac{\acute{n}_2}{2}+1, \ldots.\acute{n}_2, \end{cases}$$

and

$$V(Y_{2(e:j)}) = \begin{cases} \sigma^2_{2(1)} & \text{for} \quad j = 1, \ldots\frac{\acute{n}_2}{2} \\ \sigma^2_{2(\acute{n}_2))} & \text{for} \quad j = \frac{\acute{n}_2}{2}+1, \ldots.\acute{n}_2, \end{cases}$$

an estimate of μ_2 is:

$$\acute{\bar{y}}_{2(erss)} = \frac{1}{\acute{n}_2}\sum_{j=1}^{\acute{n}_2} Y_{(e:j)} = \frac{Y_{2(1)} + Y_{2(\acute{n}_2)}}{2},$$

where $\quad E(\acute{\bar{y}}_{2(erss)}) = \frac{\mu_{2(1)} + \mu_{2(\acute{n}_2)}}{2} \neq \mu_2.$

Hence it is a biased estimator of μ_2. It will be unbiased if $\mu_{2(1)} = \mu_{2(\acute{n}_2)}$. It is possible only in case of symmetric distribution. Now Hansen and Hurwitz (1946) estimator in ERSS, is given by

$$\bar{y}^*_{erss} = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\acute{\bar{y}}_{2(erss)}. \tag{4.9}$$

The Bias of the estimator $\bar{y}^*_{erss}$ is given by

$$Bias(\bar{y}^*_{erss}) = W_2\frac{(\mu_{2(1)} - \mu_2) + (\mu_{2(\acute{n}_2)} - \mu_2)}{2}, \tag{4.10}$$

which is almost negligible in case of near to symmetric distribution, its variance is given by

$$V(\bar{y}^*_{erss}) = \frac{\sigma^2}{n} + \frac{W_2(k-1)}{n}\sigma_2^2 - \frac{W_2 k}{n}\triangle^2_{2(e)},$$

where $\triangle^2_{2(e)} = \dfrac{\triangle^2_{2(1)} + E\triangle^2_{2(n_2)}}{2}$. The above expression can also be written as :

$$V(\bar{y}^*_{erss}) = V(\bar{y}^*_{srs}) - \frac{W_2 k}{n}\triangle^2_{2(e)}\ . \tag{4.11}$$

ERSS will be preferred on RSS if $\triangle^2_{2(e)} > \triangle^2_{2(M)}$.
Bouza (2010) used another modification to RSS that include selecting medians of all ranked set samples. Assuming $\acute{n}_2$ as even, select $Y_{2(m:j)}$ such that:

$$Y_{2(m:j)} = \begin{cases} Y_{2(\frac{\acute{n}_2}{2}:j)} & \text{for } \ j = 1, ...\frac{\acute{n}_2}{2} \\ Y_{2(\frac{\acute{n}_2}{2}+1:j)} & \text{for } \ j = \frac{\acute{n}_2}{2}+1, ...\acute{n}_2 \end{cases}$$

where

$$E(Y_{2(m:j)}) = \begin{cases} \mu_{2(\frac{\acute{n}_2}{2})} & \text{for } \ j = 1, ...\frac{\acute{n}_2}{2} \\ \mu_{2(\frac{\acute{n}_2}{2}+1)} & \text{for } \ j = \frac{\acute{n}_2}{2}+1, ...\acute{n}_2 \end{cases}$$

and

$$V(Y_{2(m:j)}) = \begin{cases} \sigma^2_{2(\frac{\acute{n}_2}{2})} & \text{for } \ j = 1, ...\frac{\acute{n}_2}{2} \\ \sigma^2_{2(\frac{\acute{n}_2}{2}+1)} & \text{for } \ j = \frac{\acute{n}_2}{2}+1, ...\acute{n}_2, \end{cases}$$

The estimator for μ_2 using Median Ranked Set Sampling (MRSS) is;

$$\acute{\bar{y}}_{2(mrss)} = \frac{1}{\acute{n}_2}\sum_{j=1}^{\acute{n}_2} Y_{(m:j)} = \frac{Y_{2(\frac{\acute{n}_2}{2})} + Y_{2(\frac{\acute{n}_2+2}{2})}}{2},$$

where

$$E(\acute{\bar{y}}_{2(mrss)}) = \frac{\mu_{2(\frac{\acute{n}_2}{2})} + \mu_{2(\frac{\acute{n}_2+2}{2})}}{2} \neq \mu_2.$$

Hence it is a biased estimator of μ_2. It will be unbiased if $\mu_{2(\frac{\acute{n}_2}{2})} = \mu_{2(\frac{\acute{n}_2+2}{2})}$. Now Hansen and Hurwitz (1946) estimator in MRSS, is given by

$$\bar{y}^*_{mrss} = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\acute{\bar{y}}_{2(mrss)}, \tag{4.12}$$

The Bias of the estimator $\bar{y}^*_{mrss}$ is given by

$$Bias(\bar{y}^*_{mrss}) = W_2\frac{\left(\mu_{2(\frac{\acute{n}_2}{2})} - \mu_2\right) + \left(\mu_{2(\frac{\acute{n}_2+2}{2})} - \mu_2\right)}{2}, \tag{4.13}$$

which is almost negligible in case of approximately symmetric distribution, its variance, is given by

$$V(\bar{y}^*_{mrss}) = \frac{\sigma^2}{n} + \frac{W_2(k-1)}{n}\sigma_2^2 - \frac{W_2 k}{n}\triangle^2_{2(m)},$$

where $\triangle^2_{2(m)} = E\left\{\frac{\triangle^2_{2(\frac{\acute{n}_2}{2})} + \triangle^2_{2(\frac{\acute{n}_2+2}{2})}}{2}\right\}$. The above expression can also be written as :

$$V(\bar{y}^*_{mrss}) = V(\bar{y}^*_{srs}) - \frac{W_2 k}{n}\triangle^2_{2(m)}\,. \tag{4.14}$$

MRSS will be preferred over RSS if $\triangle^2_{2(m)} > \triangle^2_{2(M)}$. Also MRSS will be preferred over ERSS if $\triangle^2_{2(m)} > \triangle^2_{2(e)}$.

4.4 Proposed Estimators

When the study variable is sensitive in nature than non-response occurs due to sensitivity of the character under study, consequently the estimators in (4.7),(4.9) and (4.12) fail to estimate population mean of the study character as it is hard to find a sub-sample on second call. Taking motivation from Diana et al. (2014) estimator, we use a randomized response model in RSS, ERSS and MRSS for sub-sampling non-respondents to overcome this difficulty. From (4.3) a ranked set sampled j^{th} scrambled response in the j^{th} sample is given as follow when ranking is performed on Y:

$$Z_{[j:j]} = A_j Y_{(j:j)} + B_j, \qquad (j = 1, 2, ...\acute{n}_2), \tag{4.15}$$

where $E_R(Z_{[j:j]}) = \mu_A Y_{(j:j)} + \mu_B$ and variance of $Z_{[j:j]}$ is $V_R(Z_{[j:j]}) = \sigma_A^2 Y_{(j:j)}^2 + \sigma_B^2$, here E_R, V_R are expectation and variance with respect to randomization device . Let $\hat{y}_{[j;j]}$ be transformed scrambled response of the j^{th} unit in the j^{th} sample whose expectation under randomization mechanism equals to true response $y_{(j;j)}$.

$$\hat{y}_{[j:j]} = \frac{z_{[j:j]} - \mu_B}{\mu_A}, \quad E_R(\hat{y}_{[j:j]}) = y_{(j:j)}$$
$$V_R(\hat{y}_{[j:j]}) = \frac{\sigma_A^2 y_{[j:j]}^2 + \sigma_B^2}{\mu_A^2}. \tag{4.16}$$

The estimate of μ_2 using this technique is, $\acute{\hat{y}}_{2(rss)} = \frac{1}{\acute{n}_2} \sum_{j=1}^{\acute{n}_2} \hat{y}_{2[j:j]}$. The proposed estimator using this technique to sub-sample non-respondents is given by.

$$\hat{\bar{y}}_{rss}^{*} = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\acute{\hat{\bar{y}}}_{2(rss)}, \tag{4.17}$$

with expected value

$$\begin{aligned}
E(\hat{\bar{y}}^*_{rss}) =& E_1E_2\left[w_1\bar{y}_1 + w_2E_R\{\hat{\acute{\bar{y}}}_{2(rss)}\}\right] \\
=& E_1E_2\left[w_1\bar{y}_1 + w_2\acute{\bar{y}}_{2(rss)}\right], \text{ as } \ E_R(\hat{\acute{\bar{y}}}_{2(rss)}) = \acute{\bar{y}}_{2(rss)}, \\
=& \mu.
\end{aligned}$$

Hence $\hat{\bar{y}}^*_{rss}$ is an unbiased estimator of μ and variance of $\hat{\bar{y}}^*_{rss}$, can be derived as follow:

$$V(\hat{\bar{y}}^*_{rss}) = E_1\left[V_2\{E_R(\hat{\bar{y}}^*_{rss})\} + E_2\{V_R(\hat{\bar{y}}^*_{rss})\}\right] \tag{4.18}$$

Take

$$\begin{aligned}
E_1\left[V_2\{E_R(\hat{\bar{y}}^*_{rss})\}\right] =& E_1\left[V_2\left(\bar{y}^*_{rss}\right)\right], \\
=& \frac{\sigma^2}{n} + \frac{W_2(k-1)}{n}\sigma_2^2 - \frac{W_2k}{n}\triangle^2_{2(M)}\,.
\end{aligned} \tag{4.19}$$

Now for another part,

$$\begin{aligned}
V_R\left(\hat{\bar{y}}^*_{rss}\right) =& \frac{w_2^2}{\acute{n}_2^2}\sum_{j=1}^{\acute{n}_2} V_R\left[\hat{y}_{2[j:j]}\right] \\
=& \frac{w_2^2}{\acute{n}_2^2}\sum_{j=1}^{\acute{n}_2}\left[\frac{\sigma_A^2 y^2_{(j:j)} + \sigma_B^2}{\mu_A^2}\right] \\
E_2\{V_R(\hat{\bar{y}}^*_{rss})\} =& \frac{w_2^2}{\acute{n}_2^2}\left[\frac{\sigma_A^2(\sum_{j=1}^{\acute{n}_2}\sigma^2_{2(j)} + \sum_{j=1}^{\acute{n}_2}\mu^2_{2(j)}) + \acute{n}_2\sigma_B^2}{\mu_A^2}\right] \\
=& \frac{w_2^2}{\acute{n}_2^2}\left[\frac{\sigma_A^2\{\acute{n}_2\sigma_2^2 - \sum_{i=1}^{\acute{n}_2}\triangle^2_{2(j)} + \sum_{j=1}^{\acute{n}_2}\mu^2_{2(j)}\} + \acute{n}_2\sigma_B^2}{\mu_A^2}\right].
\end{aligned}$$

Hence

$$E_1E_2\{V_R(\hat{\bar{y}}^*_{rss})\} = \frac{W_2k}{n}\left[\frac{\sigma_A^2\{\sigma_2^2 + \mu^2_{2(M)}\} + \sigma_B^2}{\mu_A^2} - \frac{\sigma_A^2}{\mu_A}\triangle^2_{2(M)}\right], \tag{4.20}$$

where $\mu^2_{2(M)} = E\{\frac{1}{\acute{n}_2}\sum_{j=1}^{\acute{n}_2}\mu^2_{2(j)}\}$ and $\triangle^2_{2(M)}$ is defined earlier. Substituting (4.19) and (4.20) in (4.18), we get:

$$\begin{aligned}
V(\hat{\bar{y}}^*_{rss}) =& \frac{\sigma^2}{n} + \frac{W_2(k-1)}{n}\sigma^2_2 - \frac{W_2k}{n}\triangle^2_{2(M)} + \frac{W_2k}{n}\Bigg[\frac{\sigma^2_A\{\sigma^2_2+\mu^2_{2(M)}\}+\sigma^2_B}{\mu^2_A} \\
&- \frac{\sigma^2_A}{\mu^2_A}\triangle^2_{2(M)}\Bigg] \\
V(\hat{\bar{y}}^*_{rss}) =& V(\hat{\bar{y}}^*_{srs}) - \frac{W_2k}{n}\triangle^2_{2(M)}\,\theta, \qquad (4.21)
\end{aligned}$$

where $\theta = 1 + \frac{\sigma^2_A}{\mu^2_A}$ and $\theta > 1$. Hence

$$G_{Eff}(rss) = \frac{W_2k}{n}\triangle^2_{2(M)}\,\theta > 0,$$

where $G_{Eff}(rss))$ denotes the gain in efficiency due to RSS.

Taking idea used by Bouza (2010), we propose an estimator of population mean by using ERSS with scrambled response model. Because it is more easy to identify only extreme units from a sample than ranking all units. The scrambled response is given by

$$Z_{[e:j]} = A_jY_{(j:e)} + B_j, \qquad (j = 1, 2, ...\acute{n}_2), \qquad (4.22)$$

where $E_R(Z_{[e:j]}) = \mu_AY_{(e:j)} + \mu_B$ and variance of $Z_{[e:j]}$ is $V_R(Z_{[j:e]}) = \sigma^2_AY^2_{(j:e)} + \sigma^2_B$, Let $\hat{y}_{[e:j]}$ be transformed scrambled response from extreme units in the j^{th} sample whose expectation under randomization mechanism equals to true response $y_{(e:j)}$.

$$\begin{aligned}
\hat{y}_{[j:e]} &= \frac{z_{[e:j]}-\mu_B}{\mu_A}, \quad E_R(\hat{y}_{[e:j]}) = y_{(e:j)} \\
V_R(\hat{y}_{[e:]}) &= \frac{\sigma^2_Ay^2_{[e:j]}+\sigma^2_B}{\mu^2_A} = \phi_{[e:j]}. \qquad (4.23)
\end{aligned}$$

The estimate of μ_2 using this technique is:

$$\hat{\acute{\bar{y}}}_{2(erss)} = \frac{1}{\acute{n}_2}\sum_{j=1}^{\acute{n}_2}\hat{y}_{(e:j)} = \frac{\hat{Y}_{2(1)} + \hat{Y}_{2(\acute{n}_2)}}{2}, \quad \text{where}$$
$$E\{E_R(\hat{\acute{\bar{y}}}_{2(erss)})\} = E(\hat{\acute{\bar{y}}}_{2(erss)}) = \frac{\mu_{2(1)} + \mu_{2(\acute{n}_2)}}{2} \neq \mu_2.$$

In case of symmetric distribution it will be unbiased as $\mu_{2(1)} = \mu_{2(\acute{n}_2)}$. The proposed estimator using this technique to sub-sample non-respondents is given by

$$\hat{\bar{y}}^*_{erss} = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\hat{\acute{\bar{y}}}_{2(erss)}. \tag{4.24}$$

The Bias of proposed estimator is ,

$$\begin{aligned} E\left(\hat{\bar{y}}^*_{erss}\right) =& E_1E_2\left[w_1\bar{y}_1 + w_2E_R\{\hat{\acute{\bar{y}}}_{2(erss)}\}\right] \\ =& E_1E_2\left[w_1\bar{y}_1 + w_2\acute{\bar{y}}_{2(erss)}\right], \text{ as } E_R(\hat{\acute{\bar{y}}}_{2(erss)}) = \acute{\bar{y}}_{2(erss)}, \\ \Rightarrow \quad Bias(\hat{\bar{y}}^*_{erss}) =& W_2\frac{\left(\mu_{2(1)} - \mu_2\right) + \left(\mu_{2(\acute{n}_2)} - \mu_2\right)}{2}, \end{aligned} \tag{4.25}$$

which is almost negligible in case of near to symmetric distribution, its variance is given by

$$V(\hat{\bar{y}}^*_{erss}) = E_1\left[V_2\{E_R(\hat{\bar{y}}^*_{erss})\} + E_2\{V_R(\hat{\bar{y}}^*_{erss})\}\right]. \tag{4.26}$$

consider first part

$$\begin{aligned} E_1\left[V_2\{E_R\left(\hat{\bar{y}}^*_{erss}\right)\}\right] =& E_1\left[V_2(\bar{y}^*_{erss})\right], \\ =& \frac{\sigma^2}{n} + \frac{W_2(k-1)}{n}\sigma_2^2 - \frac{W_2k}{n}\triangle^2_{2(e)}. \end{aligned} \tag{4.27}$$

consider second part

$$\begin{aligned}
V_R\left(\hat{\bar{y}}^*_{erss}\right) =& \frac{w_2^2}{\acute{n}_2^2}\sum_{j=1}^{\acute{n}_2}\left[\frac{\sigma_A^2 y_{(j:e)}^2+\sigma_B^2}{\mu_A^2}\right]\\
E_2\{V_R\left(\hat{\bar{y}}^*_{erss}\right)\} =& \frac{w_2^2}{\acute{n}_2^2}\left[\frac{\sigma_A^2 E_2(\frac{\acute{n}_2}{2}y_{2(1)}^2+\frac{\acute{n}_2}{2}y_{2(\acute{n}_2)}^2)+\acute{n}_2\sigma_B^2}{\mu_A^2}\right]\\
=& \frac{w_2^2}{\acute{n}_2^2}\left[\frac{\sigma_A^2\{\frac{\acute{n}_2}{2}(\sigma_{2(1)}^2+\mu_{2(1)}^2)+\frac{\acute{n}_2}{2}(\sigma_{2(\acute{n}_2)}^2+\mu_{2(\acute{n}_2)}^2)\}+\acute{n}_2\sigma_B^2}{\mu_A^2}\right]\\
=& \frac{w_2^2}{\acute{n}_2^2}\left[\frac{\sigma_A^2\{\frac{\acute{n}_2}{2}\tau_1+\frac{\acute{n}_2}{2}\tau_{\acute{n}_2}\}+\acute{n}_2\sigma_B^2}{\mu_A^2}\right],
\end{aligned}$$

where $\tau_1=\sigma_2^2-\triangle_{2(1)}^2+\mu_{2(1)}^2$ and $\tau_{\acute{n}_2}=\sigma_2^2-\triangle_{2(\acute{n}_2)}^2+\mu_{2(\acute{n}_2)}^2$. Hence

$$\begin{aligned}
E_1E_2\{V_R(\hat{\bar{y}}^*_{erss})\} =& \frac{W_2k}{n}\left[\frac{\sigma_A^2}{\mu_A^2}\left\{\sigma_2^2-\frac{1}{2}\left(\triangle_{2(1)}^2+E_1\triangle_{2(\acute{n}_2)}^2\right)\right.\right.\\
&\left.\left.+\frac{1}{2}(\mu_{2(1)}^2+E_1\mu_{2(\acute{n}_2)}^2)\right\}+\frac{\sigma_B^2}{\mu_A^2}\right] \qquad (4.28)
\end{aligned}$$

Using (4.27) and (4.28) in (4.26) , we get

$$\begin{aligned}
V(\hat{\bar{y}}^*_{erss}) =& \frac{\sigma^2}{n}+\frac{W_2(k-1)}{n}\sigma_2^2-\frac{W_2k}{n}\left[\triangle_{2(e)}^2+\frac{\sigma_A^2}{\mu_A^2}\{\sigma_2^2-\triangle_{2(e)}^2+\mu_{2(e)}^2\}+\frac{\sigma_B^2}{\mu_A^2}\right]\\
=& \frac{\sigma^2}{n}+\frac{W_2(k-1)}{n}\sigma_2^2+\frac{W_2k}{n}\left[\frac{\sigma_A^2\{\sigma_2^2+\mu_{2(e)}^2\}+\sigma_B^2}{\mu_A^2}\right]\\
&-\frac{W_2k}{n}\triangle_{2(e)}^2\,\theta\\
=& V(\hat{\bar{y}}^*_{srs})-\frac{W_2k}{n}\triangle_{2(e)}^2\,\theta, \qquad (4.29)
\end{aligned}$$

where $\mu_{2(e)}^2=\dfrac{1}{2}(\mu_{2(1)}^2+E_1\mu_{2(\acute{n}_2)}^2)$ and $\triangle_{2(e)}^2$ is defined earlier in previous section. The gain in efficiency due to ERSS is

$$G_{Eff}(erss))=\frac{W_2k}{n}\triangle_{2(e)}^2\,\theta>0.$$

Since $\theta > 0$. ERSS will give more efficient result than RSS if:

$$\triangle^2_{2(e)} > \triangle^2_{2(M)}$$

We propose an estimator of population mean by using scrambled response model in MRSS. The scrambled response is given by

$$Z_{[m:j]} = A_j Y_{(m:j)} + B_j, \qquad (j = 1, 2, ...\acute{n}_2), \tag{4.30}$$

where $E_R(Z_{[m:j]}) = \mu_A Y_{(m:j)} + \mu_B$ and variance of $Z_{[m:j]}$ is

$$V_R(Z_{[m:j]}) = \sigma^2_A Y^2_{(m:j)} + \sigma^2_B.$$

Let $\hat{y}_{[m:j]}$ be transformed scrambled response from median units in the j^{th} sample whose expectation under randomization mechanism equals to true response $y_{(m:j)}$.

$$\hat{y}_{[m:j]} = \frac{z_{[m:j]} - \mu_B}{\mu_A}, \quad E_R(\hat{y}_{[m:j]}) = y_{(m:j)}$$
$$V_R(\hat{y}_{[m:j]}) = \frac{\sigma^2_A y^2_{[m:j]} + \sigma^2_B}{\mu^2_A} = \phi_{[m:j]}. \tag{4.31}$$

The estimate of μ_2 using this technique is:

$$\acute{\hat{\bar{y}}}_{2(mrss)} = \frac{1}{\acute{n}_2} \sum_{j=1}^{\acute{n}_2} \hat{y}_{(j:m)} = \frac{\hat{Y}_{2(\frac{\acute{n}_2}{2})} + \hat{Y}_{2(\frac{\acute{n}_2}{2}+1)}}{2}, \quad \text{where}$$
$$E\{E_R(\acute{\hat{\bar{y}}}_{2(mrss)})\} = E(\acute{\hat{\bar{y}}}_{2(mrss)}) = \frac{\mu_{2(\frac{\acute{n}_2}{2})} + \mu_{2(\frac{\acute{n}_2}{2}+1)}}{2} \neq \mu_2$$

It will be unbiased if $\mu_{2(\frac{\acute{n}_2}{2}+1)} = \mu_{2(\frac{\acute{n}_2}{2})}$ which is possible only in case of symmetric distribution. The proposed estimator using this technique to sub-sample non-respondents given bellow.

$$\hat{\bar{y}}^{*}_{mrss} = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\acute{\hat{\bar{y}}}_{2(mrss)}. \tag{4.32}$$

The Bias of proposed estimator $\hat{\bar{y}}^{*}_{mrss}$ is given by

$$\begin{aligned} E(\hat{\bar{y}}^{*}_{mrss}) =& E_1E_2\left[w_1\bar{y}_1 + w_2E_R(\acute{\hat{\bar{y}}}_{2(mrss)})\right] \\ =& E_1E_2\left[w_1\bar{y}_1 + w_2(\acute{\bar{y}}_{2(mrss)})\right], \text{ as } E_R(\acute{\hat{\bar{y}}}_{2(mrss)}) = \acute{\bar{y}}_{2(mrss)}, \\ \Rightarrow \quad Bias(\hat{\bar{y}}^{*}_{mrss}) =& W_2\frac{\left(\mu_{2(\frac{\acute{n}_2}{2})} - \mu_2\right) + \left(\mu_{2(\frac{\acute{n}_2}{2}+1)} - \mu_2\right)}{2}, \end{aligned} \tag{4.33}$$

which is almost negligible when the distribution tends to symmetric, its variance using law of total variance, is given by

$$V(\hat{\bar{y}}^{*}_{mrss}) = E_1\Bigg[V_2\{E_R(\hat{\bar{y}}^{*}_{mrss})\} + E_2\{V_R(\hat{\bar{y}}^{*}_{mrss})\}\Bigg]. \tag{4.34}$$

First part of (4.34) is:

$$E_1[V_2\{E_R(\hat{\bar{y}}^{*}_{mrss})\}] = \frac{\sigma^2}{n} + \frac{W_2(k-1)}{n}\sigma_2^2 - \frac{W_2k}{n}\{\triangle^2_{2(m)}\}. \tag{4.35}$$

Also second part of (4.34) is

$$\begin{aligned} V_R(\hat{\bar{y}}^{*}_{mrss}) =& \frac{w_2^2}{\acute{n}_2^2}\sum_{j=1}^{\acute{n}_2}\{\frac{\sigma_A^2 y^2_{(j:m)} + \sigma_B^2}{\mu_A^2}\} \\ E_2\{V_R(\hat{\bar{y}}^{*}_{mrss})\} =& \frac{w_2^2}{\acute{n}_2^2}\left[\frac{\sigma_A^2\{\frac{\acute{n}_2}{2}T_1 + \frac{\acute{n}_2}{2}T_2\} + \acute{n}_2\sigma_B^2}{\mu_A^2}\right], \end{aligned}$$

where $T_1 = \sigma_2^2 - \triangle^2_{2(\frac{\acute{n}_2}{2})} + \mu^2_{2(\frac{\acute{n}_2}{2})}$ and $T_2 = \sigma_2^2 - \triangle^2_{2(\frac{\acute{n}_2}{2}+1)} + \mu^2_{2(\frac{\acute{n}_2}{2}+1)}$. Therefore

$$E_1E_2\{V_R(\hat{\hat{y}}^*_{mrss})\} = \frac{W_2k}{n}\left[\frac{\sigma_A^2}{\mu_A^2}\left\{\sigma_2^2 - \triangle^2_{2(m)} + \mu^2_{2(m)}\right\} + \frac{\sigma_B^2}{\mu_A^2}\right] \tag{4.36}$$

where $\mu^2_{2(m)} = \frac{1}{2}(\mu^2_{2(\frac{\acute{n}_2}{2})} + E_1\mu^2_{2(\frac{\acute{n}_2}{2}+1)}$ and $\triangle^2_{2(m)}$ is defined in previous section. Using (4.35) and (4.36) in (4.34) , we get

$$\begin{aligned} V(\hat{\hat{y}}^*_{mrss}) =& \frac{\sigma^2}{n} + \frac{W_2(k-1)}{n}\sigma_2^2 + \frac{W_2k}{n}\left[\frac{\sigma_A^2\{\sigma_2^2 + \mu^2_{2(m)}\} + \sigma_B^2}{\mu_A^2}\right] \\ &- \frac{W_2k}{n}\triangle^2_{2(m)}\,\theta \\ V(\hat{\hat{y}}^*_{mrss}) =& V(\hat{\hat{y}}^*_{srs}) - \frac{W_2k}{n}\triangle^2_{2(m)}\,\theta. \end{aligned} \tag{4.37}$$

Gain in efficiency due to MRSS is

$$G_{Eff}(mrss)) = \frac{W_2k}{n}\triangle^2_{2(m)}\,\theta > 0.$$

MRSS will give more efficient result than RSS if: $\triangle^2_{2(m)} > \triangle^2_{2(M)}$.
Also MRSS will give more efficient result than ERSS if: $\triangle^2_{2(m)} > \triangle^2_{2(e)}$.

4.5 Empirical Study

To compare the efficiency of proposed estimators in RSS to the corresponding estimators in SRSWR, we conduct a simulation study. The hypothetical population consist of 1000 observations on two variables X and Y. The values of X is generated from Normal ($\mu = 0, \sigma^2 = 1$), exponential ($\lambda = 5$) and Uniform ($a = 0, b = 1$) distributions. After that Y is computed such that $Y = rX + e$, where r is coefficient of correlation between X and Y and $e \sim N(0, 1)$ is the error term. Table 1 gives relative efficiency of different estimators for the above mentioned distributions of X using different combinations of k and W_2.

The last three columns of Table 1 give relative efficiency of proposed estimators with respect to Diana et al. (2014) estimator. We can see that relative efficiency of proposed estimators tend to decreases for larger k. This means for small sub-sample size proposed estimators give greater precision. It can also be inferred from Table 1 that relative RSS perform better in case of Uniform distribution as compared to other two distributions. In all cases MRSS perform better than RSS under scrambled response model. But relative efficiency of ERSS is smaller than RSS and MRSS for all cases.

Table 4.1: Efficiency comparison values

Distribution	W_2	k	RSS(1)	ERSS(1)	MRSS(1)	RSS(2)	ERSS(2)	MRSS(2)
Normal	0.1	2	1.1250	1.0609	1.0935	1.3513	1.2867	1.3946
		4	1.1013	1.1049	1.2241	1.2122	1.2088	1.2862
		6	1.1013	1.1049	1.2241	1.2122	1.2088	1.2862
	0.2	2	1.2728	1.1763	1.2138	2.1318	1.6430	2.2981
		4	1.4389	1.2135	1.3260	1.9032	1.6060	2.1399
		6	1.3291	1.1242	1.3359	1.5354	1.2804	1.7114
	0.3	2	1.4455	1.0917	1.1437	3.1526	1.5960	3.4264
		4	1.6438	1.4535	1.5430	2.7307	1.9072	3.2724
		6	1.6769	1.2640	1.4570	2.2765	1.8958	2.7598
	0.4	2	1.6160	1.2031	1.2515	4.5716	2.1163	5.1349
		4	1.9228	1.6177	1.7300	3.4769	2.1427	4.3755
		6	2.0375	1.2540	1.4467	3.0449	1.5971	3.8519
Exponential	0.1	2	1.1170	1.0543	1.0828	1.6186	1.6780	1.6661
		4	1.1123	1.1098	1.2265	1.3107	1.5030	1.4821
		6	1.1123	1.1098	1.2265	1.3107	1.5030	1.4821
	0.2	2	1.2596	1.1517	1.1909	3.1984	1.8218	3.3736
		4	1.3659	1.1088	1.2214	2.4734	2.4532	2.7109
		6	1.3584	1.1875	1.3860	1.9270	1.7200	2.0568
	0.3	2	1.4439	1.1188	1.1682	5.2756	1.4882	5.6374
		4	1.5923	1.4473	1.5372	3.8759	2.0581	4.4753
		6	1.7389	1.3048	1.5143	3.0151	2.8765	3.4168
	0.4	2	1.6419	1.1481	1.1986	7.7175	1.6036	8.6775
		4	1.9136	1.5654	1.6774	5.0912	1.8970	5.9705
		6	1.9693	1.3389	1.5237	3.9496	1.7195	4.7537
Uniform	0.1	2	1.1138	1.0674	1.0977	1.6765	1.6903	1.7198
		4	1.1204	1.1017	1.2161	1.3577	1.5053	1.4567
		6	1.1204	1.1017	1.2161	1.3577	1.5053	1.4567
	0.2	2	1.2533	1.1726	1.2050	3.2788	1.9403	3.5413
		4	1.3812	1.1515	1.2621	2.4932	2.4591	2.7205
		6	1.3049	1.1039	1.3151	1.8869	1.7240	2.0569
	0.3	2	1.4491	1.1234	1.1726	5.5418	1.6480	6.1193
		4	1.6808	1.5261	1.6343	4.0407	2.2174	4.7350
		6	1.5325	1.2022	1.3913	2.8012	2.8716	3.3526
	0.4	2	1.6165	1.2146	1.2635	8.2002	1.6974	9.1755
		4	1.9746	1.6667	1.7758	5.5181	2.1727	6.3218
		6	2.0436	1.2798	1.4815	4.2700	1.8340	5.0032

4.6 Conclusion

This chapter presented an estimation of population mean in non-response where one can obtain twin objectives of survey sampling; (i) one is to give greater confidentiality to the respondents which results increment of response rate, (ii) another is gain in precision of estimates involve in study. By assuming that non-response is due to sensitivity of the study character we proposed three estimators using scrambled response model by selecting a ranked set sample, extreme ranked set sample and median ranked set sample. It is proved both mathematically and numerically that the estimators of population mean perform better in RSS, ERSS and MRSS than SRS.

Chapter 5

Conclusion and Suggestion

5.1 Conclusion of the study

This study includes estimation of finite population mean of sensitive study variable in presence of non-response. In case of sensitive variable, non-response (or refusal) occurs because people hesitate while giving direct response to the questions asked by interviewer. Therefore it is hard to get truthful response again on second call. We discussed two important aspects one is use of randomization technique for privacy protection, which help us in obtaining truthful response on second call, and the other is reduction in variation in estimation of finite population mean as the variance increased due to use of randomized response model. Use of randomization mechanism results increment in response rate and it is inevitable on second call to collect information from those respondents who refused to answer on first call. In this study we only used a linear scrambled response model to protect confidentiality of respondents but our concern is reduction in variance. We have made three basic attempts to improve efficiency of the estimator of finite population mean using scrambled response model to sub-sample non-respondents on second call. From Chapter 2, we conclude that the proposed estimator using known coefficient of variation of the study variable and generalized ratio and regression type estimators, constructed on the basis of the proposed estimator, perform better than Diana et al. (2014) estimator in term of efficiency for a fix level of privacy protection. In Chapter 3, it is obvious that use of stratified sampling results in gain in efficiency of the proposed estimators when population of interest is heterogeneous. The generalized ratio and regression estimators are more efficient than the corresponding simple mean estimator in both SRSWOR and stratified random sampling .This implies that use of auxiliary variable (with two-phase sampling scheme when population mean of the auxiliary variable is not known) renders us more efficiency. Chapter 4 provide that use of RSS, ERSS and MRSS with randomized response technique for sub-sampling non-respondents utilizing the information on first call for ranking purpose. It is shown both mathematically and empirically that ranked set sampling is at least as efficient as SRSWR so one should prefer to select a sample using RSS when ranking is easy and less expensive.

5.2 Suggestion for further Study

Further work can be done to improve privacy protection by using different scrambled response models i.e. Additive, Subtractive and Optional scrambled response model for sub-sampling purpose on second call. This work can also be extended to further reduce variance of mean estimator using different ranked set schemes with auxiliary variable i.e one can construct ratio, product and regression type estimators.

Appendix A

Derrivation of $Bias$ and MSE

Bias of Proposed Estimator

$$\begin{aligned}
\hat{\bar{y}}^{**} &= \left[1 + \{1 + \frac{n_2}{n}(k-1)\}\frac{C_y^2}{n} + \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right]^{-1}\hat{\bar{y}}^{*} \\
&= \left[1 - \left\{1 + \frac{n_2}{n}(k-1)\right\}\frac{C_y^2}{n} - \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right]\left\{\bar{y} + \frac{n_2}{n}\left(\hat{\acute{\bar{y}}}_2 - \bar{y}_2\right)\right\} \\
E(\hat{\bar{y}}^{**}) &= E\Bigg[\bar{y} + \frac{n_2}{n}\left\{E_R(\hat{\acute{\bar{y}}}_2) - \bar{y}_2\right\} - \frac{C_y^2}{n}\left\{\bar{y} + \frac{n_2}{n}\left(E_r(\hat{\acute{\bar{y}}}_2) - \bar{y}_2\right)\right\} \\
&\quad + \left(\frac{n_2}{n}\right)(k-1)\left\{\left(E_R(\hat{\acute{\bar{y}}}_2) - \bar{y}_2\right) + \frac{n_2}{n}(k-1)\bar{y}\right\} \\
&\quad - \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\left(\bar{y} + \frac{n_2}{n}\left(E_R(\hat{\acute{\bar{y}}}_2) - \bar{y}_2\right)\right)\Bigg] \\
E(\hat{\bar{y}}^{**} - \bar{Y}) &= -\bar{Y}\left[\frac{C_y^2}{n}\left\{1 + W_2(k-1)\right\} + \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right]
\end{aligned}$$

Hence Bias of $\hat{\bar{y}}^{**}$ is

$$Bias(\hat{\bar{y}}^{**}) = -\bar{Y}\left[\frac{C_y^2}{n}\left\{1 + W_2(k-1)\right\} + \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right] \tag{A.1}$$

MSE of Proposed Estimator

$$
\begin{aligned}
MSE(\hat{\bar{y}}^{**}) &= E\left[\hat{\bar{y}}^{**} - \bar{Y}\right]^2 \\
&= E\left[\hat{\bar{y}}^{**2} + \bar{Y}^2 - 2\bar{Y}\hat{\bar{y}}^{**}\right]^2 \\
&= E(\hat{\bar{y}}^{**2}) + \bar{Y}^2 - 2\bar{Y}\left[\bar{Y} - \frac{C_y^2}{n}\left\{1 + W_2(k-1)\right\} - \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right] \qquad \text{(A.2)}
\end{aligned}
$$

Now consider $E(\hat{\bar{y}}^{**2})$ only,

$$
\begin{aligned}
E(\hat{\bar{y}}^{**2}) &= E\left[\left\{1 - \left(1 + \frac{n_2}{n}(k-1)\right)\frac{C_y^2}{n} - \frac{k}{nN}\frac{\sigma_r^2}{\bar{Y}^2}\right\}\left\{\bar{y} + \frac{n_2}{n}\left(\hat{\acute{\bar{y}}}_2 - \bar{y}_2\right)\right\}\right]^2 \\
&= E\Bigg[\Bigg\{1 + \frac{C_y^4}{n^2}\left(1 + w_2(k-1)\right)^2 + \frac{k^2}{n^2N^2}C_r^4 - 2\frac{C_y^2}{n}\left(1 + w_2(k-1)\right) \\
&\quad - 2\frac{k}{nN}C_r^2 + 2\frac{k}{nN}C_r^2\frac{C_y^2}{n}\left(1 + w_2(k-1)\right)\Bigg\}\left\{\bar{y}^2 + \left(\frac{n_2}{n}\right)^2 E_R\left(\hat{\acute{\bar{y}}}_2 - \bar{y}_2\right)^2 \\
&\quad + \frac{n_2}{n}\bar{y}\left(E_R(\hat{\acute{\bar{y}}}_2) - \bar{y}_2\right)\right\}\Bigg] \\
&= E_1\Bigg\{1 + \frac{C_y^4}{n^2} + \frac{C_y^4}{n^2}w_2^2(k-1)^2 + 2\frac{C_y^4}{n^2}w_2(k-1) + \frac{k^2}{n^2N^2}C_r^4 - 2\frac{C_y^2}{n} \\
&\quad - 2\frac{C_y^2}{n}w_2(k-1) - 2\frac{k}{nN}C_r^2 + 2\frac{k}{nN}C_r^2\frac{C_y^2}{n}\left(1 + w_2(k-1)\right)\Bigg\}\left\{\frac{S_y^2}{n} + \bar{Y}^2\right\} \\
&\quad + E_1E_2\left(\frac{n_2}{n}\right)^2 E_R\left(\hat{\acute{\bar{y}}}_2 - \bar{y}_2\right)^2 - 2\frac{C_y^2}{n}E_1E_2\left(\frac{n_2}{n}\right)^2 E_R\left(\hat{\acute{\bar{y}}}_2 - \bar{y}_2\right)^2 \\
&\quad - 2\frac{C_y^2}{n}(k-1)E_1E_2\left(\frac{n_2}{n}\right)^3 E_R\left(\hat{\acute{\bar{y}}}_2 - \bar{y}_2\right)^2 - 2\frac{k}{nN}C_r^2E_1E_2\left(\frac{n_2}{n}\right)^2 E_R\left(\hat{\acute{\bar{y}}}_2 - \bar{y}_2\right)^2 \\
E(\hat{\bar{y}}^{**2}) &= \bar{Y}^2 + \frac{C_y^2}{n}\frac{S_y^2}{n} + \frac{C_y^2}{n}\frac{S_y^2}{n}W_2^2(k-1)^2 + 2\frac{C_y^2}{n}\frac{S_y^2}{n}W_2(k-1) + \frac{k^2}{n^2N^2}C_r^2\sigma_r^2 - 2\bar{Y}^2\frac{C_y^2}{n} \\
&\quad - 2\bar{Y}^2\frac{C_y^2}{n}W_2(k-1) - 2\frac{k}{nN}\bar{Y}^2C_r^2 + 2\frac{kC_r^2}{nN}\frac{S_y^2}{n} + 2\frac{kC_r^2}{nN}W_2(k-1)S_y^2 + \frac{S_y^2}{n} \\
&\quad - 2\frac{C_y^2}{n}\frac{S_y^2}{n} - 2W_2(k-1)\frac{C_y^2}{n}\frac{S_y^2}{n} - 2\frac{kC_r^2}{nN}\frac{S_y^2}{n} + \left\{\frac{k\sigma_r^2}{nN} + \frac{W_2(k-1)}{n}S_{y(2)}^2\right\} \\
&\quad - 2\frac{C_y^2}{n}\left\{\frac{k\sigma_r^2}{nN} + \frac{W_2(k-1)}{n}S_{y(2)}^2\right\} - 2\frac{C_y^2W_2(k-1)}{n}\left\{\frac{k}{nN}\sigma_r^2 + \frac{W_2(k-1)}{n}S_{y(2)}^2\right\} \\
&\quad - 2\frac{kC_r^2}{nN}\left\{\frac{k\sigma_r^2}{nN} + \frac{W_2(k-1)}{n}S_{y(2)}^2\right\} \qquad \text{(A.3)}
\end{aligned}
$$

Using this value of $E(\hat{\bar{y}}^{**2})$ in A.3 and canceling out same terms we get we get MSE of $\hat{\bar{y}}^{**}$ as :

$$MSE(\hat{\bar{y}}^{**}) = \frac{S_y^2}{n}(1 - A^*) + \frac{W_2(k-1)}{n}\sigma_{y(2)}^2(1 - 2B^*) + \frac{k\sigma_r^2}{nN}\sigma_r^2(1 - \frac{k}{nN}C_r^2\sigma_r^2) \tag{A.4}$$

Bibliography

Bouza, C. (2002a). Estimation of the mean in ranked set sampling with non-responses. *Metrika,*, 56:171–179.

Bouza, C. (2009). Ranked set sampling and randomized response procedures for estimating the mean of a sensitive quantitative character. *Metrika*, 6:0184–0191.

Bouza, C. (2010). Ranked set sampling procedure for the estimation of the population mean under non-response : A comparison. *Revista Investigacion Operational*, 31(2):140–150.

Dell, T. and Clutter, J. (1972). Ranked set sampling theory with order statistics. background. *Biometrics*, 28:545–555.

Diana, G. and Perri, P. (2009). Estimating a sensitive proportion of randomized response procedure based on auxiliary information. *Statistical Papers*, 50:661–672.

Diana, G. and Perri, P. (2010). New scrambled response model for estimating the mean of a sensitive quantitative character . *Journal of Applied Statistics.*, 37:1875–1890.

Diana, G. and Perri, P. (2011). A class of estimators for quantitative sensitive data. *Statistical Papers*, 52:633–650.

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

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

Hansen, M. and Hurwitz, W. (1946). The problem of non-response in sample surveys. 41(236):517–529.

Kadilar, C. Cingi, H. (2003). Ratio estimator in stratified sampling. *Biometrical Journal.*, 45(2):218–225.

Kadilar, C. Cingi, H. (2005). A new ratio estimator in stratified sampling. *Communication in Statistics-Theory and Method.*, 34:597–602.

Khare, B. (1987). Allocation in stratified sampling in presence of non-response. *Metron.*, 41(I/II):213–221.

Khare, B. (1995). 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.

Khare, B. (2013). Seperate generalized estimators for population mean in the presence of non-response in stratified random sampling. *Advance in Statistics and Optimization.*, pages 150–158.

Khare, B. and Kumar, S. (2009). Utilization of coefficient of variation in the estimation of population mean using auxiliary character in presence of non response. *National Academy of Science letters-India*, 32:235–241.

Khare, B. and Kumar, S. (2011). 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.

Khare, B. and Sinha, R. (2009). 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.

Khare, B. and Srivastava, S. (1993). Estimation of population mean using auxiliary charactersin presence of non response. *National Academy of Science letters-India,*, 16(3):111–114.

Khare, B. and Srivastava, S. (1995). 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.

Mangat, N. and Singh, R. An alternative randomized response procedure. *Biometrica*, 77:439–442.

McIntyre, G. (1952). A method of unbiased selective sampling using ranked sets. *Australian Journal of Agriculureal Research*, 3:385–390.

Muttlak, H. (1996). Median ranked set sampling. *Journal of Applied statistical Science*, 6:91–98.

Okafor, F. (1996). On double sampling for stratification with subsampling the non-respondents. *Journal of Indian Social and Agriculture Statistics.*, 48:105–113.

Patil, G. (2002). Ranked set sampling. *Encyclopedia of Enviromentrics*, 3:1684–1690.

Pollock, K. and Bek, Y. (1976). A comparison of three randomized response models for quantitative data . *Journal of The American Statistical Association.*, 71:884–886.

Samawi, H. Abu-Dayyeh, W. and Ahmed, S. (1996). Extreme ranked set sampling. *Biometrical Journal,*, 30:577–586.

Searls, D. (1964). The utilization of a known coefficient of variation in the estimation procedure. *Journal of The American Statistical Association*, 41:517–529.

Shabbir, J. and Gupta, S. (2005). On modified randomized device of warner's model. *Pakistan Journal of Statistics*, 21:123–129.

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

Singh, R. and Sukhatme, B. (1969). Optimum stratification. *Annals of the Institute of Statistical Mathematics (AISM).*, 21:515–528.

Singh, R. and Sukhatme, B. (1973). Optimum stratification with ratio and regression method of estimation. *Annals of the Institute of Statistical Mathematics (AISM).*, 25:627–633.

Warner, S. (1965). Randomized response: A survey technique for eliminating evasive answer bias . *Journal of The American Statistical Association.*, 60:63–69.