SOME CONTRIBUTION TO THE THEORY OF NON-RESPONSE

THESIS

SUBMITTED TO

BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY

(A CENTRAL UNIVERSITY)

LUCKNOW

BABASAHEB
BHIMRAO
AMBEDKAR
UNIVERSITY

LUCKNOW

प्रज्ञा शील करूणा

ESTABLISHED 1996

FOR THE DEGREE OF

Doctor of Philosophy

IN

APPLIED STATISTICS

Submitted by

ABHAY PRATAP PANDEY

Supervisor
Dr. SURINDER KUMAR

Co-Supervisor
Dr. SHASHI BHUSHAN

DEPARTMENT OF APPLIED STATISTICS
SCHOOL FOR PHYSICAL SCIENCES
BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
(A CENTRAL UNIVERSITY)
VIDYA VIHAR, RAEBARELI ROAD
LUCKNOW-226 025

ENROLMENT NUMBER: 395/11

2017

Dedicated to
My Beloved Parents

DECLARATION

I, **Abhay Pratap Pandey**, Enrolment No. 395/11, hereby declare that the work which is being presented in the thesis entitled **"Some Contribution to the Theory of Non-Response"** in partial fulfilment of the requirements for the award of the Degree of Doctor of Philosophy and submitted in the department of Applied Statistics of the Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow is an authentic record of my own work carried out during a period from August, 2013 to August, 2017 under the supervision of Dr. Surinder Kumar, Associate Professor & Head, Department of Applied Statistics, School for Physical Sciences, Babasaheb Bhimrao Ambedkar University, Lucknow and co-supervision of Dr. Shashi Bhushan, Associate Professor & Head, Department of Mathematics and Statistics, Dr. Shakuntala Misra National Rehabilitation University, Lucknow.

The matter presented in this thesis has not been submitted by me for the award of any other degree or diploma of this or any other Institute.

Date:

(Abhay Pratap Pandey)
Research Scholar
Department of Applied Statistics
School for Physical Sciences
Babasaheb Bhimrao Ambedkar University
Lucknow-226025, India

CERTIFICATE

This is to certify that the thesis titled **"Some Contribution to the Theory of Non-Response"** submitted by **Mr. Abhay Pratap Pandey** is an original research work and has not been previously submitted in part or full for the award of any other degree or diploma to this or any other university.

The thesis submitted to Babasaheb Bhimrao Ambedkar University, Lucknow satisfies all the requirements as stipulated in the *Doctor of Philosophy (Ph.D.) regulations -1999 as amended in 2010* and it is fit for submission and evaluation for the award of the degree of Doctor of Philosophy of the University.

Date: Co-Supervisor

Supervisor

Head of the Department

Acknowledgement

The completion of this thesis was possible with the support and guidance of many people who were always there to help when I needed them. I take this opportunity to express my sincere gratitude to all of them.

First and foremost I am grateful to the God for all the countless gifts you have offered me and thanks to my parents for their love and support. I would like to start with the person who made the biggest difference in my life, my mentor and my co-supervisor, Dr. Shashi Bhushan, Head, Department of Mathematics and Statistics, Dr. Shakuntala Misra National Rehabilitation University, Lucknow. He has been there, in front of my eyes for last four years, motivating and inspiring every bit of me towards new possibilities in life. He has been a living role model to me, taking up new challenges every day, tackling them with all his grit and determination and always thriving to come out victorious. It's his vigor and hunger to perform in adverse situation, which has inspired me to thrive for excellence and nothing less. His attitude of living every moment as it comes, making serendipitous observation and converting them to new possibilities, correlating ideas and understanding the obvious has helped me come a long way and will always guide me in future. He made me realized what I had only heard, "When the going gets tough, the tough gets going". I will never find words to tell what I owe to him, and if I start doing it, I would not know where to stop... thanks for everything Sir... to me, professionally, you are perfection personified.

I would like to express my sincerest gratitude to my supervisor Dr. Surinder Kumar, Head, Department of Applied Statistics, Babasaheb Bhimrao Ambedkar University, Lucknow, for his unconditional support, motivation and guidance throughout the research work. He has always been there to take up all the challenges and tackling them with all his grit and determination. It is also a great pleasure for me to express my appreciation to Dr.

Jahangir Sabbir Khan, Department of Statistics and Operations Research, Aligarh Muslim University, for his continuous encouragement, guidance and motivation during my research work. I thank the faculty members, Dr. Amit Kumar Misra and Dr. Rahul Varshney for their encouraging and supportive nature. They were always been there to guide and motivate me whenever I encountered any difficulty. I also express my sincere gratitude to the Prof. Kamaan Singh Dean, School for Physical Sciences for his continuous evaluation of my research progress and providing the necessary guidance and support from time to time.

I am thankful to my seniors, Dr. Nazia Naqvi, Dr. Mukesh Kumar, Dr. Arun Kumar, Ms. Nazia Wahid, Mr. Ajay Kumar and my research colleagues Sumit Kumar and Mayank Vaish for their cooperative nature and moral support. Thanks also goes to my juniors Chandni Kumari, Vaishali Gupta, Ruby Chanchal, Mradula and Shailja Pandey for their cooperation during my research work. I also appreciate the non-teaching staff of the department, Ms. Saumya Trivedi, Dr. Amrendra Pratap Bahadur Singh and Mr. Nirmal Singh, colleagues in the university, well-wishers for their any direct or indirect help and support during my research work.

I would like to acknowledge my friends for their moral support and motivation, which drives me to give my best. Shahban Ali, Rohini Pokhrel, Chandraketu, Supriya, Anshita, Kranthi, Ratindra, Surya, Utkarsh, Satya, Neeraj… the list is endless… thanks to one and all. My special gratitude to Sumit Kumar and Vineet Kumar for being with me in thicks and thins of life, I find myself lucky to have friends like them in my life.

Last but not least, I would like to acknowledge the people who mean world to me, my parents, Smt. Vijay Laxmi Pandey and Shri Pooja Ram Pandey, my sisters and brothers, Ms. Soni Pandey, Anita Pandey, Mr. Pankaj Kumar Pandey, Mithlesh Pandey and Manoj Pandey. I don't imagine a life without their love and blessings. Thank you mom and dad for showing faith in me and for encouraging me to achieve my goals and aspirations. I consider myself the luckiest in the world to have such a supportive family, standing behind me with their love and support.

Lucknow

August…, 2017 (Abhay Pratap Pandey)

List of Publications

1. Bhushan, S. and Pandey, A. P. (2016): Optimality of the ratio type estimation methods in presence of missing data. *Communication in Statistics-Theory and Methods*, DOI 10.1080/03610926.2016.1167906.

2. Bhushan, S. and Pandey, A. P. (2016): Optimal imputation of the missing data for estimation of population mean. *Journal of Statistics and Management System*, 19(6), 755-769, DOI 10.1080/09720510.2016.1220099.

3. Bhushan, S. and Pandey, A. P. (2015): Family of estimators of ratio of a finite population using auxiliary information under random non-response. *Mathematical Sciences International Research Journal*, 4(1), 112-117.

4. Bhushan, S. and Pandey, A. P. (2016): A generalized family double sampling type estimators for estimation of finite population variance in presence of random non-response. *Recent Advances in Applied Statistics and its Applications Eds. Bhushan and Kumar*, 54-63, Lambart Publishing Inc, Germany.

5. Bhushan, S. and Pandey, A. P. (2017): Improved chain type imputation methods for estimating population mean *(Submitted)*.

6. Bhushan, S. and Pandey, A. P. (2017): On Optimality of imputation methods for estimation of population mean using higher order moments of an auxiliary variable *(Submitted)*.

7. Bhushan, S. and Pandey, A. P. (2017): Optimal imputation of the missing data using multi auxiliary information *(Submitted)*.

8. Bhushan, S. and Pandey, A. P. (2017): Improved estimation of population mean under non-response *(Submitted)*.

9. Bhushan, S. and Pandey, A. P. (2017): On efficient estimation of population mean under non-response *(Submitted)*.

10. Bhushan, S. and Pandey, A. P. (2017): Efficient classes of estimators using two auxiliary variables under non-response *(Submitted)*.

11. Bhushan, S. and Pandey, A. P. (2017): Improved estimation of population mean presence of random non-response *(Submitted)*.

12. Bhushan, S. and Pandey, A. P. (2017): Improved estimators for estimation of population variance presence of random non-response *(Submitted)*.

List of Tables

Contents

Chapter 1

Review of literature

1.1 Introduction

The statistical theory mainly deals with problem of drawing the inference about the characteristics of a population based on a representative sample. The sampling theory provides us with the tools of drawing inference about certain characteristics of the population. While proposing a sampling strategy for the given problem at hand we use different sampling schemes along with various estimation methods for studying the characteristics of interest. In estimation of population characteristics ancillary information closely related to the main characteristic plays a very important role. In estimation of population characteristics, the population parameters can be estimated more accurately by making use of information on an auxiliary variable that is correlated with study variable. The ratio and regression methods of estimation are good examples in this context. In sampling theory, it is well known that the efficiency of the estimators of unknown population parameters of the study variable y can be increased by suitably using known information on an auxiliary variable x, which may be positively or negatively correlated with y.

In sampling theory mainly two types of errors are committed, one is sampling error and other is non-sampling error. The sampling error can be minimized by increasing the size of the sample but non sampling errors pose more severe problem in estimation part of sampling theory. The non-sampling error can occur at any one or more stages of the survey. These errors are broadly classified as non-response error, response error, tabulation error etc. In presence of non-response, the estimate of parameter of interest under conventional sampling strategies lead to serious bias as some units fail to respond thereby not covering the targeted population as per the sampling strategy.

The problem of non response was first studied by Hansen and Hurwitz (1946) related to a mailed questionnaire. Hansen and Hurwitz suggested a technique for adjustment of bias due to non-response. The procedure suggested by Hansen and Hurwitz results in an unbiased estimator of population mean. Hansen and Hurwitz suggested a technique for dealing with unit non-response under deterministic non-response model. Later, a concept of random non-response was studied by Singh and Singh (1979, 85), Singh and Joarder (1998) among others. Alternatively, a situation when item non-response present in data set, a technique known as imputation was suggested by Rubin (1976) to mitigate the problem of missing data. Let us first review the relevant developments relating to imputation.

1.2 Imputation

In survey sampling, the problem of missing data is a major problem encountered by the statistician. The problem of missing data adversely affect the estimate of the population characteristics in survey sampling. Many authors dealt with the problem of non-response or missing data and suggested different type of method or estimators to estimate the population characteristics. In presence of missing data in survey sampling, methods of imputation for substitution of missing data is frequently used. Many Statisticians discussed the various methods of imputation for estimation of population mean of study variable in presence of missing data by using auxiliary information. To overcome the problem of missing data of the study variable, we use the auxiliary information about the study variable for imputation of the missing data. To circumvent the problem of missing data, imputation is used for substitution of missing data in the sampling theory. Rubin (1976) had suggested imputation methods which make incomplete data sets structurally complete and addressed three concepts: missing at random (MAR), observed at random (OAR) and parameter distribution (PD).

Missing at random (MAR):

The data are MAR if the probability of the observed missingness pattern given the observed and unobserved data, does not depend on the values of the unobserved data. We shall put all such cases where the data are missing only due to chance factors. It will therefore include cases where the enumerator is not able to contact the respondents only by chance and had he been able to contact, the data would have been collected. For example, when the information is kept on the punched cards, the non-response due to the accidental loss of one or more cards is of the first category. This type of non-response is called random non-response.

Observed at random (OAR):

The data are OAR, if for every possible value of the missing data, the probability of the observed missing-ness pattern, given the observed and unobserved data, does not depend on the values of the observed data.

Missing completely at random (MCAR):

The combination of MAR and OAR is called MCAR.

For dealing with the problem of missing data Sande (1979) and Kalton et al. (1981) have suggested some imputation methods. Thereafter, Kalton and Kasprzyk (1982), Singh and Singh (1991), Lee et al. (1994, 1995) also suggested some estimators under the problem of missing data. Heitzan and Basu (1996) have distinguished the meaning of missing at random (MAR) from missing completely at random (MCAR) in very eloquent manner. Many Statisticians like Toutenberg and Srivastava (1998), Singh and Horn (2000), Singh and Deo (2003), Rueda and Gonzalez (2004), Rueda, Gonzalez and Arcos (2005), Ahmed et al. (2006), Toutenberg, Srivastava and Shalabh (2008), Kadilar and Chingi (2008), Singh et al. (2009) and Singh et al. (2010), Diana and Perri (2010) etc. have discussed various methods of imputation for estimation of population mean under the problem of missing data by using auxiliary information about the study variable. The auxiliary information is generally used to increase the accuracy of any estimator which may either be adopted at estimation stage or sampling stage or both. Many authors including Kadilar and Chingi (2008) and Diana and Perri (2010) adopted the MCAR approach. We implicitly assume a MACR approach for imputation of the missing data in chapter 2, 3 and 4.

In order to overcome the problem of missing observations via non-response in sample survey, the technique of imputation is frequently used to deal with missing values effectively. Kalton et al. (1981) described some issues of non-response and imputation in the survey of income and program participation. Sande (1979) suggested hot deck approach to automatic edit and imputation, Lee et al. (1994, 1995) took up experiments with variance estimation from survey data with imputed values and variance estimation in the presence of imputed data for the generalized estimation system and used the information on an auxiliary variate for the purpose of imputation. Later, Singh and Horn (2000) proposed a compromised imputation in survey sampling, Ahmed et al. (2006) suggested the estimation of a population mean using different imputation methods, Thakur et al. (2012) studied some imputation methods under double sampling scheme for estimation of population mean, Singh et al. (2013) considered to assess the non-response with ratio method of imputation in successive sampling over two occasions.

Notations and results

Let y be the study variable defined on a finite population Ω of N identifiable units with values $y_i, i \in \Omega$. Let $\bar{Y} = N^{-1}\sum_{\Omega} y_i$ be the mean of the finite population under consideration over study variable y. Let x an auxiliary variable, correlated with y, which is effectively used to estimate the unknown mean of y. Further, the auxiliary information $x_i, i \in \Omega$ on x can be available on the entire population even when the information on study variable is missing.

A simple random sample without replacement (SRSWOR), s, of size n is drawn from $\Omega = \{1, 2, ..., N\}$ to estimate the population mean $\bar{Y}$. Let r be the number of responding units out of sampled n units. Let the set of responding units be denoted by A and that of non-responding units be denoted by $\bar{A}$. For every unit, $i \in A$, the value y_i is observed, but for the units $i \in \bar{A}$ the values are missing and imputed values must be derived to complete the y-data. The imputation is carried out with the aid of an auxiliary variable, x, such that x_i, the value of x for unit i, is known and positive for every $i \in s$ i.e. the data $x_s = \{x_i; i \in s\}$ are known.

Lee et al. (1994) suggested the notion for single value imputation, wherein if the ith sample unit is missing and requires imputation, the value $\hat{b}x_i$ is imputed, where $\hat{b} =$

$\sum_{i\in A} y_i/\sum_{i\in A} x_i$. In chapter 2, 3 and 4, we have focussed on improved single value imputation procedures. Some popular single value imputation procedures are described below. The mean method of imputation is given by

$$y_{.im} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r & if \quad i \in \bar{A} \end{cases} \tag{1.2.1}$$

The resultant estimator of population mean is given by

$$t_m = \frac{1}{n}\sum_{i\varepsilon s} y_{.im} = \bar{y}_r$$

The Ratio method of imputation is given by

$$y_{.irat} = \begin{cases} y_i & if \quad i \in A \\ \hat{b}x_i & if \quad i \in \bar{A} \end{cases} \tag{1.2.2}$$

The resultant estimator of population mean is given by

$$t_{rat} = \frac{1}{n}\sum_{i\varepsilon s} y_{.ir} = \bar{y}_r\frac{\bar{x}_n}{\bar{x}_r}$$

The Regression method of imputation is given by

$$y_{.ireg} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r + \hat{\beta}\left(x_i - \bar{x}_r\right) & if \quad i \in \bar{A} \end{cases} \tag{1.2.3}$$

The resultant estimator of population mean is given by

$$t_{reg} = \frac{1}{n}\sum_{i\varepsilon s} y_{.ireg} = \bar{y}_r + \hat{\beta}\left(\bar{x}_n - \bar{x}_r\right)$$

Singh and Horn (2000) proposed compromised imputation procedure such that

$$y_{.icom} = \begin{cases} \frac{\alpha n y_i}{r} + (1-\alpha)\,\hat{b}x_i & if \quad i \in A \\ (1-\alpha)\,\hat{b}x_i & if \quad i \in \bar{A} \end{cases} \tag{1.2.4}$$

The estimator under the compromised method of imputation is given by

$$t_{COMP} = \alpha \bar{y}_r + (1-\alpha)\, \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r}$$

where α is suitably chosen constant. It is important to note that, in this method, we are using information from imputed values for the responding units in addition to non-responding units. Therefore, this was an important result in search of the optimum imputation methods.

In presence of missing data, Singh and Deo (2003) suggested an imputation procedure by using power transformation. The power transformation imputed estimator of population mean remains better than the ratio and mean imputed estimators. In case of this method of imputation, the data after imputation take form

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r \left[n \left(\frac{\bar{x}_n}{\bar{x}_r} \right)^{\alpha} - r \right] \frac{x_i}{\sum\limits_{i\varepsilon \bar{A}} x_i} & if \quad i \in \bar{A} \end{cases} \tag{1.2.5}$$

where α is a suitably chosen constant. The point estimators of population mean $\bar{Y}$ becomes

$$t_{pt} = \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r} \right)^{\alpha}$$

If $\alpha = 1$, then the suggested power method of imputation leads to the ratio method of imputation described above. Also, if $\alpha = -1$, then the suggested power method of imputation leads to the product method of imputation and data after imputation take the form

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r \left[n \left(\frac{n\bar{x}_r - r\bar{x}_n}{\bar{x}_n} \right) - r \right] \frac{x_i}{\sum\limits_{i\varepsilon \bar{A}} x_i} & if \quad i \in \bar{A} \end{cases} \tag{1.2.6}$$

The point estimators of population mean $\bar{Y}$ becomes

$$t_p = \bar{y}_r \left(\frac{\bar{x}_r}{\bar{x}_n} \right)$$

They also suggested a method of imputation for population mean by using multi-auxiliary information, given by

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r \prod_{j=1}^{p} \left[n \left(\frac{\bar{x}_{nj}}{\bar{x}_{rj}} \right)^{\alpha_j} - r \right] \frac{\sum_{j=1}^{p} x_{ij}}{\sum_{i \varepsilon \bar{A}} \sum_{j=1}^{p} x_i} & if \quad i \in \bar{A} \end{cases} \tag{1.2.7}$$

The point estimator becomes

$$t_{mult} = \bar{y}_r \prod_{j=1}^{p} \left(\frac{\bar{x}_{nj}}{\bar{x}_{rj}} \right)^{\alpha_j}$$

One can easily observe that if $\alpha_j = 1, \; \forall j = 1, 2, 3, ..., p$ then the suggested method of imputation becomes

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r \prod_{j=1}^{p} \left[n \left(\frac{\bar{x}_{nj}}{\bar{x}_{rj}} \right) - r \right] \frac{\sum_{j=1}^{p} x_{ij}}{\sum_{i \varepsilon \bar{A}} \sum_{j=1}^{p} x_i} & if \quad i \in \bar{A} \end{cases} \tag{1.2.8}$$

The point estimator becomes

$$t_{mult1} = \bar{y}_r \prod_{j=1}^{p} \left(\frac{\bar{x}_{nj}}{\bar{x}_{rj}} \right)$$

where $\prod_{j=1}^{p} xj = x_1 x_2 \dots x_p$ denote the product of p terms.

Motivated by Lee et al. (1994), Singh and Horn (2000) and Singh and Deo (2003), Ahmed et al. (2006) suggested some more methods of imputation of population mean using auxiliary information in presence of missing data stated below.

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \frac{1}{n-r} \left[n \bar{y}_r \left(\frac{\bar{X}}{\bar{x}_n} \right) - r \bar{y}_r \right] & if \quad i \in \bar{A} \end{cases} \tag{1.2.9}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \dfrac{1}{n-r}\left[n\bar{y}_r\left(\dfrac{\bar{X}}{\bar{x}_n}\right)^{\beta_1} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases} \tag{1.2.10}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \dfrac{1}{n-r}\left[n\bar{y}_r\left(\dfrac{\bar{X}}{\bar{x}_r}\right) - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases} \tag{1.2.11}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \dfrac{1}{n-r}\left[n\bar{y}_r\left(\dfrac{\bar{X}}{\bar{x}_r}\right)^{\beta_2} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases} \tag{1.2.12}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \frac{1}{n-r}\left[\dfrac{n\bar{y}_r\bar{X}}{\theta_1\bar{x}_n + (1-\theta_1)\bar{X}} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases} \tag{1.2.13}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \dfrac{1}{n-r}\left[\dfrac{n\bar{y}_r\bar{X}}{\theta_2\bar{x}_r + (1-\theta_2)\bar{X}} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases} \tag{1.2.14}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \dfrac{1}{n-r}\left[n\bar{y}_r\left(\dfrac{\bar{x}_n}{\bar{x}_r}\right) - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases} \tag{1.2.15}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \dfrac{1}{n-r}\left[n\bar{y}_r\left(\dfrac{\bar{x}_n}{\bar{x}_r}\right)^{\beta_3} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases} \tag{1.2.16}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \left[\dfrac{\bar{y}_r\left(x_i + \frac{r}{n-r}\right)}{\theta_3\bar{x}_r + (1-\theta_3)\bar{x}_n} - \dfrac{r}{n-r}\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases} \tag{1.2.17}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \dfrac{1}{n-r}\left[n\bar{y}_r\left(\dfrac{\bar{x}_n}{\bar{x}_r}\right)^{\beta_4}\left(\dfrac{\bar{X}}{\bar{x}_n}\right)^{\beta_5} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases} \tag{1.2.18}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \dfrac{\bar{y}_r\bar{X}\left(x_i + \dfrac{r}{n-r}\bar{x}_r\right)}{\left\{\theta_4\bar{x}_r + (1-\theta_4)\bar{x}_n\right\}\left\{\theta_5\bar{x}_n + (1-\theta_5)\bar{X}\right\}} - \dfrac{r}{(n-r)}\bar{y}_r & if \quad i \in \bar{A} \end{cases} \tag{1.2.19}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r + b_1\left(\bar{X} - \bar{x}_n\right) & if \quad i \in \bar{A} \end{cases} \tag{1.2.20}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r + b_2\left(\bar{X} - \bar{x}_r\right) & if \quad i \in \bar{A} \end{cases} \tag{1.2.21}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r + b_3\left(\bar{x}_n - \bar{x}_r\right) & if \quad i \in \bar{A} \end{cases} \tag{1.2.22}$$

$$y_{.i} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r + k_1\left(\bar{X} - \bar{x}_n\right) + k_2\left(\bar{x}_n - \bar{x}_r\right) & if \quad i \in \bar{A} \end{cases} \tag{1.2.23}$$

The resultant point estimators are given below

$$t_1 = \bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)$$

$$t_2 = \bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)^{\beta_1}$$

$$t_2 = \bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)^{\beta_1}$$

$$t_3 = \bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)$$

$$t_4 = \bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^{\beta_2}$$

$$t_5 = \frac{\bar{y}_r\bar{X}}{\theta_1\bar{x}_n + (1-\theta_1)\,\bar{X}}$$

$$t_6 = \frac{\bar{y}_r\bar{X}}{\theta_2\bar{x}_r + (1-\theta_2)\,\bar{X}}$$

$$
\begin{aligned}
t_7 &= \bar{y}_{rat} = \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r}\right) \\
t_8 &= \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\beta_3} \\
t_9 &= \frac{\bar{y}_r \bar{x}_n}{\theta_3 \bar{x}_r + (1-\theta_3)\, \bar{x}_n} \\
t_{10} &= \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\beta_4} \left(\frac{\bar{X}}{\bar{x}_n}\right)^{\beta_5} \\
t_{11} &= \bar{y}_r \left[\frac{\bar{x}_n}{\theta_4 \bar{x}_r + (1-\theta_4)\, \bar{x}_n}\right] \left[\frac{\bar{X}}{\theta_5 \bar{x}_n + (1-\theta_5)\, \bar{X}}\right] \\
t_{12} &= \bar{y}_r + b_1 \left(\bar{X} - \bar{x}_n\right) \\
t_{13} &= \bar{y}_r + b_2 \left(\bar{X} - \bar{x}_r\right) \\
t_{14} &= \bar{y}_r + b_3 \left(\bar{x}_n - \bar{x}_r\right) \\
t_{15} &= \bar{y}_r + k_1 \left(\bar{X} - \bar{x}_n\right) + k_2 \left(\bar{x}_n - \bar{x}_r\right)
\end{aligned}
$$

In order to estimate the mean of study variable in the presence of missing data when the population mean of the auxiliary information are assumed to be known, Kadilar and Cingi (2008) suggested some regression-cum-ratio type imputation methods based on their earlier work Kadilar and Cingi (2004). It is noteworthy that Kadilar and Chingi (2008) suggested the estimators for population mean only and not the imputation criteria for obtaining the missing observations which may be required for further analysis. Kadilar and Cingi (2004) proposed a regression-cum-ratio type estimator in the case of no missing data given by

$$
t_{KC} = \frac{\bar{y}_r + b_1 \left(\bar{X} - \bar{x}_n\right)}{\bar{x}_n} \bar{X} \tag{1.2.24}
$$

where $b = \frac{s_{xy}}{s_x^2}$ is the regression coefficient of y on x. It is important to observe that classical regression estimator $t_{lr} = \bar{y}_r + b_1\left(\bar{X} - \bar{x}_n\right)$, is more efficient than estimator proposed by Kadilar and Cingi (2004) t_{KC}, for $\bar{Y}$. Kadilar and Cingi (2008) proposed three estimators by using regression-cum-ratio method of imputation in the presence of missing data are as follows

$$t_{KC_1} = \frac{\bar{y}_r + b_1\left(\bar{X} - \bar{x}_n\right)\bar{X}}{\bar{x}_n} \tag{1.2.25}$$

$$t_{KC_2} = \frac{\bar{y}_r + b_1\left(\bar{X} - \bar{x}_r\right)\bar{X}}{\bar{x}_r} \tag{1.2.26}$$

$$t_{KC_3} = \frac{\bar{y}_r + b_1\left(\bar{x}_n - \bar{x}_r\right)\bar{X}}{\bar{x}_r} \tag{1.2.27}$$

Diana and Perri (2010) proposed three regression type estimators motivated by estimators suggested by Diana and Tommasi (2003) and Diana and Perri (2007), who proposed general unbiased class(es) of estimators, when the information on auxiliary variable(s) is available and have proved that the best estimator in the class is regression-type estimator up to the first order approximation. Diana and Perri (2010) proposed the estimators by using the same amount information as Kadilar and Cingi (2008). Diana and Perri proposed three estimators as by using different regression-type methods of imputation such that the imputed data is given by

$$y_{.iDP_1} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r + b_1\left(\bar{X} - \bar{x}_n\right) & if \quad i \in \bar{A} \end{cases} \tag{1.2.28}$$

The resultant estimator is given by

$$t_{DP_1} = t_{12} = \bar{y}_r + b_1\left(\bar{X} - \bar{x}_n\right)$$

$$y_{.iDP_2} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r + b_2\left(\bar{X} - \bar{x}_r\right) & if \quad i \in \bar{A} \end{cases} \tag{1.2.29}$$

The resultant estimator is given by

$$t_{DP_2} = t_{13} = \bar{y}_r + b_2\left(\bar{X} - \bar{x}_r\right)$$

$$y_{.iDP_3} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r + b_3\left(\bar{x}_n - \bar{x}_r\right) & if \quad i \in \bar{A} \end{cases} \tag{1.2.30}$$

The resultant estimator is given by

$$t_{DP_3} = t_{14} = \bar{y}_r + b_3\left(\bar{x}_n - \bar{x}_r\right)$$

It can be easily observed that the t_{DPi} uses the same amount of information as that of t_{KCi}, $(i = 1, 2, 3)$ hence they should only be compared with the corresponding Kadilar and Cingi (2008) estimator, so that valid conclusions may be drawn.

In their article, Diana and Perri highlighted the role of the auxiliary information in improving the estimates of the population mean in presence of the missing data and proved that the suggested estimators are more efficient than the estimators suggested by Kadilar and Cingi (2008), thereby proving that their estimators are optimal. Diana and Perri also emphasised on the fact that the estimators proposed by them are structurally simpler because they only consider the regression type component while excluding the ratio-type component introduced by Kadilar and Cingi (2008).

Let $\varepsilon_0 = \dfrac{\bar{y}_r}{\bar{Y}} - 1$, $\varepsilon_1 = \dfrac{\bar{x}_r}{\bar{X}} - 1$, and $\varepsilon_2 = \dfrac{\bar{x}_n}{\bar{X}} - 1$. Then using the concept of two phase sampling following Rao and Sitter (1995) and the mechanism of MCAR, for given r and n, we have:

$E\left(\varepsilon_0\right) = E\left(\varepsilon_1\right) = E\left(\varepsilon_2\right) = 0$

and

$E\left(\varepsilon_0^2\right) = \left(\dfrac{1}{r} - \dfrac{1}{N}\right)C_y^2$, $E\left(\varepsilon_1^2\right) = \left(\dfrac{1}{r} - \dfrac{1}{N}\right)C_x^2$, $E\left(\varepsilon_2^2\right) = \left(\dfrac{1}{n} - \dfrac{1}{N}\right)C_x^2$,

$E\left(\varepsilon_0\varepsilon_1\right) = \left(\dfrac{1}{r} - \dfrac{1}{N}\right)\rho_{yx}C_yC_x$, $E\left(\varepsilon_0\varepsilon_2\right) = \left(\dfrac{1}{n} - \dfrac{1}{N}\right)\rho_{yx}C_yC_x$, $E\left(\varepsilon_1\varepsilon_2\right) = \left(\dfrac{1}{n} - \dfrac{1}{N}\right)C_x^2$.

where $C_y^2 = \dfrac{S_y^2}{\bar{Y}^2}$, $C_x^2 = \dfrac{S_x^2}{\bar{X}^2}$, $\rho_{yx} = \dfrac{S_{xy}}{S_xS_y}$ and S_y^2, S_x^2 and S_{xy} have their usual meanings.

Also define

$\bar{y}_r = r^{-1}\sum_{i\varepsilon R} y_i$, $\bar{x}_r = r^{-1}\sum_{i\varepsilon R} x_i$, $\bar{x}_n = n^{-1}\sum_{i\varepsilon s} x_i$ $s^*_{xy} = (r-1)^{-1}\sum_{i=1}^{r}(y_i - \bar{y}_r)$, $s^{*2}_x = (r-1)^{-1}\sum_{i=1}^{r}(x_i - \bar{x}_r)^2$, $s^{*2}_y = (r-1)^{-1}\sum_{i=1}^{r}(y_i - \bar{y}_r)^2$, $c^2_y = \frac{s^{*2}_y}{\bar{y}^2_r}$, $c^2_x = \frac{s^{*2}_x}{\bar{x}^2_r}$, $\hat{\rho}_{yx} = \frac{s^*_{xy}}{s^*_x S^*_y}$.
The minimum mean square error (MSE) of existing estimators are given below

$$Var(t_m) = \left(\frac{1}{r} - \frac{1}{N}\right) S^2_y$$

$$MSE(t_{rat}) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2_y + \left(\frac{1}{r} - \frac{1}{N}\right)\left[S^2_y + R^2 S^2_x - 2RS_{xy}\right]$$

$$min.MSE(t_{comp}) = MSE(t_{rat}) - \left(\frac{1}{r} - \frac{1}{N}\right)\left(1 - \rho\frac{C_y}{C_x}\right)^2 \bar{Y}^2 C^2_x$$

$$MSE(t_{pt}) = MSE(t_{rat}) - \left(\frac{1}{r} - \frac{1}{n}\right) S^2_x \left(\frac{S_{xy}}{S^2_x} - \frac{\bar{Y}}{\bar{X}}\right)^2$$

$$MSE(t_{mult}) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2_y + \left(\frac{1}{r} - \frac{1}{n}\right) S^2_y \left(1 - R_{y.x_1x_2...x_p}\right)$$

$$MSE(t_p) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2_y + \left(\frac{1}{r} - \frac{1}{n}\right)\left[S^2_y + R^2 S^2_x + 2RS_{xy}\right]$$

$$min.MSE(t_2) = min.MSE(t_5) = min.MSE(t_{12}) = \left(\frac{1}{r} - \frac{1}{N}\right) S^2_y - \left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_{xy}}{S^2_x}$$

$$min.MSE(t_4) = min.MSE(t_6) = min.MSE(t_{13}) = min.MSE(t_{14}) = \left(\frac{1}{r} - \frac{1}{N}\right) S^2_y \left(1 - \rho^2\right)$$

$$min.MSE(t_8) = min.MSE(t_9) = min.MSE(t_{14}) = \left(\frac{1}{r} - \frac{1}{N}\right) S^2_y - \left(\frac{1}{r} - \frac{1}{n}\right)\frac{S_{xy}}{S^2_x}$$

optimum value of the constants using Ahmed et al. (2006) are given below
$\beta_1 = \beta_2 = \beta_3 = \theta_1 = \theta_2 = \theta_3 = \rho\frac{C_y}{C_x}$, $b_1 = b_2 = b_3 = k_1 = k_2 = \frac{S_{xy}}{S^2_x}$.

$$min.MSE(t_{KC_1}) = \left(\frac{1}{r} - \frac{1}{N}\right) S^2_y + \left(\frac{1}{n} - \frac{1}{N}\right) S^2_x \left(R^2 - B^2\right)$$

$$min.MSE(t_{KC_2}) = \left(\frac{1}{r} - \frac{1}{N}\right)\left(S^2_y + R^2 S^2_x - BS_{xy}\right)$$

$$min.MSE(t_{KC_3}) = \left(\frac{1}{r} - \frac{1}{N}\right) S^2_y + \left(\frac{1}{r} - \frac{1}{n}\right)\left[(R+B)^2 S^2_x - 2(R+B)S_{xy}\right]$$

where
$R = \dfrac{\bar{Y}}{\bar{X}}$ and $B = \dfrac{S_{xy}}{S_x^2}$.

$$min.MSE(t_{DP_1}) = \left(\frac{1}{r} - \frac{1}{N}\right) S_y^2 - \left(\frac{1}{n} - \frac{1}{N}\right) \frac{S_{xy}}{S_x^2}$$

$$min.MSE(t_{DP_2}) = \left(\frac{1}{r} - \frac{1}{N}\right) S_y^2 \left(1 - \rho^2\right)$$

$$min.MSE(t_{DP_3}) = \left(\frac{1}{r} - \frac{1}{N}\right) S_y^2 - \left(\frac{1}{r} - \frac{1}{n}\right) \frac{S_{xy}}{S_x^2}$$

1.3 Non-response

The problem of non response was first studied by Hansen and Hurwitz (1946) related to a mailed questionnaire. Hansen and Hurwitz (1946) suggested a technique for adjustment of bias due to non-response. The procedure suggested by Hansen and Hurwitz results in an unbiased estimator. In this procedure we first take a simple random sample without replacement of n possible respondent and mail a survey schedule to all of them with a deadline for response. When the deadline of reply is over, we identify the non respondents. Subsequently, we select a subsample by simple random sample without replacement in the non-response class and obtain information by personal interview. Finally, we pool the results from both classes to estimate the population mean (total) thereby obtaining an unbiased estimate of population mean.

The non-response adversely affects the estimate of population mean and population variance and many authors have suggested a number of estimators for estimating population parameter and their variance under the non-response for various situations. The problem of deterministic non-response in sample surveys is common and is more prevalent in mail surveys than in personal interviews (Srinath, 1971). El-Badry (1956) extended Hansen and Hurwitz's technique.

Notations and results

Consider a finite population of size N and a random sample of size n drawn without replacement. In surveys on human populations, it is often the case that n_1 units respond to the items under examination, but the remaining $(n-n_1)$ units do not provide any response. The initial survey may be conducted through the mail or by telephone, perhaps computer aided; see Rao (1986). In the case of non-response in the initial attempt, Hansen and Hurwitz (1946) suggested a double sampling scheme for estimating the population mean comprising the following steps:

(1) A simple random sample of size n is selected and the questionnaire is mailed to the sampled units.

(2) A subsample of size $r = \frac{n_2}{k}, (k \geq 0)$ from the non-responding units in the initial attempt is contacted through personal interviews.

In the Hansen and Hurwitz method, the population of size N is supposed to be composed of two strata, namely respondents and non-respondents, having size N_1 and $N_2 = (N - N_1)$. Thus, we label the data as $y_1, ..., y_{N_1}$ for the response group. Let $\bar{Y} = \sum_{i=1}^{N} y_i/N$ and $S_y^2 = \sum_{i=1}^{N} \left(y_i - \bar{Y}\right)^2/(N-1)$ denote the population mean and variance, respectively. Let $\bar{Y}_1 = \sum_{i=1}^{N_1} y_i/N_1$ and $S_{y_1}^2 = \sum_{i=1}^{N_1} \left(y_i - \bar{Y}_1\right)^2/(N_1-1)$ denote the mean and variance of the response group, respectively, and similarly, let $\bar{Y}_2 = \sum_{i=1}^{N_2} y_i/N_2$ and $S_{y_2}^2 = \sum_{i=1}^{N_2} \left(y_i - \bar{Y}_2\right)^2/(N_2-1)$ denote the mean and variance of the non-response group. The population mean can be written as $\bar{Y} = W_1\bar{Y}_1 + W_2\bar{Y}_2$, where $W_1 = N_1/N$ and $W_2 = N_2/N$. The sample mean $\bar{y}_1 = \sum_{i=1}^{n_1} y_i/n_1$ is unbiased for $\bar{Y}_1$, but has a bias equal to $W_2(\bar{Y}_1 - \bar{Y}_2)$ in estimating the population mean $\bar{Y}$.

The sample mean $\bar{y}_{2r} = \sum_{i=1}^{r} y_i/r$ is unbiased for the mean $\bar{y}_2$ of the n_2 units. An unbiased estimator for the population mean $\bar{Y}$ is

$$\bar{y}^* = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\bar{y}_{(2r)} \tag{1.3.1}$$

$$\bar{y}^* = w_1\bar{y}_1 + w_2\bar{y}_{2r}$$

where $w_1 = n_1/n$ and $w_2 = n_2/n$.

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

$$Var(\bar{y}^*) = \left(\frac{1-f}{n}\right) S_y^2 + \frac{W_2(k-1)}{n} S_{y_2}^2$$

where $f = n/N$.

Let $x_i, (i = 1, 2, \ldots, N)$ denote an auxiliary variate correlated with the study variate $y_i, (i = 1, 2, \ldots, N)$. The population mean of the auxiliary variate x is $\bar{X} = \sum_{i=1}^{N} x_i/N$. Let $\bar{X}_1$ and $\bar{X}_2$ denote the means of the response and non-response groups, respectively. Let $\bar{x}$ denote the mean of all n units. Let $\bar{x}_1$ and $\bar{x}_2$ denote the means of n_1 responding units and the n_2 non-responding units, respectively. Furthermore, let $\bar{x}_{2r} = \sum_{i=1}^{r} x_i/r$ denote the mean of the subsampled units. The population variances of x and y are denoted by S_x^2 and S_y^2, and the population covariance by S_{xy}. The population correlation coefficient is $\rho_{yx} = S_{xy}/S_x S_y$. The unbiased estimator of the population mean $\bar{X}$ of the auxiliary variate x is

$$\bar{x}^* = \frac{n_1}{n}\bar{x}_1 + \frac{n_2}{n}\bar{x}_{(2r)} \tag{1.3.2}$$

$$\bar{x}^* = w_1\bar{x}_1 + w_2\bar{x}_{2r}$$

The variance of $\bar{x}^*$ is given by

$$Var(\bar{x}^*) = \left(\frac{1-f}{n}\right) S_x^2 + \frac{W_2(k-1)}{n} S_{x_2}^2$$

with $(\bar{x}_1, \bar{y}_1)$ and $(\bar{x}_{2r}, \bar{y}_{2r})$ are the sample means based on n_1 units and sub sample means based on r units of the variates (x, y), respectively, $b^* = \left(s_{xy}^*/s_x^{*2}\right)$ is an estimator population regression coefficient $\beta = (S_{xy}/S_x^2)$.

where

$s_{xy}^* = \frac{1}{(n-1)}\left(\sum_{u_1} x_j y_j + r\sum_{u_{2m}} x_j y_j - n\bar{x}\bar{y}^*\right)$, $s_x^{*2} = \frac{1}{(n-1)}\left(\sum_{u_1} x_j^2 + r\sum_{u_{2m}} x_j^2 - n\bar{x}\bar{x}^*\right)$,

$S_{xy} = \frac{1}{N-1}\sum_{j=1}^{N}(x_j - \bar{X})(y_j - \bar{Y})$, $S_x^2 = \frac{1}{N-1}\sum_{j=1}^{N}(x_j - \bar{X})^2$, $b^{**} = \left(s_{xy}^*/s_x^2\right)$,

$S_{x_2}^2 = \sum_{i=1}^{N_2}\left(x_i - \bar{X}_2\right)^2/(N_2 - 1)$.

It is well known that the use of auxiliary information improves the precision of the estimates under non-response see Khare and Srivastava (1993) Cochran (1977), and Rao (1983, 1986) suggested the use of the ratio method of estimation for population mean of the study variate with sub-sampling from the amongst the non-respondents. Khare and Srivastava (1993) suggested an estimation procedure of population mean using auxiliary character in presence of non-response, Khare and Srivastava (1995) proposed to study the conventional and alternative two phase sampling ratio product and regression estimators in presence of non-response. Khare and Srivastava (1997) proposed transformed ratio type estimators for the population mean in the presence of non-response. Okafor and Lee (2000) proposed a double sampling scheme for ratio and regression estimation with sub sampling the non-respondent are also dealing with non-response problem. Singh and Kumar (2009a) proposed a general class of estimators of population mean in survey sampling using auxiliary information with sub sampling the non-respondent. Also, Singh and Kumar and Kozak (2010) suggested a number of estimators for estimating population mean under non-response. It is also observed that dealing with non-response problem, most of these works done for regression and ratio type estimators to estimate parameters were improved by Singh and Kumar (2008, 2010) in both single phase and two phase sampling.

If the non-response occurs on both the study variable y and auxiliary variable x, and the population mean $\bar{X}$ of the auxiliary variable is known. Then, in this situation, the usual ratio, product and regression estimators for the population mean $\bar{Y}$ of y are respectively defined (see Cochran, 1977) by

$$t_{rat_1} = \bar{y}^* \left(\frac{\bar{X}}{\bar{x}^*}\right) \tag{1.3.3}$$

$$t_{p_1} = \bar{y}^* \left(\frac{\bar{x}^*}{\bar{X}}\right) \tag{1.3.4}$$

$$t_{lr_1} = \bar{y}^* + b^* \left(\bar{X} - \bar{x}^*\right) \tag{1.3.5}$$

where $b^* = s^*_{xy}/s^{*2}_x$ and the estimates s^*_{xy} and s^{*2}_x are based on the available data under the given sampling design.

When non-response occurs on the study variable y, information on the auxiliary variable x is obtained from all the sample units (i.e. the initial units), and the population mean $\bar{X}$ of the auxiliary variable is known. Then, in this situation, the usual ratio, product and regression estimators for the population mean $\bar{Y}$ of y are respectively defined (see Rao, 1986) by

$$t_{rat_2} = \bar{y}^* \left(\frac{\bar{X}}{\bar{x}}\right) \tag{1.3.6}$$

$$t_{p_2} = \bar{y}^* \left(\frac{\bar{x}}{\bar{X}}\right) \tag{1.3.7}$$

$$t_{lr_2} = \bar{y}^* + b\left(\bar{X} - \bar{x}\right) \tag{1.3.8}$$

where $b = s^*_{xy}/s^2_x$, $s^2_x = \sum\limits_{i=1}^{n}(x_i - \bar{x})^2/(n-1)$ and the estimates s^*_{xy} and s^2_x are based on the available data under the given sampling design.

Also, when non-response occurs on the study variable y, information on the auxiliary variable x is obtained from all the sample units, and the population mean $\bar{X}$ of the auxiliary variable x is not known. Then, in this situation, the usual ratio, product and regression estimators for the population mean $\bar{Y}$ of y are respectively defined by

$$t_{rat_3} = \bar{y}^* \left(\frac{\bar{x}}{\bar{x}^*}\right) \tag{1.3.9}$$

$$t_{p_3} = \bar{y}^* \left(\frac{\bar{x}^*}{\bar{x}}\right) \tag{1.3.10}$$

$$t_{lr_3} = \bar{y}^* + b_{(2r)}\left(\bar{x} - \bar{x}^*\right), \tag{1.3.11}$$

where $b_{(2r)} = s_{xy_{(2r)}}/s^2_{x_{(2r)}}$, $s_{xy_{(2r)}} = \sum\limits_{i=1}^{r}(x_i - \bar{x}_{(2r)})(y_i - \bar{y}_{(2r)})/(r-1)$ and $s^2_{x_{(2r)}} = \sum\limits_{i=1}^{r}(x_i - \bar{x}_{(2r)})^2/(r-1)$.

Okafor and Lee (2000) suggested the usual ratio, product and regression estimators using two phase sampling in presence of non-response are, respectively, defined as

$$t_{rat_4} = \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}^*}\right) \tag{1.3.12}$$

$$t_{p_4} = \bar{y}^* \left(\frac{\bar{x}^*}{\bar{x}'}\right) \tag{1.3.13}$$

$$t'_{lr_4} = \bar{y}^* + b^*(\bar{x}' - \bar{x}^*) \tag{1.3.14}$$

and alternative two-phase ratio, product and regression estimators are given by

$$t_{rat_5} = \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}}\right) \tag{1.3.15}$$

$$t_{p_5} = \bar{y}^* \left(\frac{\bar{x}}{\bar{x}'}\right) \tag{1.3.16}$$

$$t_{lr_5} = \bar{y}^* + b^{**}(\bar{x}' - \bar{x}) \tag{1.3.17}$$

where $\bar{X}$ is the sample mean of the auxiliary character x based on n units and $\bar{x}'$ is the sample mean of x based on n' units.

When information on the auxiliary variable x is obtained from all the sample units and population mean $\bar{X}$ of the auxiliary variable is known, but some sample units fail to supply information on the study variable y, Singh and Kumar (2008) proposed some new estimators under situation given by

$$t_{rat_6} = \bar{y}^* \left(\frac{\bar{X}}{\bar{x}^*}\right) \left(\frac{\bar{X}}{\bar{x}}\right) \tag{1.3.18}$$

$$t_{p_6} = \bar{y}^* \left(\frac{\bar{x}^*}{\bar{X}}\right) \left(\frac{\bar{x}}{\bar{X}}\right) \tag{1.3.19}$$

$$t_g = \bar{y}^* \left(\frac{\bar{X}}{\bar{x}^*}\right)^{\alpha_1} \left(\frac{\bar{X}}{\bar{x}}\right)^{\alpha_2} \tag{1.3.20}$$

$$t_d = \bar{y}^* + d_1(\bar{x} - \bar{x}^*) + d_2(\bar{X} - \bar{x}) \tag{1.3.21}$$

Singh and Kumar (2010) proposed some modified ratio, product and regression estimators using double sampling in presence of non-response.

$$t_d^{(r)} = \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}^*}\right) \left(\frac{\bar{x}'}{\bar{x}}\right) \tag{1.3.22}$$

$$t_d^{(p)} = \bar{y}^* \left(\frac{\bar{x}^*}{\bar{x}'}\right) \left(\frac{\bar{x}}{\bar{x}'}\right) \tag{1.3.23}$$

$$t_{\alpha_1}^{\alpha_2} = \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}^*}\right)^{\alpha_1} \left(\frac{\bar{x}'}{\bar{x}}\right)^{\alpha_2} \tag{1.3.24}$$

$$t_{dd} = \bar{y}^* + d_1\left(\bar{x} - \bar{x}^*\right) + d_2\left(\bar{x}' - \bar{x}\right) \tag{1.3.25}$$

Let us consider

$\bar{y}^* = \bar{Y} + \varepsilon_0^*,\ \bar{x}^* = \bar{X} + \varepsilon_1^*,\ \bar{x} = \bar{X} + \varepsilon_1,\ \bar{x}' = \bar{X} + {\varepsilon_1}'.$

such that

$E(\varepsilon_0^*) = E(\varepsilon_1^*) = E(\varepsilon_1) = E({\varepsilon_0}') = 0,$
$E(\varepsilon_0^{*2}) = \left(\frac{1-f}{n}\right) S_y^2 + \frac{W_2(k-1)}{n} S_{y_2}^2,\ E(\varepsilon_1^{*2}) = \left(\frac{1-f}{n}\right) S_x^2 + \frac{W_2(k-1)}{n} S_{x_2}^2,$
$E(\varepsilon_1^2) = \left(\frac{1-f}{n}\right) S_x^2,\ E(\varepsilon_1'^2) = \left(\frac{1-f'}{n'}\right) S_x^2,\ E(\varepsilon_0^*\varepsilon_1^*) = \left(\frac{1-f}{n}\right) S_{xy} + \frac{W_2(k-1)}{n} S_{xy_2},$
$E(\varepsilon_0^*\varepsilon_1) = \left(\frac{1-f}{n}\right) S_{xy},\ E(\varepsilon_0^*{\varepsilon_1}') = \left(\frac{1-f'}{n'}\right) S_{xy},\ E(\varepsilon_0^*\varepsilon_1) = \left(\frac{1-f}{n}\right) S_x^2,$
$E(\varepsilon_1^*{\varepsilon_1}') = \left(\frac{1-f'}{n'}\right) S_x^2,\ E(\varepsilon_1{\varepsilon_1}') = \left(\frac{1-f'}{n'}\right) S_x^2.$

where

$W_2 = N_2/N,\ S_r^2 = \left(S_y^2 + R^2 S_x^2 - 2R\rho S_y S_x\right) = \left[S_y^2 + R^2 S_x^2 (R - 2\beta)\right],$
$S_{r_2}^2 = \left[S_{y_2}^2 + R^2 S_{x_2}^2 (R - 2\beta_2)\right],$
$S_p^2 = \left(S_y^2 + R^2 S_x^2 + 2R\rho S_y S_x\right) = \left[S_y^2 + R^2 S_x^2 (R + 2\beta)\right],$
$S_{p_2}^2 = \left[S_{y_2}^2 + R^2 S_{x_2}^2 (R + 2\beta_2)\right],$
$S_{\beta_2}^2 = \left[S_{y_2}^2 + \beta^2 S_{x_2}^2 (\beta - \beta_2)\right],\ \beta_2 = S_{xy(2)}/S_{x_2}^2,$
$Sxy_2 = \frac{1}{N_2-1} \sum_{j=1}^{N} (x_j - \bar{X}_2)(y_j - \bar{Y}_2).$

The MSE, upto the first order of approximation, of the above mentioned estimators are given by

$$MSE(t_{rat_1}) = \frac{(1-f)}{n}\left(S_y^2 + R^2 S_x^2 - 2RS_{xy}\right) + \frac{W_2(k-1)}{n}\left(S_{y_2}^2 + R^2 S_{x_2}^2 - 2RS_{xy_2}\right)$$

$$MSE(t_{p_1}) = \frac{(1-f)}{n}\left(S_y^2 + R^2 S_x^2 + 2RS_{xy}\right) + \frac{W_2(k-1)}{n}\left(S_{y_2}^2 + R^2 S_{x_2}^2 + 2RS_{xy_2}\right)$$

$$MSE(t_{lr_1}) = \frac{(1-f)}{n} S_y^2 \left(1 - \rho^2\right) + \frac{W_2(k-1)}{n}\left(S_{y_2}^2 + \beta^2 S_{x_2}^2 - 2\beta S_{xy_2}\right)$$

$$MSE(t_{rat_2}) = \frac{(1-f)}{n}\left(S_y^2 + R^2 S_x^2 - 2RS_{xy}\right) + \frac{W_2(k-1)}{n} S_{y_2}^2$$

$$MSE(t_{p_2}) = \frac{(1-f)}{n}\left(S_y^2 + R^2 S_x^2 + 2RS_{xy}\right) + \frac{W_2(k-1)}{n} S_{y_2}^2$$

$$MSE(t_{lr_2}) = \frac{(1-f)}{n} S_y^2 \left(1 - \rho^2\right) + \frac{W_2(k-1)}{n} S_{y_2}^2$$

$$MSE(t_{rat_3}) = \frac{(1-f)}{n}\left(S_y^2 + R^2S_x^2 - 2RS_{xy}\right) + \frac{W_2(k-1)}{n}S_{y_2}^2$$

$$MSE(t_{p_3}) = \frac{(1-f)}{n}\left(S_y^2 + R^2S_x^2 + 2RS_{xy}\right) + \frac{W_2(k-1)}{n}S_{y_2}^2$$

$$MSE(t_{lr_3}) = \frac{(1-f)}{n}S_y^2\,(1-\rho^2) + \frac{W_2(k-1)}{n}S_{y_2}^2$$

$$MSE(t_{rat_4}) = \left(\frac{1}{n'} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)S_r^2 + \frac{W_2(K-1)}{n}S_{r_2}^2$$

$$MSE(t_{p_4}) = \left(\frac{1}{n'} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)S_p^2 + \frac{W_2(K-1)}{n}S_{p_2}^2$$

$$MSE(t_{lr_4}) = \left(\frac{1}{n'} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)S_y^2(1-\rho^2) + \frac{W_2(K-1)}{n}S_{\beta_2}^2$$

$$MSE(t_{rat_5}) = \left(\frac{1}{n'} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)S_r^2 + \frac{W_2(K-1)}{n}S_{y_2}^2$$

$$MSE(t_{p_5}) = \left(\frac{1}{n'} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)S_p^2 + \frac{W_2(K-1)}{n}S_{y_2}^2$$

$$MSE(t_{lr_5}) = \left(\frac{1}{n'} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)S_y^2(1-\rho^2) + \frac{W_2(K-1)}{n}S_{y_2}^2$$

$$MSE(t_{rat_6}) = \left(\frac{1-f}{n}\right)\left[S_y^2 + 4R^2S_x^2(R-\beta)\right] + \frac{W_2(k-1)}{n}\left[S_{y_2}^2 + R^2S_{x_2}^2(R-2\beta_2)\right]$$

$$MSE(t_{p_6}) = \left(\frac{1-f}{n}\right)\left[S_y^2 + 4R^2S_x^2(R+\beta)\right] + \frac{W_2(k-1)}{n}\left[S_{y_2}^2 + R^2S_{x_2}^2(R+2\beta_2)\right]$$

$$MSE(t_d) = \left(\frac{1-f}{n}\right)\left[S_y^2 + d_2S_x^2(d_2-2\beta)\right] + \frac{W_2(k-1)}{n}\left[S_{y_2}^2 + d_1S_{x_2}^2(d_1-2\beta_2)\right]$$

$$MSE(t_g) = \left(\frac{1-f}{n}\right)\left[S_y^2 + R\theta S_x^2(\theta-2\beta)\right] + \frac{W_2(k-1)}{n}\left[S_{y_2}^2 + R\alpha_1S_{x_2}^2(R\alpha_1-2\beta_2)\right]$$

$$MSE(t_{d_d}) = \left[\left(\frac{1}{n} - \frac{1}{n'}\right)S_y^2 + d_2S_x^2(d_2-2\beta) + \frac{W_2(k-1)}{n}S_{y_2}^2 + d_1S_x^2(d_1-2\beta_2) + \left(\frac{1}{n'} - \frac{1}{N}\right)S_y^2\right]$$

$$MSE(t_d^r) = \left[\left(\frac{1}{n} - \frac{1}{n'}\right)S_y^2 + 4RS_x^2(R-\beta) + \frac{W_2(k-1)}{n}S_{y_2}^2 + RS_x^2(R-2\beta_2) + \left(\frac{1}{n'} - \frac{1}{N}\right)S_y^2\right]$$

$$MSE(t_d^p) = \left[\left(\frac{1}{n}-\frac{1}{n'}\right)S_y^2 + 4RS_x^2(R+\beta) + \frac{W_2(k-1)}{n}S_{y_2}^2 + RS_x^2(R+2\beta_2) + \left(\frac{1}{n'}-\frac{1}{N}\right)S_y^2\right]$$

$$MSE(t_{\alpha_1}^{\alpha_2}) = \left[\left(\frac{1}{n}-\frac{1}{n'}\right)S_y^2 + R\alpha_2 S_x^2(R\alpha_2-2\beta) + \frac{W_2(k-1)}{n}S_{y_2}^2 + RS_x^2(R-2\beta_2) + \left(\frac{1}{n'}-\frac{1}{N}\right)S_y^2\right]$$

and the minimum MSE of t_g and t_d are same and is given by

$$min.MSE(t_g) = min.MSE(t_d) = \left(\frac{1-f}{n}\right)S_y^2(1-\rho^2) + \frac{W_2(k-1)}{n}S_{y_2}^2(1-\rho_2^2)$$

and the minimum MSE of $t_{\alpha_1}^{\alpha_2}$ and t_{dd} are same and is given by

$$min.MSE(t_{\alpha_1}^{\alpha_2}) = min.MSE(t_{dd}) = \left(\frac{1-f}{n}\right)S_y^2(1-\rho^2) + \frac{W_2(k-1)}{n}S_{y_2}^2(1-\rho_2^2)$$

The optimum value of the constants are

$d_1 = \beta_2$, $d_2 = \beta$, $\alpha_1 = \beta_2/R$ and $\alpha_2 = (\beta-\beta_2)/R$.

Singh and Kumar (2010b) proposed the following estimators using information based on two auxiliary variables x and z for estimating the population mean of the study variable y under different situation is given by

$$\bar{y}_{SK}^1 = \bar{y}^* + b_{yx}^*(\bar{x}'-\bar{x}^*)\frac{\bar{Z}}{\bar{Z}+n_1(\bar{z}^*-\bar{Z})} \tag{1.3.26}$$

$$\bar{y}_{SK}^2 = \bar{y}^* + b_{yx}^*(\bar{x}'-\bar{x}^*)\frac{\bar{z}'}{\bar{z}'+n_2(\bar{z}^*-\bar{z}')} \tag{1.3.27}$$

$$\bar{y}_{SK}^3 = \bar{y}^* + b_{yx}^{**}(\bar{x}'-\bar{x})\frac{\bar{Z}}{\bar{Z}+n_3(\bar{z}-\bar{Z})} \tag{1.3.28}$$

$$\bar{y}_{SK}^4 = \bar{y}^* + b_{yx}^{**}(\bar{x}'-\bar{x}^*)\frac{\bar{z}'}{\bar{z}'+n_4(\bar{z}^*-\bar{z}')} \tag{1.3.29}$$

where

$\bar{y}^* = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\bar{y}_{(2r)}$, $\bar{x}^* = \frac{n_1}{n}\bar{x}_1 + \frac{n_2}{n}\bar{x}_{(2r)}$, $\bar{z}^* = \frac{n_1}{n}\bar{z}_1 + \frac{n_2}{n}\bar{z}_{(2)}^*$, $\bar{x}' = \frac{1}{n'}\sum x_i$, $\bar{z}' = \frac{1}{n'}\sum z_i$, $\bar{x} = \frac{1}{n}\sum x_i$ and $\bar{z} = \frac{1}{n}\sum z_i$ with $(\bar{y}_1, \bar{x}_1, \bar{z}_1)$ and $(\bar{y}_2^*, \bar{x}_2^*, \bar{z}_2^*)$ being the sample means based on n_1 units and sub-sample means based on r units of the variates $(\bar{y},\ \bar{x},\ \bar{z})$ respectively and $n_1,\ n_2,\ n_3,\ n_4$ are suitably chosen constants.

Further, Shabbir and Khan (2013) proposed the following estimators using information based on two auxiliary variables x and z for estimating the population mean of the study variable y under different situations is given by

$$\bar{y}_{sk}^{5} = \bar{y}^{*}\left(\frac{\bar{x}'}{\bar{x}^{*}}\right)^{\alpha_1}\left(\frac{\bar{Z}}{\bar{z}^{*}}\right)^{\alpha_2} + b_{yx}^{*}(\bar{x}' - \bar{x}^{*}) + b_{yz}^{*}(\bar{Z} - \bar{z}^{*}) \tag{1.3.30}$$

$$\bar{y}_{sk}^{6} = \bar{y}^{*}\left(\frac{\bar{x}'}{\bar{x}^{*}}\right)^{\alpha_3}\left(\frac{\bar{z}'}{\bar{z}^{*}}\right)^{\alpha_4} + b_{yx}^{*}(\bar{x}' - \bar{x}^{*}) + b_{yz}^{*}(\bar{z}' - \bar{z}^{*}) \tag{1.3.31}$$

$$\bar{y}_{sk}^{7} = \bar{y}^{*}\left(\frac{\bar{x}'}{\bar{x}}\right)^{\alpha_5}\left(\frac{\bar{Z}}{\bar{z}}\right)^{\alpha_6} + b_{yx}^{**}(\bar{x}' - \bar{x}) + b_{yz}^{*}(\bar{Z} - \bar{z}) \tag{1.3.32}$$

$$\bar{y}_{sk}^{8} = \bar{y}^{*}\left(\frac{\bar{x}'}{\bar{x}}\right)^{\alpha_7}\left(\frac{\bar{z}'}{\bar{z}}\right)^{\alpha_8} + b_{yx}^{**}(\bar{x}' - \bar{x}) + b_{yz}^{**}(\bar{z}' - \bar{z}) \tag{1.3.33}$$

where $\alpha_1,\ \alpha_2,\ \alpha_3,\ \alpha_4,\ \alpha_5,\ \alpha_6,\ \alpha_7,\ \alpha_8$ are the characterizing scalars to be chosen suitably and $b_{yx}^{*} = s_{yx}^{*}/s_{x}^{*2},\ b_{yz}^{*} = s_{yz}^{*}/s_{z}^{*2},\ b_{yx}^{**} = s_{yx}^{*}/s_{x}^{2},\ b_{yz}^{**} = s_{yz}^{*}/s_{z}^{2}$ are the sample regression coefficient.

The minimum MSE's of these estimators are given below

$$MSE(\bar{y}_{sk}^{1})_{min} = MSE(\bar{y}_{lr_1}) - \bar{Y}^2\left[\frac{D_1^2}{D_2}\right]$$

$$MSE(\bar{y}_{sk}^{2})_{min} = MSE(\bar{y}_{lr_1}) - \bar{Y}^2\left[\frac{D_3^2}{D_2}\right]$$

$$MSE(\bar{y}_{sk}^{3})_{min} = MSE(\bar{y}_{lr_2}) - \bar{Y}^2\left[\frac{D_4^2}{D_5}\right]C_z^2$$

$$MSE(\bar{y}_{sk}^{4})_{min} = MSE(\bar{y}_{lr_2}) - \bar{Y}^2 f' A^{*2} C_z'^2$$

$$MSE(\bar{y}_{sk}^{5})_{min} = MSE(\bar{y}^{*}) - \bar{Y}^2\frac{(m_2 m_3^2 + m_1 m_4^2 - 2m_3 m_4 m_5)}{(m_1 m_2 - m_5^2)}$$

$$MSE(\bar{y}_{sk}^{6})_{min} = MSE(\bar{y}^{*}) - \bar{Y}^2\frac{(m_2' m_3^2 + m_1 m_4'^2 - 2m_3 m_4 m_5)}{(m_1 m_2' - m_5^2)}$$

$$MSE(\bar{y}_{sk}^{7})_{min} = MSE(\bar{y}^{*}) - \bar{Y}^2 f_n\left[\frac{f_n\rho_{yz}^2 + f_n'(\rho_{yx}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz})}{f_n - (f_n - f_n')\rho_{yx}^2}\right]C_y^2$$

$$MSE(\bar{y}^8_{sk})_{min} = MSE(\bar{y}^*) - \bar{Y}^2(f_n - f_n{}') \left[\frac{\rho^2_{yx} + \rho^2_{yz} - 2\rho_{yx}\rho_{yz}\rho_{xz}}{1-\rho^2_{xz}}\right] C_y^2$$

where

$R = \frac{\bar{Y}}{\bar{X}}$, $C_x^2 = \frac{S_x^2}{\bar{X}^2}$, $C_{x_2}^2 = \frac{S_{x_2}^2}{\bar{X}^2}$, $C_y^2 = \frac{S_y^2}{\bar{Y}^2}$, $C_{y_2}^2 = \frac{S_{y_2}^2}{\bar{Y}^2}$, $\rho_{yx} = \frac{S_{yx}}{S_y S_x}$, $\rho_{yx_2} = \frac{S_{yx_2}}{S_{y_2} S_{x_2}}$, $\rho_{zx} = \frac{S_{zx}}{S_z S_x}$, $\rho_{zx_2} = \frac{S_{zx_2}}{S_{z_2} S_{x_2}}$, $\rho_{yz} = \frac{S_{yz}}{S_y S_z}$, $\rho_{yz_2} = \frac{S_{yz_2}}{S_{y_2} S_{z_2}}$, $K_{yx} = \frac{\rho_{yx}}{C_y C_x}$, $K_{yx_2} = \frac{\rho_{yx_2}}{C_{y_2} C_{x_2}}$, $K_{zx} = \frac{\rho_{zx}}{C_z C_x}$, $K_{zx_2} = \frac{\rho_{zx_2}}{C_{z_2} C_{x_2}}$, $K_{yz} = \frac{\rho_{yz}}{C_y C_z}$, $K_{yz_2} = \frac{\rho_{yz_2}}{K_{y_2} K_{z_2}}$, $A^* = K_{yz} - K_{yx}K_{xz}$, $B^* = K_{yz_2} - K_{yx}K_{xz_2}$.
$m_1 = (f_n - f_n{}')C_x^2 + g_n C_{x_2}^2$, $m_2 = f_n C_z^2 + g_n C_{z_2}^2$, $m_3 = (f_n - f_n{}')\rho_{yx} C_y C_x + g_n \rho_{yx_2} C_{y_2} C_{x_2}$, $m_4 = f_n \rho_{yz} C_y C_z + g_n \rho_{yz_2} C_{y_2} C_{z_2}$, $m_5 = (f_n - f_n{}')\rho_{zx} C_z C_x + g_n \rho_{zx_2} C_{z_2} C_{x_2}$, $m_6 = f_n C_z^2$, $m_7 = f_n \rho_{yz} C_y C_z$, $m_1{}' = (f_n - f_n{}')C_x^2$, $m_2{}' = (f_n - f_n{}')C_z^2 + g_n C_{z_2}^2$, $m_3{}' = (f_n - f_n{}')\rho_{yx} C_y C_x$, $m_4{}' = (f_n - f_n{}')\rho_{yz} C_y C_z + g_n \rho_{yz_2} C_{y_2} C_{z_2}$, $m_5{}' = (f_n - f_n{}')\rho_{zx} C_z C_x$, $m_6{}' = (f_n - f_n{}')C_z^2$, $m_7{}' = (f_n - f_n{}')\rho_{yz} C_y C_z$, $D_1 = f_n{}' K_{yz} C_z^2 + (f_n - f_n{}')A^* C_z^2 + g_n B^* C_{z_2}^2$, $D_2 = f_n C_z^2 + g_n C_{z_2}^2$, $D_3 = (f_n - f_n{}')A^* C_z^2 + g_n B^* C_{z_2}^2$, $D_4 = f_n{}' K_{yz} + (f_n - f_n{}')A^*$, $D_5 = f_n$.

Following the work of Singh and kumar (2010b) and Shabbir and Khan (2013), Bhushan and Naqvi (2015) proposed the following cost efficient estimators under certain regularity condition using information based on two auxiliary variables for estimating the population mean of the study variable in presence of non-response under different situations.

In first situation, non-response occurs on the study variables as well as on the both the auxiliary variables and yhe population mean $\bar{X}$ of the first auxiliary variable x is unknown however the population mean $\bar{Z}$ of the second auxiliary variable z is known. The estimator for estimating the population mean in this situation is given by

$$t_{g_{n_1}} = f_1(\bar{y}^*, \bar{x}^*, \bar{z}^*, u) \tag{1.3.34}$$

where $u = \frac{\bar{z}^*}{\bar{Z}}$.

In second situation, the non-response occurs on the study variables as well as on the both the auxiliary variables and the population mean $\bar{X}$ of the first auxiliary variable x and population mean $\bar{Z}$ of the second auxiliary variable z both are unknown. The estimator for estimating the population mean in this situation is given by

$$t_{g_{n_2}} = f_2(\bar{y}^*, \bar{x}^*, \bar{x}', \bar{z}^*, \bar{z}') \tag{1.3.35}$$

In third situation, the non-response occurs on the study only and the population mean $\bar{X}$ of the first auxiliary variable x is unknown however the population mean $\bar{Z}$ of the second auxiliary variable z is known. The estimator for estimating the population mean in this situation is given by

$$t_{g_{n_3}} = f_3(\bar{y}^*, \bar{x}, \bar{x}', v) \tag{1.3.36}$$

where $u = \frac{\bar{z}}{\bar{Z}}$.

In fourth situation, the non-response occurs on the study variables only and population mean of both the auxiliary variables are unknown. The estimator for estimating the population mean in this situation is given by

$$t_{g_{n_4}} = f_4(\bar{y}^*, \bar{x}, \bar{x}', \bar{z}, \bar{z}') \tag{1.3.37}$$

satisfying the following conditions
$f_1\left(\bar{Y}, \bar{X}, \bar{X}, 1\right) = \bar{Y}$, $f_1^0 = 1$, $f_1^1 = -f_1^2$, $f_2\left(\bar{Y}, \bar{X}, \bar{X}, \bar{Z}, \bar{Z}\right) = \bar{Y}$, $f_2^0 = 1$, $f_2^1 = -f_2^2$, $f_2^3 = -f_2^4$, $f_3\left(\bar{Y}, \bar{X}, \bar{X}, 1\right) = \bar{Y}$, $f_3^0 = 1$, $f_3^1 = -f_3^2$, $f_4\left(\bar{Y}, \bar{X}, \bar{X}, \bar{Z}, \bar{Z}\right) = \bar{Y}$, $f_4^0 = 1$, $f_4^1 = -f_4^2$ and $f_4^3 = -f_4^4$.

The minimum mean square error of the estimators are given below

$$MSE(t_{g_{n_1}})_{min} = MSE(\bar{y}^*) - \bar{Y}^2 \frac{(2m_3m_4m_5 - m_1m_4^2 - m_2m_3^2)}{(m_5^2 - m_1m_2)}$$

$$MSE(t_{g_{n_2}})_{min} = MSE(\bar{y}^*) - \bar{Y}^2 \frac{(2m_3{m_4}'m_5 - m_1m_4'^2 - {m_2}'m_3^2)}{(m_5^2 - m_1{m_2}')}$$

$$MSE(t_{g_{n_3}})_{min} = MSE(\bar{y}^*) - \bar{Y}^2 \frac{(2{m_3}'{m_5}'m_7 - {m_1}'m_7^2 - m_6m_3'^2)}{(m_5'^2 - {m_1}'m_6)}$$

$$MSE(t_{g_{n_4}})_{min} = MSE(\bar{y}^*) - \bar{Y}^2 \frac{(2{m_3}'{m_5}'{m_7}' - {m_1}'m_7'^2 - {m_6}'m_3'^2)}{(m_5'^2 - {m_1}'{m_6}')}$$

1.4 Random non-response

The problem of random non-response was discussed by several authors. Some authors suggested that by using the auxiliary information, we can improve the efficiency of the estimators, see Singh and Joarder (1998) suggested an estimator of finite population variance using random non-response in survey sampling, Singh and Singh (1979, 1985) suggested an estimators by using double sampling and unequal sampling under random non-response, Singh et al. (2000) proposed a regression type estimators for random non-response in survey sampling. Singh and Tracy (2001) suggested an estimator of population mean in presence of random non-response, Singh et al. (2007) and Singh et al. (2012) suggested a families of estimators for mean, ratio and product of finite population and finite population variance under random non-response.

Notations and results

Let $\Omega = (1, 2, ..., N)$ denote a population of N units from which a simple random sample of size n is drawn without replacement. If $r\,(r = 0, 1, 2, ..., (n-2))$ denotes the number of sampling units on which information could not be obtained due to a random non-response, then the remaining $(n-r)$ units can be treated as a simple random sample from the population. It is assumed that r is less than $(n-1)$ i.e. $0 \leq r \leq (n-2)$ for obvious reasons. Singh and Joarder (1998) assumed that r has the following discrete distribution as

$$P(r) = \frac{n-r}{(nq+2p)}\binom{n-2}{r}p^r q^{n-2-r} \tag{1.4.1}$$

where p is the probability of non-response, $q = 1-p$ and $\binom{n-2}{r}$ represents that total number ways to obtain r non-response out of a possible $(n-2)$. We define

For the variate y_i and x_i

$\bar{X} = N^{-1}\sum\limits_{i=1}^{N} x_i$: Population mean of the ith variate x_i.

$\bar{Y} = N^{-1}\sum\limits_{i=1}^{N} y_i$: Population mean of the ith variate y_i.

C_x, C_y: Population coefficient of variation (CV) of the variate y and x.

$\rho_{il} = S_{yx}/S_y S_x$: Population correlation coefficient between the variates y and x.

$(n-1)\,S_{yx} = \sum\limits_{i=1}^{N}\left(y_i - \bar{Y}_i\right)\left(x_i - \bar{X}\right)$;$K_{xy} = \rho_{xy}\frac{C_y}{C_x}$

$K = (K_{12})$,$d = (\lambda_{12}C_y + \lambda_{12}C_x)$

$\lambda_{w_1,w_2,} = \mu_{w_1,w_2,}/\,(\mu_{20})^{\frac{w_1}{2}}\,(\mu_{02})^{\frac{w_2}{2}}$

$(N-1)\,\mu_{w_1,w_2} = \sum\limits_{i=1}^{N}\left(y_i - \bar{Y}\right)^{w_1}\left(x_i - \bar{X}\right)^{w_2}$

where w_1 and w_2 are non-negative integers

Define the following terms

$B = C_x^2K^2 + \frac{(C_xK\lambda_{03}-d)^2}{\left(\lambda_{04}-\lambda_{03}^2-1\right)}$ or $B = (C_x^2K^2 + \Delta B_1^2)$

where $B_1 = \frac{\Delta_2}{\Delta}$, $\Delta_2 = (C_xK\lambda_{03} - d)$ and $\Delta = (\lambda_{04} - \lambda_{03}^2 - 1)$

Let n units be selected from N units by SRSWOR and without loss of generality, we assume that first n units have been selected in the sample thus for the variate y and x, we define

$\bar{y} = n^{-1}\sum\limits_{i=1}^{n} y_i$, $\bar{y}^* = (n-r)^{-1}\sum\limits_{i=1}^{n-r} y_i$, $s_x^2 = (n-1)^{-1}\sum\limits_{i=1}^{n}(x_i - \bar{x})^2$,

$s_x^{*2} = (n-1)^{-1}\sum\limits_{i=1}^{n-r}(x_i - \bar{x}^*)^2$ are conditionally unbiased estimators of S_x^2 respectively.

where $\theta^* = \left(\frac{1}{nq+2p} - \frac{1}{N}\right)$and $\theta = \left(\frac{1}{n} - \frac{1}{N}\right)$

Singh and Joarder (1998) obtained the following maximum likelihood estimator of p (probability of non-response), given by

$$\hat{p} = \frac{(n-1+r) - \sqrt{(n-1+r)^2 - \frac{4rn(n-3)}{(n-2)}}}{2\,(n-3)} \tag{1.4.2}$$

If $r = 0$ then $\hat{p} = 0$, and if $r = n-2$ then $\hat{p} = 1$; thus, $\hat{p}$ is an admissible estimators of response probability p.

Let us define

$\bar{y}^* = \bar{Y}\,(1+\varepsilon_0)$, $\bar{x}^* = \bar{X}\,(1+\varepsilon_1)$, $\bar{x} = \bar{X}\,(1+\varepsilon_2)$, $s_x^{*2} = S_x^2\,(1+\varepsilon_3)$, $s_x^2 = S_x^2\,(1+\varepsilon_4)$ and $s_y^{*2} = S_y^2\,(1+\varepsilon_5)$

Then, under the model

$E\,(\varepsilon_i) = 0, (i = 0, 1, ..., 5)$

$E\,(\varepsilon_0^2) = \theta^*C_y^2$, $E\,(\varepsilon_1^2) = \theta^*C_x^2$, $E\,(\varepsilon_2^2) = \theta C_x^2$, $E\,(\varepsilon_3^2) = \theta^*\,(\lambda_{04}-1)$, $E\,(\varepsilon_4^2) = \theta\,(\lambda_{04}-1)$, $E\,(\varepsilon_5^2) = \theta^*\,(\lambda_{40}-1)$, $E\,(\varepsilon_0\varepsilon_1) = \theta^*\rho_{yx}C_yC_x$, $E\,(\varepsilon_0\varepsilon_2) = \theta\rho_{yx}C_yC_x$, $E\,(\varepsilon_0\varepsilon_3) = \theta^*\lambda_{12}C_y$, $E\,(\varepsilon_0\varepsilon_4) = \theta\lambda_{12}C_y$, $E\,(\varepsilon_1\varepsilon_2) = \theta C_x^2$, $E\,(\varepsilon_1\varepsilon_3) = \theta^*\lambda_{03}C_x$, $E\,(\varepsilon_1\varepsilon_4) = \theta\lambda_{03}C_x$, $E\,(\varepsilon_1\varepsilon_5) = \theta^*\lambda_{21}C_x$, $E\,(\varepsilon_2\varepsilon_3) = \theta\lambda_{03}C_x$, $E\,(\varepsilon_2\varepsilon_4) = \theta\lambda_{03}C_x$, $E\,(\varepsilon_2\varepsilon_5) = \theta\lambda_{21}C_x$, $E\,(\varepsilon_3\varepsilon_4) = \theta\,(\lambda_{04}-1)$,

$E(\varepsilon_3\varepsilon_5) = \theta^*(\lambda_{22} - 1)$ and $E(\varepsilon_4\varepsilon_5) = \theta(\lambda_{22} - 1)$

The estimated value of the parameters are given below

$\hat{\Delta} = \left(\hat{\lambda}_{04} - \hat{\lambda}_{03}^2 - 1\right)$, $\hat{\Delta}^* = \left(\hat{\lambda}_{04}^* - \hat{\lambda}_{03}^{*2} - 1\right)$, $\hat{\Delta}_1^* = \left[\hat{d}^*\hat{\lambda}_{03}^* - \hat{K}^*\left(\hat{\lambda}_{04}^* - 1\right)C_x\right]$, $\hat{\Delta}_2^* = \left[\hat{d}^*\hat{\lambda}_{03}^* C_x - \hat{d}^*\right]$, $\hat{K}^* = \hat{K}_{yx}^*$, $\hat{K}_{yx}^* = \dfrac{\hat{\rho}_{yx}^*\hat{C}_y^*}{C_x}$, $\hat{\rho}_{yx}^* = \dfrac{\hat{\mu}_{11}^*}{\sqrt{\hat{\mu}_{20}^*\hat{\mu}_{02}^*}}$, $\hat{C}_y^* = \sqrt{\dfrac{\hat{\mu}_{20}^*}{\bar{y}^*}}$ and $\hat{C}_x^* = \sqrt{\dfrac{\hat{\mu}_{02}^*}{\bar{x}^*}}$.

We have studied the effect of random non-response on the study and auxiliary variables of several estimators of variance under the following three strategies considered by Singh and Joarder (1998).

Strategy I: We are considering the situation when random non-response exists on both the study variable y and the auxiliary variable x and population variance S_x^2 ofthe auxiliary character is known.

Strategy II: Here we are considering the situation when information on variable y could not be obtained for r units while information on variable x is available and population variance S_x^2 of the auxiliary variable is known.

Strategy III: Here we again consider the situation when information on variable y could not be obtained for r units while information on the variable x is obtained for all the sample units, but the population variance S_x^2 of the auxiliary variable is unknown.

Let us define

$s_y^{*2} = S_y^2(1 + \varepsilon_0)$, $s_x^{*2} = S_x^2(1 + \varepsilon_1)$ and $s_x^2 = S_x^2(1 + \varepsilon_2)$

Then, under the model

$E(\varepsilon_i) = 0, (i = 0, 1, 2)$ $E(\varepsilon_0^2) = \theta^*(\lambda_{40} - 1)$, $E(\varepsilon_1^2) = \theta^*(\lambda_{04} - 1)$, $E(\varepsilon_2^2) = \theta(\lambda_{04} - 1)$, $E(\varepsilon_0\varepsilon_1) = \theta^*(\lambda_{22} - 1)$, $E(\varepsilon_0\varepsilon_2) = \theta(\lambda_{22} - 1)$ and $E(\varepsilon_1\varepsilon_2) = \theta(\lambda_{04} - 1)$.

Strategy I: When s_y^{*2}, s_x^{*2} and S_x^2 are used.

Under this strategy, Singh and Joarder (1998) suggested the following estimator of finite population variance.

$$t_1' = s_y^{*^2}\left(\frac{S_x^2}{s_x^{*2}}\right) \tag{1.4.3}$$

For this strategy, Ahmeda et al. (2005) proposed the following estimators for finite population variance

$$t_{v_1} = s_y^{*^2}\left(\frac{S_x^2}{s_x^{*2}}\right)^{\alpha_1} \quad (1.4.4)$$

$$t_{v_2} = s_y^{*^2} + k_1\left(S_x^2 - s_x^{*2}\right) \quad (1.4.5)$$

$$t_{v_3} = \frac{s_y^{*^2} S_x^2}{\theta_1 s_x^{*2} + (1-\theta_1) S_x^2} \quad (1.4.6)$$

where α_1,k_1 and θ_1 are suitably chosen constants.

Strategy II: When s_y^{*2}, s_x^2 and S_x^2 are used.

Under this strategy, Singh and Joarder (1998) proposed the following estimator of finite population variance

$$t_2^{'} = s_y^{*^2}\left(\frac{S_x^2}{s_x^2}\right) \quad (1.4.7)$$

$$t_4^{'} = s_y^{*^2}\left(\frac{S_x^2}{s_x^2}\right) + \alpha\left(\frac{s_x^{*2}}{S_x^2} - 1\right) \quad (1.4.8)$$

$$t_5^{'} = s_y^{*^2}\left(\frac{S_x^2}{s_x^2}\right)\left(\frac{s_x^{*2}}{S_x^2}\right)^{\alpha} \quad (1.4.9)$$

under this strategy, Ahmed et al. (2005) proposed the following estimators for finite population variance

$$t_{v_4} = s_y^{*^2}\left(\frac{S_x^2}{s_x^2}\right)^{\alpha_2} \quad (1.4.10)$$

$$t_{v_5} = s_y^{*^2} + k_2\left(S_x^2 - s_x^2\right) \quad (1.4.11)$$

$$t_{v_6} = \frac{s_y^{*^2} S_x^2}{\theta_2 s_x^2 + (1-\theta_2) S_x^2} \quad (1.4.12)$$

where α_2,k_2 and θ_2 are suitably chosen constants.

Strategy III: When s_y^{*2}, s_x^{*2} and s_x^2 are used.

Under strategy III, Singh and Joarder (1998) proposed the following estimator of finite population variance

$$t_3^{'} = s_y^{*^2}\left(\frac{s_x^2}{s_x^{*2}}\right) \quad (1.4.13)$$

For this strategy, Ahmed et al. (2005) proposed the following estimators for finite population variance

$$t_{v_7} = s_y^{*^2}\left(\frac{s_x^2}{s_x^{*2}}\right)^{\alpha_1} \tag{1.4.14}$$

$$t_{v_8} = s_y^{*^2} + k_3\left(s_x^2 - s_x^{*2}\right) \tag{1.4.15}$$

$$t_{v_9} = \frac{s_y^{*^2} s_x^2}{\theta_3 s_x^{*2} + (1-\theta_3)\, s_x^2} \tag{1.4.16}$$

Similarly, a class of estimator for mean under non- response model. Singh et al. (2007) suggested a family of estimators of mean, ratio and product of a finite population under three different strategy. The usual estmators of the ratio $R_{(\alpha)} = \frac{\bar{Y}_0}{\bar{Y}_1^{\alpha}}(\bar{Y}_1 \neq 0)$ is defined by

$$\hat{R}_{(\alpha)} = \frac{\bar{y}_0}{\bar{y}_1^{\alpha}}(\bar{y}_1 \neq 0) \tag{1.4.17}$$

where α is a scalar which takes the values 0, 1 and -1. It is to be mentioned that:

(1). For $\alpha = 0$, $R_{(\alpha)} \to R_{(0)} = \bar{Y}_0$, and its estimator $\hat{R}_{(\alpha)} \to \hat{R}_{(0)} = \bar{y}_0$,

(2). For $\alpha = 1$, $R_{(\alpha)} \to R_{(1)} = \frac{\bar{Y}_0}{\bar{Y}_1}$, and its estimator $\hat{R}_{(\alpha)} \to \hat{R}_{(1)} = \frac{\bar{y}_0}{\bar{y}_1} = \hat{R}$,

(2). For $\alpha = -1$, $R_{(\alpha)} \to R_{(-1)} = \bar{Y}_0\bar{Y}_1$, and its estimator $\hat{R}_{(\alpha)} \to \hat{R}_{(-1)} = \bar{y}_0\bar{y}_1 = \hat{P}$.

Under ***Strategy I***, we consider the situation when random non-response for r units on study variable y_0, y_1 and auxiliary variable x are present in the sample, and the population mean $\bar{X}$ and the variance S_x^2 of x are known, we defined a family of estimators of $R_{(\alpha)}$ as

$$t_{m_1} = R_{(\alpha)}^{*} f\left(u^{\bullet}, v^{\bullet}\right) \tag{1.4.18}$$

where $f\left(u^{\bullet}, v^{\bullet}\right)$ is a function of $(u^{\bullet}, v^{\bullet})$, where $u^{\bullet} = \frac{\bar{x}^*}{\bar{X}}$ and $v^{\bullet} = \frac{s_x^{*2}}{S_x^2}$, such that $f\left(1,1\right) = 1$ and satisfying certain regularity conditions, as defined below

1. Whatever the sample $(u^{\bullet}, v^{\bullet})$ assumes values in a bounded, closed convex subset, S, of the two dimenstional real space containing the point $(1,1)$.

2. In S the function $f\left(u^{\bullet}, v^{\bullet}\right)$ is continuous and bounded.

3. The first and second order partial derivatives of $f\left(u^{\bullet}, v^{\bullet}\right)$ exist as well as are continuous and bounded in S.

Under ***Strategy II***, we consider the situation when information on variable y_0 and y_1 can not be obtained for r units, while the population mean $\bar{X}$ and the variance S_x^2 of the auxiliary variable x are known. we propose the following family of estimator of $R_{(\alpha)}$.

$$t_{m_2} = R^*_{(\alpha)}\phi\left(u, v\right) \tag{1.4.19}$$

where $\phi\left(u, v\right)$ is a function of $\left(u, v\right)$, where $u = \frac{\bar{x}}{\bar{X}}$ and $v = \frac{s_x^2}{S_x^2}$, such that $\phi\left(1, 1\right) = 1$ and satisfying certain regularity conditions, as defined above strategy I.

Under ***Strategy III***, we consider the situation when information on study variable y_0 and y_1 can not be obtained for r units while information on the auxiliary variable x is obtained for all the sample units. but the population mean $\bar{X}$ and the variance S_x^2 of the auxiliary character x are not known. under these circumstances we define the following class of estimator of $\bar{Y}$

$$t_{m_3} = R^*_{(\alpha)}\psi\left(u^*, v^*\right) \tag{1.4.20}$$

where $\psi\left(u^*, v^*\right)$ is a function of $\left(u^*, v^*\right)$, where $u = \frac{\bar{x}^*}{\bar{x}}$ and $v = \frac{s_x^{*2}}{s_x^2}$, such that $\psi\left(1, 1\right) = 1$ and satisfying certain regularity conditions, as defined above strategy 1.

Similarly, Singh et al. (2012) suggested a families of estimators of finite population variance using a random non-response in survey sampling under three different strategy.

Under ***Strategy I***, we consider the situation when random non-response for r units on study variable y and auxiliary variable x are present in the sample, and the population mean $\bar{X}$ and the variance S_x^2 of x are known, we defined a family of estimators of S_y^2 as

$$t_{v_{11}} = s_y^{*2} f\left(u^\bullet, v^\bullet\right) \tag{1.4.21}$$

where $f\left(u^\bullet, v^\bullet\right)$ is a function of $\left(u^\bullet, v^\bullet\right)$ (where $u^\bullet = \frac{\bar{x}^*}{\bar{X}}$ and $v^\bullet = \frac{s_x^{*2}}{S_x^2}$),such that $f\left(1, 1\right) = 1$ and satisfying certain regularity conditions, as defined below

1. Whatever the sample $\left(u^\bullet, v^\bullet\right)$ assumes values in a bounded, closed convex subset, S, of the two dimenstional real space containing the point $\left(1, 1\right)$.

2. In S the function $f\left(u^\bullet, v^\bullet\right)$ is continuous and bounded.

3. The first and second order partial derivatives of $f\left(u^\bullet, v^\bullet\right)$ exist as well as are continuous

and bounded in S.

Under ***Strategy II***, we consider the situation when information on variable y can not be obtained for r units, while the population mean $\bar{X}$ and the variance S_x^2 of the auxiliary variable x are known. we propose the following family of estimator of S_y^2

$$t_{v_{12}} = s_y^{*2}\phi(u, v) \tag{1.4.22}$$

where $\phi(u, v)$ is a function of (u, v), where $u = \frac{\bar{x}}{\bar{X}}$ and $v = \frac{s_x^2}{S_x^2}$, such that $\phi(1, 1) = 1$ and satisfying certain regularity conditions, as defined above strategy I.

Under ***Strategy III***, we consider the situation when information on study variable y can not be obtained for r units while information on the auxiliary variable x is obtained for all the sample units, but the population mean $\bar{X}$ and the variance S_x^2 of the auxiliary character x are not known. Under these circumstances we define the following class of estimator of S_y^2

$$t_{v_{13}} = s_y^{*2}\psi(u^*, v^*) \tag{1.4.23}$$

where $\psi(u^*, v^*)$ is a function of (u^*, v^*) where $u = \frac{\bar{x}^*}{\bar{x}}$ and $v = \frac{s_x^{*2}}{s_x^2}$, such that $\psi(1, 1) = 1$ and satisfying certain regularity conditions, as defined above strategy I.

The MSE of the above discussed estimators are given below

$MSE(t_1^{'}) = \theta^* S_y^4 (\lambda_{40} + \lambda_{04} - 2\lambda_{22})$

$minMSE(t_{v_1}) = minMSE(t_{v_2}) = minMSE(t_{v_3}) = \theta^* S_y^4 \left(\lambda_{40} - \frac{\lambda_{22}^2}{\lambda_{04}}\right)$

$MSE(t_2^{'}) = \theta S_y^4 (\lambda_{40} + \lambda_{04} - 2\lambda_{22}) + \theta^* S_y^4$

$$min.MSE(t_4^{'}) = MSE(t_2^{'}) - \frac{\left[\frac{1}{(nq+2p)} - \frac{1}{n}\right]^2 S_y^4(\lambda_{22}-1)^2}{\theta^*(\lambda_{04}-1)}$$

$$min.MSE(t_{v_4}) = min.MSE(t_{v_5}) = min.MSE(t_{v_6}) = S_y^4\left(\theta^*\lambda_{40} - \theta\frac{\lambda_{22}^2}{\lambda_{04}}\right)$$

$$MSE(t_3^{'}) = \left[\theta^*(\lambda_{04} - 2\lambda_{22} + 1) + \left(\frac{1}{n} + \frac{1}{(nq+2p)} - \frac{2}{N}\right)(\lambda_{40} - 1)\right] S_y^4$$

$$min.MSE(t_{v_7}) = min.MSE(t_{v_8}) = min.MSE(t_{v_9}) = S_y^4\left[\theta^*\lambda_{40} - \tfrac{(\theta^*-\theta)^2\lambda_{22}^2}{(\theta^*+\theta)\lambda_{04}-2\theta\lambda_{22}}\right]$$

optimum values of the constant are given below

$$\alpha_1 = \alpha_2 = \frac{\lambda_{22}^2}{\lambda_{04}} \text{ and } \alpha_3 = \frac{(\theta^*-\theta)\lambda_{22}}{(\theta^*+\theta)\lambda_{04} - 2\theta\lambda_{22}}.$$

$$k_1 = k_2 = \frac{S_y^2\lambda_{22}^2}{S_x^2\lambda_{04}} \text{ and } k_3 = \frac{S_y^2(\theta^*-\theta)\lambda_{22}}{S_x^2(\theta^*+\theta)\lambda_{04} - 2\theta\lambda_{22}}.$$

$$\theta_1 = \theta_2 = \frac{\lambda_{22}^2}{\lambda_{04}} \text{ and } \theta_3 = \frac{(\theta^*-\theta)\lambda_{22}}{(\theta^*+\theta)\lambda_{04} - 2\theta\lambda_{22}}.$$

The MSE are given by

$$MSE(\hat{R}_\alpha^*) = \theta^*\hat{R}_\alpha^2\left[C_0^2 + \alpha C_1^2(\alpha - 2K_{01})\right]$$

$$min.MSE(t_{m_1}) = MSE(\hat{R}_\alpha^*) - \theta^* R_\alpha^2 B_\alpha$$

$$min.MSE(t_{m_2}) = MSE(\hat{R}_\alpha^*) - \theta R_\alpha^2 B_\alpha$$

$$min.MSE(t_{m_3}) = MSE(\hat{R}_\alpha^*) - (\theta^*-\theta) R_\alpha^2 B_\alpha$$

$$min.MSE(t_{v_{11}}) = \theta^* S_y^4\left[\lambda_{40} - 1 - D\right]$$

$$min.MSE(t_{v_{12}}) = MSE(d_2) + (\theta^*-\theta)(\lambda_{40}-1)S_y^4$$

$$min.MSE(t_{v_{13}}) = S_y^4\left[\theta^*(\lambda_{40}-1) - (\theta^*-\theta)\left(\lambda_{21}^*2 + \frac{(\lambda_{21}^*\lambda_{03} - \lambda_{22}^* + 1)^2}{\lambda_{04} - \lambda_{03}^2 - 1}\right)\right]$$

1.5 Thesis plan

In this chapter, we provide some introductory review on the theory of non-response, imputation and random non-response. Some relevant and important results are stated in this chapter. This chapter also consists of the basic definitions and notations relevant to this thesis.

In **Chapter 2**, we have proposed various Searls-type ratio and difference imputation methods on the lines of Ahmad et al. (2006). It is a well known fact that the optimal ratio type estimator attains the MSE of regression estimator (or optimal difference estimator) but while using Searls-type transformation (1964) this may not always happen. These imputation methods are shown to perform better than the imputation procedures of Ahmed et al. (2006). These difference imputation methods also perform better than the imputation procedure of Diana and Perri (2010). The proposed ratio imputation methods may even outperform the proposed difference imputation methods under certain optimality condition. This study is concluded with the numerical study alongside the theoretical comparison.

Chapter 3 introduces some improved methods of imputation using higher order moment of an auxiliary variable while imputing missing values. The performance of the proposed imputation methods are investigated relative to the estimators proposed by Mohamed et al. (2017) and by Bhushan and Pandey (2016). A comparative study has been carried out and it has been shown that the proposed estimators perform better in comparison to estimators proposed by Mohamed et al. (2017) and Bhushan and Pandey (2016). The theoretical findings are supported by an empirical study on real populations and a simulation study using hypothetical situation.

The **Chapter 4** proposes some new imputation methods by extending the work of Bhushan and Pandey discussed in chapter 2 and chapter 3 using multi-auxiliary information. The popularly used imputation like mean imputation, ratio method of imputation, regression method of imputation and power transformation method are special cases of the proposed methods apart from being less efficient than the proposed methods. The proposed imputation methods can be considered as an efficient extension to the work of Singh and Deo (2003), Singh (2009), Ahmed et al. (2006), Diana and Perri (2010) and Bhushan and Pandey (2016). The theoretical results are derived and comparative study is conducted and the results are found to be quite encouraging providing the improvement over the all the discussed works.

In **Chapter 5**, we have proposed some efficient estimators of population mean in presence of non-response. These estimators are suggested for both single phase sampling and two phase sampling. The estimators of mean obtained from proposed technique remains better than the estimators obtained by Cochran (1977), Khare and Srivastava (1995), Rao (1983, 86) and Singh and Kumar (2008, 10) under the derived optimality condition. The mean squared error of the resultant estimators is found to be less than that of the MSE of the estimators proposed Cochran (1977), Khare and Srivastava (1995), Rao (1983, 86) and Singh and Kumar (2008, 10) in an empirical study conducted on three populations.

The **Chapter 6** proposes some new classes of estimators of population mean under non-response using bivariate auxiliary information. We have considered two auxiliary variables which are more efficient for the study character. It is observed that the proposed estimators are better perform as comparison to conventional estimators proposed by Singh and Kumar (2010b), Sabbir and Khan (2013) and Bhushan and Naqvi (2015) in order to support the results a comparative study is also carried out both theoretically as well as empirically.

In **Chapter 7**, we have constructed some improved difference and ratio type estimators of the population mean under two different situation of random non-response considered by Tracy and Osahan (1994) using Searls (1964) philosophy. The proposed difference and ratio type estimators remain better than the estimators obtained by Singh et al. (2007) in presence of random non-response. A comparative study has been performed and obtained some optimality conditions under which the conventional conclusion is reversed. It has been shown that the proposed estimators perform better in comparison to conventional estimators.

Chapter 8 proposes some new estimators of the population variance in presence of random non-response based on Searls (1964) philosophy. The proposed estimators remain better than the estimators obtained by Ahmed et al. (2005) in presence of random non-response using auxiliary variables. A comparative study has been performed and conditions for optimality have been obtained. It has been shown that the proposed estimators perform better in comparison to conventional estimators.

Chapter 2

On optimal imputation methods for estimation of population mean

2.1 Introduction

In many sample surveys, we encounter the problem of missing data due to various reasons. One of the popular way to tackle the problem of missing data is through imputation. The problem of missing data can be dealt either at the estimation stage using imputation or at the sampling design stage by using a sub-sampling design. The problem of missing data was first tackled by Hansen and Hurwitz (1946) at the sampling design stage and later modified and extended by El-badry et al. (1956), Srinath (1971) etc. among others in a mailed questionaire survey. The missing data adversely affects the accuracy of the results. The problem of missing data was tackled at estimation stage by various authors and different type of imputation procedures were adopted to impute (fill-in) the missing data at the estimation stage. Many authors have stressed over the role of auxiliary information for imputation in the presence of missing data.

Rubin (1976) and Kalton et al. (1981) have suggested various imputation methods which make the data structurally complete. Rubin (1976) in his seminal work introduced the important concept of missing at random (MAR). Later, Kalton and Kasprzyk (1982), Lee et al. (1994, 1995), Singh and Horn (2000), Singh and Deo (2003), Ahmed et al. (2006)

and Singh (2009) suggested the number of imputation methods under MCAR and the corresponding imputed estimators for dealing with the problem of missing data. Heitzan and Basu (1996) have distinguished the meaning of missing at random (MAR) and missing completely at random (MCAR). In the present work we implicitly assume MCAR approach for imputation of the missing data.

Let $\Omega = \{1, 2, ..., N\}$ be the finite population of N units and a simple random sample without replacement (SRSWOR), s, of size n is drawn from to estimate the population mean $\bar{Y} = N^{-1}\sum_{\Omega} y_i$ of the study variable. Also, let us define $\bar{X} = N^{-1}\sum_{\Omega} x_i$, $K = \rho_{yx}S_y/S_x$, $\rho_{yx} = S_{xy}/S_xS_y$, $C_x = S_x/\bar{X}$, $C_y = S_y/\bar{Y}$, $S_y^2 = (N-1)^{-1}\sum_{\Omega}\left(y_i - \bar{Y}\right)^2$ and $S_x^2 = (N-1)^{-1}\sum_{\Omega}\left(x_i - \bar{X}\right)^2$. Further, let r be the number of responding units out of sampled n units. Let the set of responding units be denoted by A and that of non-responding units be denoted by $\bar{A}$. For every unit, $i \in A$, the value y_i is observed, but for the units $i \in \bar{A}$, the values are missing and imputed values must be derived to complete the y-data. The imputation is carried out with the aid of an auxiliary variable, x, such that x_i, the value of x for unit i, is known and positive for every $i \in s$ i.e. the data $x_s = \{x_i; i \in s\}$ are known. In this chapter, apart from comparision of these imputation procedures, we also propose some Searls type variations of these imputation estimators. Recently, there has been a keen interest in improving the estimators by using Searls technique without even acknowledging the Searls work, for instance, see Koyuncu and Kadilar (2010), Gupta and Shabbir (2008) etc.

2.2 Proposed imputation methods

In the present work, we propose to study the Searls type ratio imputation methods for estimation of population mean $\bar{Y}$ based on the imputation methods suggested by Ahmed et al. (2006), Diana and Perri (2010) from the viewpoint of optimality. In case of conventional imputation procedure the optimal results are provided by regression imputation methods as suggested by Diana and Perri and any generalised ratio type imputation, like Ahmad et al. (2005), Singh and Horn (2000), Singh and Deo (2003) , Singh (2009) etc., under optimal conditions attain the MSE of its respective regression imputation. But the situation changes under STT which are generally better than the conventional imputation procedures but may or may not attain the minimum MSE of STDIM. A Searls type transformation uses the second order terms of sampling errors while calculating the MSE and hence most of time both the characterizing scalars cannot be solved simultaneously and hence an alternative optimal solution is provided. These STRIM are put to test against our previous Searls type difference imputation methods with the following objective in mind:

(i) Whether the Searls type imputation methods are better than the conventional methods of imputation;

(ii) Whether any ratio type imputation methods provides optimal solution under STT;

(iii) Whether under STT, the optimal ratio type imputation can outperform optimal difference type imputation.

(iv) Whether the available auxiliary information has some role in optimality of ratio type imputation methods.

It may be stressed again that while using Searls type transformation, the optimal STRIM may outperform or under perform or perform at par with optimal STDIM. Therefore, the present study is significant in search of optimal ratio type imputation method under STT. As it would be more appropriate to compare the imputation methods which utilize the same auxiliary information.

Searls (1964) proposed a technique for improving the conventional estimators by multipling a tuning constant term α whose optimum value depends on the coefficient of variation, which is a fairly stable quantity. We refer this technique of multiplication by a tuning constant as Searls type transformation (STT). Using STT, we define Searls type mean imputation as

$$y_{.i} = \begin{cases} \alpha y_i \; if \; i \in \; A \\ \alpha \bar{y}_r \; if \; i \in \; \bar{A} \end{cases}$$

which results into the following Searls type estimator

$$T_r = \alpha \bar{y}_r$$

It is important to point out that it is a well accepted "thumb rule" not to modify the respondant data unless it is necessary to maintain the internal consistency of multivariate dataset but here, in this chapter, we do so as the intention is to estimate the population mean in presence of missing data. Therefore, we use Searls technique in ratio type imputation methods for estimation of population mean $\bar{Y}$. We have divided the proposed imputation methods into the following four categories.

Strategy I: When $\bar{X}$ is known and $\bar{x}_n$ is used.

$$y_{1i} = \begin{cases} \alpha_1 y_i & if \quad i \in A \\ \frac{\alpha_1}{n-r}\left[n\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right) - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{2i} = \begin{cases} \alpha_2 y_i & if \quad i \in A \\ \frac{\alpha_2}{n-r}\left[n\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)^{\beta_1} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{5i} = \begin{cases} \alpha_5 y_i & if \quad i \in A \\ \frac{\alpha_5}{n-r}\left[\frac{n\bar{y}_r\bar{X}}{\theta_1\bar{x}_n + (1-\theta_1)\bar{X}} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{12i} = \begin{cases} \alpha_{12} y_i & if \quad i \in A \\ \alpha_{12}\bar{y}_r + \frac{nd_1}{n-r}\left(\bar{X} - \bar{x}_n\right) & if \quad i \in \bar{A} \end{cases}$$

the resultant point estimators are given by

$$T_1 = \alpha_1\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)$$

$$T_2 = \alpha_2\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)^{\beta_1}$$

$$T_5 = \frac{\alpha_5\bar{y}_r\bar{X}}{\theta_1\bar{x}_n + (1-\theta_1)\bar{X}}$$

$$T_{12} = \alpha_{12}\bar{y}_r + d_1\left(\bar{X} - \bar{x}_n\right)$$

Strategy II: When $\bar{X}$ is known and $\bar{x}_r$ is used.

$$y_{3i} = \begin{cases} \alpha_3 y_i & if \quad i \in A \\ \dfrac{\alpha_3}{n-r}\left[n\bar{y}_r\left(\dfrac{\bar{X}}{\bar{x}_r}\right) - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{4i} = \begin{cases} \alpha_4 y_i & if \quad i \in A \\ \dfrac{\alpha_4}{n-r}\left[n\bar{y}_r\left(\dfrac{\bar{X}}{\bar{x}_r}\right)^{\beta_2} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{6i} = \begin{cases} \alpha_6 y_i & if \quad i \in A \\ \dfrac{\alpha_6}{n-r}\left[\dfrac{n\bar{y}_r\bar{X}}{\theta_2\bar{x}_r + (1-\theta_2)\bar{X}} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{13i} = \begin{cases} \alpha_{13} y_i & if \quad i \in A \\ \alpha_{13}\bar{y}_r + \dfrac{nd_2}{n-r}\left(\bar{X} - \bar{x}_r\right) & if \quad i \in \bar{A} \end{cases}$$

the resultant point estimators are given by

$$T_3 = \alpha_3\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)$$

$$T_4 = \alpha_4\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^{\beta_2}$$

$$T_6 = \frac{\alpha_6\bar{y}_r\bar{X}}{\theta_2\bar{x}_r + (1-\theta_2)\bar{X}}$$

$$T_{13} = \alpha_{13}\bar{y}_r + d_2\left(\bar{X} - \bar{x}_r\right)$$

Strategy III: When $\bar{X}$ is unknown and $\bar{x}_r$, $\bar{x}_n$ are used.

$$y_{7i} = \begin{cases} \alpha_7 y_i & if \quad i \in A \\ \dfrac{\alpha_7}{n-r}\left[n\bar{y}_r\left(\dfrac{\bar{x}_n}{\bar{x}_r}\right) - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{8i} = \begin{cases} \alpha_8 y_i & if \quad i \in A \\ \dfrac{\alpha_8}{n-r}\left[n\bar{y}_r\left(\dfrac{\bar{x}_n}{\bar{x}_r}\right)^{\beta_3} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{9i} = \begin{cases} \alpha_9 y_i & if \quad i \in A \\ \alpha_9 \left[\dfrac{\bar{y}_r \left(x_i + \frac{r}{n-r} \bar{x}_r \right)}{\theta_3 \bar{x}_r + (1-\theta_3)\bar{x}_n} - \dfrac{r}{(n-r)} \bar{y}_r \right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{14i} = \begin{cases} \alpha_{14} y_i & if \quad i \in A \\ \alpha_{14} \bar{y}_r + d_3 \left(x_i - \bar{x}_r \right) & if \quad i \in \bar{A} \end{cases}$$

the resultant point estimators are given by

$$T_7 = \alpha_7 \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r} \right)$$

$$T_8 = \alpha_8 \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r} \right)^{\beta_3}$$

$$T_9 = \frac{\alpha_9 \bar{y}_r \bar{x}_n}{\theta_3 \bar{x}_r + (1-\theta_3)\bar{x}_n}$$

$$T_{14} = \alpha_{14} \bar{y}_r + d_3 \left(\bar{x}_n - \bar{x}_r \right)$$

Strategy IV: When $\bar{X}$ is known and $\bar{x}_r$ and $\bar{x}_n$ are used.

$$y_{10i} = \begin{cases} \alpha_{10} y_i & if \quad i \in A \\ \dfrac{\alpha_{10}}{n-r} \left[n\bar{y}_r \left(\dfrac{\bar{x}_n}{\bar{x}_r} \right)^{\beta_4} \left(\dfrac{\bar{X}}{\bar{x}_n} \right)^{\beta_5} - r\bar{y}_r \right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{11i} = \begin{cases} \alpha_{11} y_i & if \quad i \in A \\ \alpha_{11} \left[\dfrac{\bar{y}_r \bar{X} \left(x_i + \dfrac{r}{n-r} \bar{x}_r \right)}{\left\{ \theta_4 \bar{x}_r + (1-\theta_4)\bar{x}_n \right\} \left\{ \theta_5 \bar{x}_n + (1-\theta_5)\bar{X} \right\}} - \dfrac{r}{(n-r)} \bar{y}_r \right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{15i} = \begin{cases} \alpha_{15} y_i & if \quad i \in A \\ \alpha_{15} \bar{y}_r + \dfrac{n}{n-r} k_1 \left(\bar{X} - \bar{x}_n \right) + k_2 \left(x_i - \bar{x}_r \right) & if \quad i \in \bar{A} \end{cases}$$

the resultant point estimators are given by

$$
\begin{aligned}
T_{10} &= \alpha_{10}\bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\beta_4} \left(\frac{\bar{X}}{\bar{x}_n}\right)^{\beta_5} \\
T_{11} &= \alpha_{11}\bar{y}_r \left[\frac{\bar{x}_n}{\theta_4\bar{x}_r + (1-\theta_4)\,\bar{x}_n}\right] \left[\frac{\bar{X}}{\theta_5\bar{x}_n + (1-\theta_5)\,\bar{X}}\right] \\
T_{15} &= \alpha_{15}\bar{y}_r + k_1\left(\bar{X} - \bar{x}_n\right) + k_2\left(\bar{x}_n - \bar{x}_r\right)
\end{aligned}
$$

Theorem 2.2.1. *The bias and minimum MSE of the proposed STR estimators* $(j = 1, 2, ..., 11)$ *are given by*

$$Bias\,(T_j) = \bar{Y}\,(\alpha_j - 1) + \alpha_j Bias\,(t_j) \tag{2.2.1}$$

and

$$\min MSE\,(T_j) = \bar{Y}^2\left(1 - \frac{B_j^2}{A_j}\right) \tag{2.2.2}$$

where $Bias\,(T_j)$ *is the first order bias with parameter* α_j *and* $Bias\,(t_j)$ *is the first order bias of the non-Searls counterpart with* $\alpha_j = 1$.

Proof. The derivations can be very easily followed on the lines of Singh and Deo (2003), Singh (2009), Diana and Perri (2010), Kadilar and Chingi (2008) etc. The MSE of T_2 is given by

$$
\begin{aligned}
MSE\,(T_2) = \bar{Y}^2 \Bigg[1 + \alpha_2^2 \left\{1 + \left(\frac{1}{r} - \frac{1}{N}\right) C_y^2 + 2\beta_1^2\left(\frac{1}{n} - \frac{1}{N}\right) C_x^2 + \beta_1\left(\frac{1}{n} - \frac{1}{N}\right)\left(C_x^2 - 4\rho_{yx}C_yC_x\right)\right\} \\
-2\alpha_2\left\{1 + \frac{\beta_1^2}{2}\left(\frac{1}{n} - \frac{1}{N}\right) C_x^2 + \frac{\beta_1}{2}\left(\frac{1}{n} - \frac{1}{N}\right)\left(C_x^2 - 2\rho_{yx}C_yC_x\right)\right\}\Bigg]
\end{aligned}
$$

which can be expressed as

$$MSE\,(T_2) = \bar{Y}^2\left[1 + \alpha_2^2 A_2 - 2\alpha_2 B_2\right]$$

For optimum value of α_2 differentiating the $MSE\,(T_2)$ with respect to α_2 and equating to zero we get,

$\alpha_{2opt} = \dfrac{B_2}{A_2}$

substituting the optimum value of α_2 in $MSE\left(T_2\right)$ we get minimum MSE

$$\min MSE\left(T_2\right) = \bar{Y}^2\left(1 - \frac{B_2^2}{A_2}\right)$$

The derivations for other estimators $T_j(j = 1, 3, ..., 11)$ can be done on similar lines. In general, we have

$$MSE\left(T_j\right) = \bar{Y}^2\left[1 + \alpha_j^2 A_j - 2\alpha_j B_j\right] \qquad \blacksquare$$

It is important to mention here that simultaneous optimization w.r.t, α_j and β_j of the expression of MSE is not possible and we use optimum value of $\beta_j = \beta_{jopt}$ when $\alpha_j = 1$ and use this within $\alpha_j = \alpha_{jopt}$ to obtain (2.2.2) as used recently by various authors including Singh and Solanki (2013).

The optimum values of scalars involved are tabulated below for ready reference:

$$\alpha_{jopt} = \frac{B_j}{A_j}\ ; (j = 1, 2, ..., 11)$$

$$A_1 = \left\{1 + f_r C_y^2 + f_n\left(3C_x^2 - 4\rho_{yx}C_y C_x\right)\right\}$$

$$B_1 = \left\{1 + f_n\left(C_x^2 - \rho_{yx}C_y C_x\right)\right\}$$

$$A_2 = \left\{1 + f_r C_y^2 + 2\beta_1^2 f_n C_x^2 + \beta_1 f_n\left(C_x^2 - 4\rho_{yx}C_y C_x\right)\right\}$$

$$B_2 = \left\{1 + \frac{\beta_1^2}{2} f_n C_x^2 + \frac{\beta_1}{2} f_n\left(C_x^2 - 2\rho_{yx}C_y C_x\right)\right\}$$

$$A_3 = \left\{1 + f_r C_y^2 + f_r\left(3C_x^2 - 4\rho_{yx}C_y C_x\right)\right\}$$

$$B_3 = \left\{1 + f_r\left(C_x^2 - \rho_{yx}C_y C_x\right)\right\}$$

$$A_4 = \left\{1 + f_r C_y^2 + 2\beta_2^2 f_r C_x^2 + \beta_2 f_r\left(C_x^2 - 4\rho_{yx}C_y C_x\right)\right\}$$

$$B_4 = \left\{1 + \frac{\beta_2^2}{2} f_r C_x^2 + \frac{\beta_2}{2} f_r \left(C_x^2 - 2\rho_{yx} C_y C_x\right)\right\}$$

$$A_5 = \left\{1 + f_r C_y^2 + 3\theta_1^2 f_n C_x^2 - 4\theta_1 f_n \rho_{yx} C_y C_x\right\}$$

$$B_5 = \{1 + \theta_1^2 f_n C_x^2 - \theta_1 f_n \rho_{yx} C_y C_x\}$$

$$A_6 = \left\{1 + f_r C_y^2 + 3\theta_2^2 f_r C_x^2 - 4\theta_2 f_r \rho_{yx} C_y C_x\right\}$$

$$B_6 = \{1 + \theta_2^2 f_r C_x^2 - \theta_2 f_r \rho_{yx} C_y C_x\}$$

$$A_7 = \left\{1 + f_r C_y^2 + f_{rn}\left(3C_x^2 - 4\rho_{yx} C_y C_x\right)\right\}$$

$$B_7 = \{1 + f_{rn}\left(C_x^2 - \rho_{yx} C_y C_x\right)\}$$

$$A_8 = \left\{1 + f_r C_y^2 + 2\beta_3^2 f_{rn} C_x^2 + \beta_3 f_{rn}\left(C_x^2 - 4\rho_{yx} C_y C_x\right)\right\}$$

$$B_8 = \left\{1 + \frac{\beta_3^2}{2} f_{rn} C_x^2 + \frac{\beta_3}{2} f_{rn}\left(C_x^2 - 2\rho_{yx} C_y C_x\right)\right\}$$

$$A_9 = \left\{1 + f_r C_y^2 + \theta_3^2 f_{rn} C_x^2 + 2\theta_3 f_{rn}\left(C_x^2 - 2\rho_{yx} C_y C_x\right)\right\}$$

$$B_9 = \{1 + \theta_3 f_r\left(C_x^2 - \rho_{yx} C_y C_x\right)\}$$

$$A_{10} = \left\{1 + f_r C_y^2 + 2\beta_4^2 f_{rn} C_x^2 + \beta_4 f_{rn}\left(C_x^2 - 4\rho_{yx} C_y C_x\right) + 2\beta_5^2 f_n C_x^2 + \beta_5 f_n\left(C_x^2 - 4\rho_{yx} C_y C_x\right)\right\}$$

$$B_{10} = \left\{1 + \frac{\beta_4^2}{2} f_{rn} C_x^2 + \frac{\beta_4}{2} f_{rn}\left(C_x^2 - 2\rho_{yx} C_y C_x\right) + \frac{\beta_5^2}{2} f_n C_x^2 + \frac{\beta_5}{2} f_n\left(C_x^2 - 2\rho_{yx} C_y C_x\right)\right\}$$

$$A_{11} = \left\{1 + f_r C_y^2 + 3\theta_4^2 f_{rn} C_x^2 - 4\theta_4 f_{rn} \rho_{yx} C_y C_x + 3\theta_5^2 f_n C_x^2 - 4\theta_5 f_n \rho_{yx} C_y C_x\right\}$$

$$B_{11} = \{1 + \theta_4^2 f_{rn} C_x^2 - \theta_4 f_{rn} \rho_{yx} C_y C_x + \theta_5^2 f_n C_x^2 - \theta_5 f_n \rho_{yx} C_y C_x\}$$

where $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 = \rho\frac{C_y}{C_x}$ are used as optimizing value of the constants in this study, which are the optimum values of these scalars when $\alpha_i = 1$ is set in the respective estimators.

Theorem 2.2.2. *The bias and minimum MSE of the STD estimators* $(k = 12, ..., 15)$ *are given by*

$$Bias\left(T_k\right) = \bar{Y}\left(\alpha_k - 1\right) \tag{2.2.3}$$

and

$$\min MSE\left(T_k\right) = \frac{\bar{Y}^2 MSE\left(t_k\right)}{\bar{Y}^2 + MSE\left(t_k\right)} \tag{2.2.4}$$

where $MSE\left(T_k\right)$ *is the first order MSE with parameter* α_k *and* $MSE\left(t_k\right)$ *is the first order MSE of the non-Searls counterpart with* $\alpha_k = 1$.

Proof. The MSE of T_{12} is given by
$MSE\left(T_{12}\right) = \left(\alpha_{12} - 1\right)^2 \bar{Y}^2 + \alpha_{12}^2 \bar{Y}^2 \left(\frac{1}{r} - \frac{1}{N}\right) C_y^2 + d_1^2 \bar{X}^2 \left(\frac{1}{n} - \frac{1}{N}\right) C_x^2 - 2\alpha_{12} d_1 \bar{X}\bar{Y}$
$\left(\frac{1}{n} - \frac{1}{N}\right) \rho_{yx} C_y C_x$
For optimum value of α_{12} and d_1 differentiating partially the $MSE\left(T_{12}\right)$ with respect to α_{12} and d_1, and equating to zero, we get

$\alpha_{12opt} = \dfrac{1}{1 + C_y^2\left\{f_{rn} + f_n\left(1 - \rho_{yx}^2\right)\right\}}$ and $d_{1opt} = K\alpha_{12opt}$

subsituting the optimum value of α_{12} and d_1 in $MSE\left(T_{12}\right)$, we get minimum MSE

$\min MSE\left(T_{12}\right) = \dfrac{\bar{Y}^2 MSE\left(t_{12}\right)}{\bar{Y}^2 + MSE\left(t_{12}\right)}$

where $K = \rho_{yx}\frac{S_y}{S_x}$, $f_n = \left(\frac{1}{n} - \frac{1}{N}\right)$, $f_r = \left(\frac{1}{r} - \frac{1}{N}\right)$ & $f_{rn} = \left(\frac{1}{r} - \frac{1}{n}\right)$

The derivation of other estimators $T_k\left(k = 13, 14, 15\right)$ can be done on similar lines to obtain the from (2.2.4). ∎

It is interesting to note that simultaneous optimization w.r.t. the characterizing scalars α_k $(k = 12, ..., 15)$, d_l $(l = 1, 2, 3)$ and k_m $(m = 1, 2)$ of the expression of MSE is possible for STD estimators but not STR estimators.

The optimum values of scalars involved are tabulated below for ready reference:

$$\alpha_{13opt} = \frac{1}{\left\{1 + f_r C_y^2 \left(1 - \rho_{yx}^2\right)\right\}} \text{ and } d_{2opt} = K\alpha_{13opt}$$

$$\alpha_{14opt} = \frac{1}{1 + C_y^2 \left\{f_n + f_{rn} \left(1 - \rho_{yx}^2\right)\right\}} \text{ and } d_{3opt} = K\alpha_{14opt}$$

$$\alpha_{15opt} = \frac{1}{1 + C_y^2 \left\{f_n \left(1 - \rho_{yx}^2\right) + f_{rn} \left(1 - \rho_{yx}^2\right)\right\}}, \; k_{1opt} = K\alpha_{15opt}$$

and $k_{2opt} = K\alpha_{14opt}$.

Theorem 2.2.3. *The STRIM* $(j = 1, 2, ..., 11)$ *are better than STDIM* $(k = 12, ..., 15)$ *iff*

$$\frac{B_j^2}{A_j} > \alpha_{kopt} \tag{2.2.5}$$

and vice versa. Otherwise both are equally efficient in case of equality in (2.2.5).

Proof. It may be easily observed from (2.2.4) that the MSE of proposed STD estimators $(k = 12, ..., 15)$ are given by

$$\min MSE\left(T_k\right) = \bar{Y}^2 \left(1 - \alpha_{kopt}\right) \tag{2.2.6}$$

Comparing (2.2.6) with (2.2.2), we have the theorem. ∎

The only way ascertain (2.2.5) whether this holds in practice or not is through an empirical study.

2.2.1 Note on practicability

It is evident by observing the optimal values of the tuning constant α's, β's and θ's from Appendix 2 that the optimum values of these tuninig constants depend on the quantities like C_y^2, C_x^2, ρ_{yx} and K. The stability of these quantities has been discussed by authors like

Murthy (1977), Reddy (1978), Singh (2009) etc. among various others. Therefore, some prior information about these quantities may be readily available from some previous study or a pilot survey which may be used to harness the gains in efficiency.

2.3 Numerical study

1. *Population 1*, the relative comparison among the estimators is given using a real data set. The data for the illustration has been taken from the Department of Statistics (Jordan), Healthcare Utilization and expenditure survey, 2000. The population that we like to study contains 8306 household. We consider the variables y and x where y is the expenditure of the household and x is the income of the household. The following values were reported by Ahmed et al. (2006):
   S_y^2=338006, S_x^2=862017, S_{xy}=281892, $\bar{Y}$=253.75, $\bar{X}$=343.316, C_y=2.29116, C_x=2.70436, ρ=0.52223, n=200, r=180.

2. *Population 2*, we use same data reported in Kadilar and Chingi (2008) concerning the level of apple y and the number of apple trees x. Essential population statistics for calculation are N=19, $\bar{Y}$=575, $\bar{X}$=13573.68, S_y=858.36, S_x=12945.38, ρ=0.88. sample sizes are assumed to be n=10 and r=8.

3. *Population 3*, we use same data reported in Diana and Perri (2010) concerning the sale area y (in square meters) and the number of employees x by a market research company. Essential population statistics for calculation are N=2376, $\bar{Y}$=1701.95, $\bar{X}$=40.62, S_y=2195.52, S_x=95.46, ρ=0.90. sample sizes are assumed to be n=300 and r=100.

In the given table, we give the mean square error (MSE) and percentage relative efficiency (PRE) for the estimators under consideration with respect to the estimator $\bar{y}_r$. We have also compared the proposed estimators with the MSE of the estimators suggested by Ahmed et al. (2006) for the same data set.

Table 2.1: MSE and PRE of the estimators

Estimator	MSE	P.R.E. w.r.t. t_r	Proposed Estimator	MSE	P.R.E. w.r.t. t_r	P.R.E. of T_i w.r.t. t_i
t_r	1837.11	100	T_r	1786.155	102.852	102.85
t_{RAT}	1867.229	98.387	T_{RAT}	1806.332	101.704	103.37
Strategy I						
t_1	2101.646	87.413	T_1	1935.923	94.895	108.56
t_2	1387.297	132.423	T_2	1345.223	**136.565**	103.13
t_5	1387.297	132.423	T_5	1358.037	135.277	102.15
t_{12}	1387.301	132.423	T_{12}	1358.037	135.277	102.15
Strategy II						
t_3	2131.764	86.178	T_3	1949.2	94.249	109.37
t_4	1336.084	137.410	T_4	1294.967	**141.865**	103.18
t_6	1336.084	137.410	T_6	1308.926	140.352	102.07
t_{13}	1336.089	137.410	T_{13}	1308.924	140.352	102.08
Strategy III						
t_7	1867.229	98.387	T_7	1806.332	101.704	103.37
t_8	1785.898	102.867	T_8	1735.992	105.824	102.87
t_9	1785.898	102.867	T_9	1734.256	**105.930**	102.98
t_{14}	1785.904	102.867	T_{14}	1737.701	105.720	102.77
Strategy IV						
t_{10}	1336.084	137.410	T_{10}	1294.967	**141.865**	103.18
t_{11}	1336.084	137.410	T_{11}	1308.928	140.352	102.07
t_{15}	1336.084	137.410	T_{15}	1308.924	140.352	102.07

Table 2.2: MSE and PRE of the estimators

Estimator	MSE	P.R.E. w.r.t. t_r	Proposed Estimator	MSE	P.R.E. w.r.t. t_r	P.R.E. of T_i w.r.t. t_i
t_r	53319.79	100	T_r	45915.04	116.126	116.13
t_{RAT}	40126.58	132.878	T_{RAT}	36317.86	146.814	110.49
Strategy I						
t_1	28322.0	188.261	T_1	26810.06	198.879	105.64
t_2	26293.03	202.790	T_2	24832.94	214.713	105.88
t_5	26293.03	202.740	T_5	24356.11	**218.917**	107.95
t_{12}	26293.03	202.790	T_{12}	24356.11	**218.917**	107.95
Strategy II						
t_3	15129.02	352.433	T_3	14983.82	355.898	100.97
t_4	12028.93	443.262	T_4	11906.4	447.824	101.03
t_6	12028.93	443.262	T_6	11606.66	**459.389**	103.64
t_{13}	12028.93	443.262	T_{13}	11606.66	**459.389**	103.64
Strategy III						
t_7	40126.58	132.878	T_7	36317.86	146.814	110.49
t_8	39055.64	136.522	T_8	36982.57	144.175	105.61
t_9	39055.64	136.522	T_9	35641.86	149.598	109.58
t_{14}	39055.64	136.522	T_{14}	34929.53	**152.649**	111.81
Strategy IV						
t_{10}	12028.93	443.262	T_{10}	11906.4	447.824	101.03
t_{11}	12028.93	443.262	T_{11}	11606.66	**459.389**	103.64
t_{15}	12028.93	443.262	T_{15}	11606.66	**459.389**	103.64

Table 2.3: MSE and PRE of the estimators

Estimator	MSE	P.R.E. w.r.t. t_r	Proposed estimator	MSE	P.R.E. w.r.t. t_r	P.R.E. of T_i w.r.t. t_i
t_r	46174.33	100	T_r	45449.83	101.594	101.59
t_{RAT}	47448.16	97.315	T_{RAT}	44078.54	104.754	107.64
Strategy I						
t_1	46730.83	98.809	T_1	45077.988	102.432	103.67
t_2	34802.79	132.674	T_2	34241.99	**134.847**	101.64
t_5	34802.79	132.674	T_5	34389.6	134.268	101.20
t_{12}	34802.79	132.674	T_{12}	34389.6	134.268	101.20
Strategy II						
t_3	48004.66	96.187	T_3	42920.091	107.582	111.85
t_4	8773.123	526.315	T_4	8508.201	**542.703**	103.11
t_6	8773.123	526.315	T_6	8746.632	527.909	100.30
t_{13}	8773.123	526.315	T_{13}	8746.632	527.909	100.30
Strategy III						
t_7	47448.16	97.315	T_7	44078.54	104.754	107.64
t_8	20144.67	229.213	T_8	19763.98	233.628	101.93
t_9	20144.67	229.213	T_9	19407.12	**237.924**	103.80
t_{14}	20144.67	229.213	T_{14}	20005.54	230.807	100.70
Strategy IV						
t_{10}	8773.123	526.315	T_{10}	8508.201	**542.703**	103.11
t_{11}	8773.123	526.315	T_{11}	8746.632	527.909	100.30
t_{15}	8773.123	526.315	T_{15}	8746.632	527.909	100.30

From perusal of above results for Strategy I, a comparision of STRIM T_1, T_2, T_5 and T_{12} with the conventional imputation counterparts t_1, t_2, t_5 and t_{12} show that STRIM T_1, T_2, T_5 and T_{12} are better than t_1, t_2, t_5 and t_{12}. A similar result can be seen for all other strategic situations II and III. Hence, we conclude that all proposed Searls type estimators have higher efficiency in comparision to the conventional imputation procedures. Again on perusal of Strategy I results, a comparision of STRIM T_1, T_2, T_5 and T_{12} within themselves shows that STRIM T_2 is better than other STRIM counterparts T_1 and T_5 in populations 1 and 3 while in population 2 STRIM T_5 is better than other STRIM counterparts T_1 and T_2. In populations 1 and 3, STRIM T_2 outperforms STDIM T_{12}; while in population 2, T_5 attains the MSE of STDIM T_{12}. It is further noteworthy that the condition (2.2.5) is satisfied for T_2 in population 1 and 3 while an equality exists in condition (2.2.5) for T_5 in population 2. Further, T_4 is better in population 1 and 3 in comparision with remaining imputation methods as the condition (2.2.5) is satisfied while T_6 performs at par with T_{13} in population 2 as (2.2.5) is satisfied with an equality. Also, for strategy III, T_9 is the best estimator in populations 1 and 3 while T_{14} is best in population 2. Further, the results of strategy IV are equivalent to that of the corresponding estimators in Strategy II i.e. estimators T_{10}, T_{11} and T_{15} are equivalent to T_4, T_6 and T_{13} respectively for all the populations which is in consonance with the conventional results.

Further, one feature regarding ratio estimation, which is missed most of the time, is that the ratio estimation is feasible only when there is a strong positive linear relation between x and y variables and this is reflected through their high positive correlation. This correlation is strongest when the ratio is calculated by using the same sample size base that i.e. $\bar{y}_r/\bar{x}_r$ is used and not $\bar{y}_r/\bar{x}_n$. It is a popular notion that larger sample size yields by better results and so it is better to use $\bar{x}_n$ instead of $\bar{x}_r$. Infact, this larger sample size based estimator $\bar{x}_n$ weakens this correlation with $\bar{y}_r$, which is retrograde for ratio estimator. Thus, strategy I is not a better strategy as it ignores this important aspect and is quite clearly exihibited by the numerical study, where strategy II is always superior than strategy I. Also, Strategy IV is more complicated in the sense that it requires more computational labour, while the final outcome is same as that Strategy II.

2.4 Conclusion

Ahmed et al. (2006), Singh and Deo (2003) and Singh (2009) proposed various important conventional ratio-type imputation methods (RTIM) which attain the MSE of corresponding regression imputation method (REIM) using same amount of information. But the situation changes if we use STT as the optimality of RTIM might be lost in the process. The optimality of STRIM depends on the condition (2.2.5) of theorem 3, in other words, it may outperform the corresponding STDIM while at other times it might be inferior to STDIM and this infact depends on (2.2.5). The optimal expressions of MSE are stated in (2.2.2) and (2.2.4) for STRIM and STDIM respectively. Further, we see that a condition (2.2.5) is involved in optimality of STRIM in comparision to STDIM which can only be verified numerically. Therefore, on the basis of both theoretical as well as numerical study we make the following observations:

(i) STRIM are always better than the conventional RTIM. This was in consonance with our earlier work in which we have shown that STDIM are better than the conventional REIMs proposed by Diana and Perri (2010).

(ii) It is a well known fact that the conventional optimal RTIM are equally efficient as that of REIM, but under STT (2.2.5) should be satisfied with a equality sign so that the optimal STRIM and optimal STDIM are equally efficient.

(iii) Further, the empirical study on two cited populations clearly demonstrates that optimal STRIM depends on (2.2.5) and, therefore, STRIM may sometimes outperforms the corresponding optimal STDIM using the same amount of information.

(iv) Strategy II and Strategy IV are most efficient but strategy II is preferable in the sense that Strategy IV is substantially more complex, thus requires more computational labour, than Strategy II though their performances are exactly same.

The most important finding of this chapter is that STRIM may sometimes be better than the corresponding STDIM using the same amount of information provided (2.2.5) is satisfied. This finding is different from the conventional theory where the optimal RTIM attains at best the MSE of the corresponding REIM upto the first order of approximation.

Chapter 3

On optimal imputation methods using higher order moment

3.1 Introduction

In sampling most of the time data contains missing values and such missing values vitiates both unbiasedness and efficiency of the sampling strategy. One of the most popular ways to deal with the missing data is through imputation. The problem of missing data was tackled by various authors and different type of imputation procedures were adopted to impute or to fill-in the missing data. The role of auxiliary information related to the study variable is crucial for construction of imputation methods. Many authors like Toutenberg and Srivastava (1998), Toutenberg et al. (2008), Rueda and Gonzalez (2004), Rueda et al. (2005) etc. have proposed various methods of imputation under the problem of missing data using auxiliary information.

Rubin (1976), Kalton et al. (1981) and Kalton and Kasprzyk (1982) have suggested various imputation methods which makes the data structurally complete. In his seminal work on imputation Rubin (1976) introduced three concepts, viz, missing at random (MAR), observed at random (OAR) and parameter distribution (PD). Later, Lee et al. (1994, 1995), Singh and Horn (2000), Singh and Deo (2003) and Singh (2009) suggested the number of imputation methods and the corresponding imputed estimators for dealing

with the problem of missing data. Heitzan and Basu (1996) have distinguished the meaning of missing at random (MAR) from missing completely at random (MCAR). In the present chapter, we assume MCAR approach for imputation of the missing data.
Let $\Omega = \{1, 2, ..., N\}$ be the finite population of N identifiable units and a simple random sample without replacement (SRSWOR), s, of size n is drawn from to estimate the population mean $\bar{Y} = N^{-1}\sum_{i=1}^{N} y_i$ of the study variable. Also, let r be the number of responding units out of sampled n units. Let the responding units set be denoted by A and non-responding units set be denoted by $\bar{A}$. For every unit, $i \in A$, the value y_i is observed, but for the units $i \in \bar{A}$, the values are missing and required to be imputed to make the data structurally complete. The imputation is carried out with the aid of auxiliary variables, x and x_i, the value of x for the unit i, is known and positive for every $i \in s$ i.e. the data $x_s = \{x_i; i \in s\}$ are known.

This chapter is motivated by the quest of finding optimal imputation procedure using the higher order moments of an available auxiliary information as suggested by Mohamed et al. (2017). This chapter is arranged as follows: section 3.2, provides a brief review of some imputation methods alongwith some other important results; in section 3.3, we propose three new improved imputation methods using higher order moments of an auxiliary variable and some basic results are given; section 3.4, deals with a comparative study of the proposed new improved imputation methods using higher order moments of an auxiliary variable in comparision to conventional estimators; in section 3.5, a simulation study has been done to compare their performance and in section 3.6, some concluding remarks are made.

3.2 Method of imputation proposed by Mohamed et al. (2017)

Mohamed et al.(2017) proposed a some new methods of imputation by using higher order moment (variance) of the auxiliary variable under three different strategies and showed that improvement over the mean, ratio and regression methods of imputation.

Since Mohamed et al. (2017) used the higher order moment thus the strategies defined for ratio and regression methods of imputation in chapter 2 are redefined as follows

Strategy I: When $\bar{X}$ and S_x^2 are known and $\bar{x}_n$ and $s^2_{x(n)}$are used.

Strategy II: When $\bar{X}$ and S_x^2 are known and $\bar{x}_r$ and $s^2_{x(r)}$ are used.

Strategy III: When $\bar{X}$ and S_x^2 are unknown and $\bar{x}_n$, $\bar{x}_r$, $s^2_{x(n)}$ and $s^2_{x(r)}$ are used.

Strategy I

$$y_{.ireg1} = \begin{cases} y_i & if \quad i\epsilon A \\ \bar{y}_r + \dfrac{n\hat{\beta}_1}{n-r}\left(\bar{X}-\bar{x}_n\right) + \hat{\beta}_2\left(\dfrac{nS_x^2}{(n-r)} - \dfrac{n\sum\limits_{i\epsilon A}\left(x_i-\bar{x}_n\right)^2}{(n-r)(n-1)}\right) & if \quad i\epsilon \bar{A} \end{cases}$$

Strategy II

$$y_{.ireg2} = \begin{cases} y_i & if \quad i\epsilon A \\ \bar{y}_r + \dfrac{n\hat{\beta}_3}{n-r}\left(\bar{X}-\bar{x}_r\right) + \hat{\beta}_4\left(\dfrac{nS_x^2}{(n-r)} - \dfrac{n\sum\limits_{i\epsilon A}\left(x_i-\bar{x}_r\right)^2}{(n-r)(r-1)}\right) & if \quad i\epsilon \bar{A} \end{cases}$$

Strategy III

$$y_{.ireg3} = \begin{cases} y_i & if \quad i\epsilon A \\ \bar{y}_r + \hat{\beta}_5\left(x_i-\bar{x}_r\right) + \hat{\beta}_6\left(\dfrac{n\left(x_i-\bar{x}_n\right)}{(n-r)} + \dfrac{n\sum\limits_{i\epsilon A}\left(x_i-\bar{x}_n\right)^2}{(n-r)(n-1)} - \dfrac{n\sum\limits_{i\epsilon A}\left(x_i-\bar{x}_r\right)^2}{(n-r)(r-1)}\right) & if \quad i\epsilon \bar{A} \end{cases}$$

The resultant point estimators are given by

$$t_{reg1} = \bar{y}_r + \hat{\beta}_1\left(\bar{X}-\bar{x}_n\right) + \hat{\beta}_2\left(S_x^2 - s^2_{x(n)}\right)$$

$$t_{reg2} = \bar{y}_r + \hat{\beta}_3\left(\bar{X}-\bar{x}_r\right) + \hat{\beta}_4\left(S_x^2 - s^2_{x(r)}\right)$$

$$t_{reg3} = \bar{y}_r + \hat{\beta}_5\left(\bar{x}_n-\bar{x}_r\right) + \hat{\beta}_6\left(s^2_{x(n)} - s^2_{x(r)}\right)$$

respectively.

It may be observed that we have used only regression imputation methods of Mohamed et al. (2017) as both regression and ratio imputation method established for improved efficiency under optimal conditions.

The mean square error (MSE) of the resultant estimators based on the different methods of imputations are given below for ready references and further comparisons

$$\min MSE\left(t_{reg1}\right)=f_r S_y^2-f_n S_y^2\left[\rho^2+\frac{\left(\lambda_{12}-\rho\lambda_{03}\right)^2}{\left(\lambda_{04}-1-\lambda_{03}^2\right)}\right]$$

$$\min MSE\left(t_{reg2}\right)=f_r S_y^2\left[1-\rho^2-\frac{\left(\lambda_{12}-\rho\lambda_{03}\right)^2}{\left(\lambda_{04}-1-\lambda_{03}^2\right)}]\right]$$

$$\min MSE\left(t_{reg3}\right)=f_n S_y^2+f_{rn} S_y^2\left[1-\rho^2-\frac{\left(\lambda_{12}-\rho\lambda_{03}\right)^2}{\left(\lambda_{04}-1-\lambda_{03}^2\right)}]\right]$$

where $f_n=\left(\frac{1}{n}-\frac{1}{N}\right)$, $f_r=\left(\frac{1}{r}-\frac{1}{N}\right)$ & $f_{rn}=\left(\frac{1}{r}-\frac{1}{n}\right)$.

$\lambda_{ab}=\frac{\mu_{ab}}{\mu_{20}^{a/2}\mu_{02}^{b/2}}$, $\mu_{ab}=\frac{1}{N-1}\sum_{i\epsilon\Omega}\left(y_i-\bar{Y}\right)^a\left(x_i-\bar{X}\right)^b$; where $a,b=0,1,2,3,4$;

$C_y^2=\frac{S_y^2}{\bar{Y}^2}=\frac{\mu_{20}}{\bar{Y}^2}$; $C_x^2=\frac{S_x^2}{\bar{X}^2}=\frac{\mu_{02}}{\bar{X}^2}$; $\rho=\frac{S_{yx}}{S_y S_x}=\frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}}$ and S_{yx} have their usual meanings.

3.3 Proposed imputation methods

In this section, we propose some improved imputation methods by using higher order moment namely variance of an auxiliary variable. It has been observed by us that an improvement over regresion (difference) type estimator is possible, as described in the previous chapter. Thus in the present Chapter, we have proposed some new imputation methods for estimation of population mean $\bar{Y}$ as an improvement over Mohamed et al.(2017), using technique proposed in previous chapter.

The Searls technique (1964) is used for improving the efficiency of conventional estimators which further optimizes the result, see Bhushan and Pandey (2016). Therefore, the present study is significant in search of optimal imputation methods using higher order moment. As it would be more appropriate to compare the imputation methods which utilize the same amount auxiliary information, we have proposed three new imputation methods given below under the three different strategies

Strategy I: When $\bar{X}$ and S_x^2 are known and $\bar{x}_n$ and $s_{x(n)}^2$ are used.

$$y_{.ireg1} = \begin{cases} \gamma_1 y_i & if \quad i\epsilon A \\ \gamma_1 \bar{y}_r + \dfrac{nd_1}{n-r}\left(\bar{X}-\bar{x}_n\right) + d_2\left(\dfrac{nS_x^2}{(n-r)} - \dfrac{n\sum\limits_{i\epsilon A}(x_i-\bar{x}_n)^2}{(n-r)(n-1)}\right) & if \quad i\epsilon \bar{A} \end{cases}$$

Strategy II: When $\bar{X}$ and S_x^2 are known and $\bar{x}_r$ and $s_{x(r)}^2$ are used.

$$y_{.ireg2} = \begin{cases} \gamma_2 y_i & if \quad i\epsilon A \\ \gamma_2 \bar{y}_r + \dfrac{nd_3}{n-r}\left(\bar{X}-\bar{x}_r\right) + d_4\left(\dfrac{nS_x^2}{(n-r)} - \dfrac{n\sum\limits_{i\epsilon A}(x_i-\bar{x}_r)^2}{(n-r)(r-1)}\right) & if \quad i\epsilon \bar{A} \end{cases}$$

Strategy III: When $\bar{X}$ and S_x^2 are unknown and $\bar{x}_n$, $\bar{x}_r$, $s_{x(n)}^2$ and $s_{x(r)}^2$ are used.

$$y_{.ireg3} = \begin{cases} \gamma_3 y_i & if \quad i\epsilon A \\ \gamma_3 \bar{y}_r + d_5\left(x_i-\bar{x}_r\right) + d_6\left(\dfrac{n(x_i-\bar{x}_n)}{(n-r)} + \dfrac{n\sum\limits_{i\epsilon A}(x_i-\bar{x}_n)^2}{(n-r)(n-1)} - \dfrac{n\sum\limits_{i\epsilon A}(x_i-\bar{x}_r)^2}{(n-r)(r-1)}\right) & if \quad i\epsilon \bar{A} \end{cases}$$

the resultant point estimators are given by

$$T_{reg_1} = \gamma_1 \bar{y}_r + d_1\left(\bar{X}-\bar{x}_n\right) + d_2\left(S_x^2 - s_{x(n)}^2\right)$$

$$T_{reg_2} = \gamma_2 \bar{y}_r + d_3\left(\bar{X}-\bar{x}_r\right) + d_4\left(S_x^2 - s_{x(r)}^2\right)$$

$$T_{reg_3} = \gamma_3 \bar{y}_r + d_5\left(\bar{x}_n-\bar{x}_r\right) + d_6\left(s_{x(n)}^2 - s_{x(r)}^2\right)$$

Theorem 3.3.1. *The bias and minimum MSE of improved regression (or difference) type estimator T_{regj} is given by*

$$Bias\left(T_{regj}\right) = \bar{Y}\left(\gamma_j - 1\right) \tag{3.3.1}$$

and

$$\min MSE_{\gamma_j}\left(T_{regj}\right) = \frac{\bar{Y}^2 MSE_1\left(t_{regi}\right)}{\bar{Y}^2 + MSE_1\left(t_{regi}\right)} \tag{3.3.2}$$

where $MSE_{\gamma_1}\left(T_{regj}\right)$ is the first order MSE with parameter γ_j and $MSE_1\left(t_{regi}\right)$ is the first order MSE of the conventional estmators with $\gamma_j = 1$.

Proof. The MSE of T_{reg_j} $(j = 1, 2, 3)$ is given by

$MSE\left(T_{reg_1}\right) = \left(\gamma_1 - 1\right)^2 \bar{Y}^2 + \gamma_1^2 f r S_y^2 + d_1^2 f_n S_x^2 + d_2^2 f_n S_x^4\left(\lambda_{04} - 1\right) - 2\gamma_1 d_1 f_n \rho_{yx} S_y S_x - 2\gamma_1 d_2 f_n S_y S_x^2 \lambda_{12} + 2 d_1 d_2 f_n S_x^3 \lambda_{03}$

Optimizing $MSE\left(T_{reg_1}\right)$ with respect to γ_1, d_1 and d_2, we get

$$\min MSE\left(T_{reg_1}\right)=\frac{\bar{Y}^2\left[f_rS_y^2-f_nS_y^2\left\{\rho^2+\frac{\left(\lambda_{12}-\rho\lambda_{03}\right)^2}{\left(\lambda_{04}-1-\lambda_{03}^2\right)}\right\}\right]}{\bar{Y}^2+\left[f_rS_y^2-f_nS_y^2\left\{\rho^2+\frac{\left(\lambda_{12}-\rho\lambda_{03}\right)^2}{\left(\lambda_{04}-1-\lambda_{03}^2\right)}\right\}\right]}=\frac{\bar{Y}^2\left[MSE\left(t_{reg_1}\right)\right]}{\left[\bar{Y}^2+MSE\left(t_{reg_1}\right)\right]}$$

The optimum values of scalars involved are given below:
$\gamma_1=\dfrac{\bar{Y}^2}{\left[\bar{Y}^2+f_rS_y^2-f_nS_y^2\left\{\rho^2+\frac{\left(\lambda_{12}-\rho\lambda_{03}\right)^2}{\left(\lambda_{04}-1-\lambda_{03}^2\right)}\right\}\right]}$, $d_1=\gamma_1\dfrac{S_y\left\{\rho_{yx}\left(\lambda_{04}-1\right)-\lambda_{12}\lambda_{03}\right\}}{S_x\left(\lambda_{04}-1-\lambda_{03}^2\right)}$ and
$d_2=\gamma_1\dfrac{S_y\left(\lambda_{12}-\rho_{yx}\lambda_{03}\right)}{S_x^2\left(\lambda_{04}-1-\lambda_{03}^2\right)}$
Similarly the values of scalars involve in T_2 and T_3.
$\gamma_2=\dfrac{\bar{Y}^2}{\left[\bar{Y}^2+f_rS_y^2\left\{1-\rho^2-\frac{\left(\lambda_{12}-\rho\lambda_{03}\right)^2}{\left(\lambda_{04}-1-\lambda_{03}^2\right)}\right\}\right]}$, $d_3=\gamma_2\dfrac{S_y\left\{\rho_{yx}\left(\lambda_{04}-1\right)-\lambda_{12}\lambda_{03}\right\}}{S_x\left(\lambda_{04}-1-\lambda_{03}^2\right)}$ and
$d_4=\gamma_2\dfrac{S_y\left(\lambda_{12}-\rho_{yx}\lambda_{03}\right)}{S_x^2\left(\lambda_{04}-1-\lambda_{03}^2\right)}$
$\gamma_3=\dfrac{\bar{Y}^2}{\left[\bar{Y}^2+f_rS_y^2-f_{rn}S_y^2\left\{\rho^2+\frac{\left(\lambda_{12}-\rho\lambda_{03}\right)^2}{\left(\lambda_{04}-1-\lambda_{03}^2\right)}\right\}\right]}$, $d_5=\gamma_3\dfrac{S_y\left\{\rho_{yx}\left(\lambda_{04}-1\right)-\lambda_{12}\lambda_{03}\right\}}{S_x\left(\lambda_{04}-1-\lambda_{03}^2\right)}$
and $d_6=\gamma_3\dfrac{S_y\left(\lambda_{12}-\rho_{yx}\lambda_{03}\right)}{S_x^2\left(\lambda_{04}-1-\lambda_{03}^2\right)}$ ■

Corollary 3.3.2. *The proposed imputation method* T_{regj} $(j=1,2,3)$. *has always lesser MSE than the conventional imputation methods* t_{regi} $(i=1,2,3)$, *so that* $minMSE\left(T_j\right)\leq minMSE\left(t_i\right)(i,j=1,2,3)$.

Proof. Trivial by using (3.3.2) ■

Theorem 3.3.3. *The suggested estimators* $T_{reg(j)}$ $(j=1,2,3)$ *are better than estimators proposed by us in chapter 2,* T_k $(k=2,4,8)$ *iff*

$$\gamma_{jopt}>\frac{B_j^2}{A_j} \tag{3.3.3}$$

and vice versa. Otherwise both are equally efficient in case of equality in (3.3.3).

Proof. It may be easily observed from (3.3.2) that the MSE of proposed estimators T_{regj},

$(j = 1, 2, 3)$ are given by

$$\min MSE\left(T_{regj}\right) = \overline{Y}^2\left(1 - \gamma_{jopt}\right) \qquad (3.3.4)$$

Comparing (3.3.4) with (2.2.2), we have the theorem. ■

Theorem 3.3.4. *The suggested estimators $T_{reg(j)}$, $j = 1, 2, 3$ are better than estiamtors $T_i, (i = 12, 13, 14)$ proposed by us in chapter 2.*

Proof. Comparing optimum value of the γ_1, γ_2 and γ_3 involved in suggested estimators with the optimum value of the constants α_{12}, α_{13} and α_{14} respectively involved in estimators suggested by us in chapter 2. We observed that $\gamma_1 > \alpha_{12}$, $\gamma > \alpha_{13}$ and $\gamma_3 > \alpha_{14}$. Therefore, the proposed estimators always perform better than the estimators suggested by us in chapter 2. ■

The only way ascertain (3.3.3) whether this holds in practice or not is through an computational study.

3.4 Computational study using real populations

In order to have a better understanding about the efficiancy of the proposed methods of imputation we have conducted an empirical study on two population (data source from Advance sampling theory with applications by Sarjinder Singh), first the data from Agriculture Statistics (1999) Washington, US and second data from a team of doctors wishes to estimate the average duration of sleep (in minutes) during the night for persons aged 50 years and over in a small village in the United States and compared the proposed imputation methods with $\bar{y}_r$ and the results are reported.

1. All operating banks: Amount (in \$000) of agriculture loans outstanding in different states in 1997.

   Y_i=Amount (in \$000) of real estate farm loans in different states during 1997.

   X_i=Amount (in \$000) of non real estate farm loans in different states during 1997.

$N = 50$, $n = 20$, $r = 8$, $\bar{Y} = 555.43$, $\bar{X} = 878.16$, $S_y^2 = 342021.5$, $S_x^2 = 1176526$, $S_{xy} = 509910.41$, $\lambda_{03} = 1.5936$, $\lambda_{30} = 1.1011$, $\rho_{xy} = .8038$, $\lambda_{21} = 0.9387$, $\lambda_{12} = 1.0982$, $\lambda_{40} = 3.5822$, $\lambda_{04} = 4.5247$, $\lambda_{22} = 2.8411$, $\lambda_{31} = 2.9287$, $\lambda_{13} = 3.2561$.

2. Second population consider a small village having only 30 old persons (age more that 50 years): Approximate duration of sleep (in minutes) and age (in years) of the persons.

   Y_i=Duration of sleep (in minutes).

   X_i=Age of old person (in years).

   $N = 30$, $n = 12$, $r = 4$, $\bar{Y} = 384.2$, $\bar{X} = 67.267$, $S_y^2 = 3582.38$, $S_x^2 = 85.237$, $S_{xy} = -472.607$, $\lambda_{03} = 0.3390$, $\lambda_{30} = -0.0014$, $\rho_{xy} = -0.8552$, $\lambda_{21} = 0.1757$, $\lambda_{12} = -0.3118$, $\lambda_{40} = 2.5788$, $\lambda_{04} = 2.1643$, $\lambda_{22} = 1.9296$, $\lambda_{31} = -2.0996$, $\lambda_{13} = -1.9259$.

Mean square error (MSE) and Percent relative efficiency (PRE) of the estimators with respect to $\bar{y}_r$ is defined by Table: 3.1.

$$PRE = \frac{MSE(\bar{y}_r)}{min.MSE(T_{regi})} \times 100 \tag{3.4.1}$$

Table 3.1: MSE and PRE of the estimators

	Population 1		*Population 2*	
Estimator	*MSE*	*PRE*	*MSE*	*PRE*
$\bar{y}_r$	35912.16	100	775.9935	100
Strategy I				
$\bar{y}_{rat1}$	30681.36	117.0491	1184.906	65.4856
T_2	26498.56	135.5253	643.1707	120.6435
$\bar{y}_{reg1}$	29200.18	122.9864	644.9418	120.3122
T_{12}	26675.32	134.6273	642.1361	120.8379
t_{reg1}	28821.39	124.6028	644.86	120.3274
T_{reg1}	**26358.86**	**136.2436**	**642.0551**	**120.8531**
Strategy II				
$\bar{y}_{rat2}$	17604.10	203.9994	2547.57	30.4581
T_4	11728.98	306.1840	**207.5226**	**373.908**
$\bar{y}_{reg2}$	12419.99	289.1489	208.4440	372.2551
T_{13}	11939.32	300.7897	208.1501	372.7808
t_{reg1}	11094.23	323.7021	208.0898	372.8888
T_{reg2}	**10709.12**	**335.3429**	207.7968	373.4145
Strategy III				
$\bar{y}_{rat3}$	22835.00	157.2685	2138.607	36.2826
T_8	17741.85	202.4155	339.3727	228.6405
$\bar{y}_{reg3}$	19132.06	187.7072	339.4457	228.5913
T_{14}	18014.86	199.348	338.6669	229.117
t_{reg3}	18185.10	197.4818	339.1733	228.775
T_{reg3}	**17172.82**	**209.1226**	**338.3957**	**229.3006**

3.5 Simulation study

Following Singh and Horn (1998), we generated a population of $N = 2000$ units with two variables Y and X whose values are given by:

$$y_i = 2.8 + \left(\sqrt{1-\rho_{xy}^2}\right) y_i^* + \rho_{xy}\frac{S_y}{S_x}x_i^* \tag{3.5.1}$$

and

$$x_i = 2.4 + x_i^* \tag{3.5.2}$$

where $x_i^* \sim G(a_x, b_x)$ and $y_i^* \sim G(a_y, b_y)$ follow independently gamma distributions. In particular, we choose $a_x = 0.085$, $b_x = 1.2$, $a_y = 0.109$ and $b_y = 2.0$ and take the different value of ρ_{xy} (where $\rho_{xy} = 0.6, 0.7, 0.8, 0.9$). The higher value of $\rho_{xy} = 0.6, 0.7, 0.8, 0.9$ successively seems to be a reasonable choice for checking the behaviour of the proposed estmators. From the given population of $N = 2000$ units, using the SRSWOR scheme, we selected $NITRF = 200$ samples, each of size n units. From a given sample s_n of n units, and using SRSWOR scheme, we again selected $NITRS = 200$ samples s_r, each of r units. Thus over $200 \times 200 = 40,000$ iterations, were used to compute the percent relative efficiency of the proposed estimators $T_i, (i = 1, 2, 3)$ over the sample mean and linear regression estimator for $n = 200$ and $r = 160$, 180 as follows:

$$PRE = \frac{\sum\limits_{k=1}^{NITRF}\sum\limits_{j=1}^{NITRS}\left[t_i(j,k)-\bar{Y}\right]^2}{\sum\limits_{k=1}^{NITRF}\sum\limits_{j=1}^{NITRS}\left[T_i(j,k)-\bar{Y}\right]^2}\times 100 \tag{3.5.3}$$

Table 3.2: PRE of the estimators

	$r = 160$			
$Estimators$	$\rho = 0.6$	$\rho = 0.7$	$\rho = 0.8$	$\rho = 0.9$
$\bar{y}_r$	100	100	100	100
Strategy I				
$\bar{y}_{rat1}$	88.36655	105.1631	129.9882	171.2441
T_2	143.3762	164.7264	199.1826	251.5282
$\bar{y}_{reg1}$	144.5892	166.7752	202.7699	270.8894
T_{12}	144.593	166.7788	202.7732	270.892
t_{reg1}	145.0553	167.2627	203.2694	271.3374
T_{reg1}	**145.0623**	**167.2696**	**203.2761**	**271.3434**
Strategy II				
$\bar{y}_{rat2}$	85.98316	107.2747	142.8378	216.011
T_4	162.8979	200.714	274.8872	434.5103
$\bar{y}_{reg2}$	165.0934	205.0189	284.7328	520.9396
T_{13}	165.098	205.0237	284.7379	520.9454
t_{reg2}	165.9197	206.0451	286.158	523.547
T_{reg2}	**165.928**	**206.0539**	**286.1678**	**523.5591**
Strategy III				
$\bar{y}_{rat3}$	95.67446	100.5215	105.9176	112.0678
T_8	108.9233	111.9454	115.6831	119.7238
$\bar{y}_{reg3}$	109.1016	112.184	115.9854	120.7718
T_{14}	109.108	112.1907	115.9924	120.7792
t_{reg3}	109.1696	112.2409	116.0281	120.7958
T_{reg3}	**109.176**	**112.2477**	**116.0352**	**120.8031**

Table 3.3: PRE of the estimators

	$r = 180$			
Strategy I				
$\bar{y}_{rat1}$	100.9456	119.9386	149.534	204.416
T_2	151.9242	172.7367	205.4933	313.4592
$\bar{y}_{reg1}$	151.2385	177.2367	223.5479	330.8936
T_{12}	151.3889	177.3956	223.7196	331.0872
t_{reg1}	151.8809	177.9693	224.4236	332.0121
T_{reg1}	**152.0295**	**178.1263**	**224.5932**	**332.2031**
Strategy II				
$\bar{y}_{rat2}$	99.17297	120.4098	155.5555	228.6035
T_4	159.5551	186.9669	232.4768	414.2442
$\bar{y}_{reg2}$	159.2722	192.6087	258.0258	448.4914
T_{13}	159.4274	192.7758	258.2135	448.7278
t_{reg2}	160.0031	193.4925	259.2098	450.5494
T_{reg2}	**160.1564**	**193.6576**	**259.3952**	**450.7829**
Strategy III				
$\bar{y}_{rat3}$	97.92498	99.98358	102.2982	105.0706
T_8	103.2568	104.6444	106.0087	108.3744
$\bar{y}_{reg3}$	103.3897	104.6319	106.2434	108.4492
T_{14}	103.484	104.7172	106.3157	108.4999
t_{reg3}	103.4014	104.6414	106.2502	108.4526
T_{reg3}	**103.4957**	**104.7266**	**106.3224**	**108.5032**

3.6 Conclusion

1. The proposed imputation methods T_{regi}, $i = 1, 2, 3$ always perform better than the corresponding Mohamed et al (2017) regression (difference) type imputation methods t_{regi}, $i = 1, 2, 3$ as establish by Theorem 3.3.1. The computation result are given in Tables 3.1, 3.2 and 3.3, confirm this fact.

2. The proposed imputation methods $T_{regi}, i = 1, 2, 3$ always perform better than the corresponding regression (difference) type imputation strategy proposed us in Chapter 2, T_{12}, T_{13} and T_{14} can be easily seen on comparing (3.3.4) with (2.2.4) as established by Theorem 3.3.4. The computational result given in Table 3.1, 3.2 and 3.3 also support this fact.

3. The proposed imputation strategy $T_{regi}, i = 1, 2, 3$ out perform the ratio imputation stratrgy T_2, T_4, T_8 proposed by us in Chapter 2, provided the condition in Theorem

3.3.3 is satisfied. The same can be observed from 1, 2 and 3 a counter example in case of population 2 under strategy II described in Table 3.1, which shows t_5 is better than the proposed imputation methods as the inequality 3.3.3 is reversed.

4. Similar imputation strategies, resulting estimators and conclusion can be drawn for population total as well.

Chapter 4

Optimal imputation of the missing data in presence of multi auxiliary information

4.1 Introduction

Missing data is a common problem faced by practitioners in sample survey. Imputation is used for substitution of missing data in the sampling theory to circumvent the problem of missing data. A questionnaire contains many questions that we call items. When item non-response occurs, substantial information about the non-respondent is usually available from other items on the questionnaire. Many imputation methods in literature are used selection of these items as auxiliary variable in assigning values to the i^{th} non-response for item y.

The problem of missing data can be dealt either at the estimation stage using imputation or at the sampling design stage by using a sub-sampling design. The problem of missing data was first tackled by Hansen and Hurwitz (1946) at the sampling design stage and later modified and extended by El-badry et al. (1956), Srinath (1971) etc among others in a mailed questionaire survey. The missing data adversely affects the accuracy of the results. The problem of missing data was tackled at estimation stage by various authors and different type of imputation procedures were adopted to impute (fill-in) the missing data at the estimation stage. Many authors have stressed over the role of auxiliary information for imputation in the presence of missing data.

Rubin (1976) and Kalton et al. (1981) have suggested various imputation methods which make the data structurally complete. Rubin (1976) in his seminal work introduced three important concepts, namely, missing at random (MAR), observed at random (OAR) and parameter distribution (PD). Later, Kalton and Kasprzyk (1982), Lee et al. (1994, 1995), Singh and Horn (2000), Singh and Deo (2003), Ahmed et al. (2006) and Singh (2009) suggested the number of imputation methods and the corresponding imputed estimators for dealing with the problem of missing data. Heitzan and Basu (1996) have distinguished the meaning of missing at random (MAR) and missing completely at random (MCAR). In the present work we implicitly assume MCAR approach for imputation of the missing data. In this chapter, we have divided the chapter into various sections. In section 4.2, we have disscussed the various imputation methods suggested by the various authors. In section 4.3, we have proposed the estimators by using different imputation methods and obtained their biases and MSEs. In section 4.4, we have carried out a computational study by using four different populations and, in the final section, we have disscussed the importance of the proposed work.

4.2 Conventional imputation methods

Let $\Omega = \{1, 2, ..., N\}$ be the finite population of N units and a simple random sample without replacement (SRSWOR), s, of size n is drawn from to estimate the population mean $\bar{Y} = N^{-1} \sum_{i=1}^{N} y_i$ of the study variable. Further, let r be the number of responding units out of sampled n units. Let the set of responding units be denoted by A and that of non-responding units be denoted by $\bar{A}$. For every unit, $i \in A$, the value y_i is observed, but for the units $i \in \bar{A}$, the values are missing and imputed values must be derived to complete the data structure. Let us assume that the imputation is carried out with the aid of multiauxiliary information, $\mathbf{X}$, such that $\mathbf{x_i}$, the value of $\mathbf{x}$ for unit i, is known and positive for every $i \in s$ i.e. the data $\mathbf{X_s} = \{\mathbf{x_i}; i \in s\}$ are known.

In this chapter, we follow the notations of Lee et al. (1994) and the notion for single value imputation. In order to give due consideration to economy of space, the authors and their suggested estimators under different imputation methods are given below
Lee et al. (1994) suggested mean and ratio method of imputation

$$y_{.im} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r & if \quad i \in \bar{A} \end{cases} \tag{4.2.1}$$

$$y_{.ir} = \begin{cases} y_i & if \quad i \in A \\ \widehat{b}x_i & if \quad i \in \bar{A} \end{cases} \tag{4.2.2}$$

and their resultant estimators are given by

$$t_m = \bar{y}_r \tag{4.2.3}$$

$$t_r = \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r} \tag{4.2.4}$$

Singh and Deo (2003) proposed a new method of imputation of missing data by using the power transformation method in presence of multi-auxiliary information as

$$y_{.imult} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r \left[n \prod\limits_{j=1}^{p} \left(\frac{\bar{x}_{n_j}}{\bar{x}_{r_j}} \right)^{\alpha_j} - r \right] \frac{\sum\limits_{j=1}^{p} x_{ij}}{\sum\limits_{i \epsilon \bar{A}} \sum\limits_{j=1}^{p} x_{ij}} & if \quad i \in \bar{A} \end{cases} \tag{4.2.5}$$

where $\prod\limits_{i=1}^{p} x_i = x_1.x_2...x_p$ denote the product of p-terms. Under this method of imputation, the point estimator becomes

$$t_{r_{mult}} = \bar{y}_r \prod_{j=1}^{p} \left(\frac{\bar{x}_{n_j}}{\bar{x}_{r_j}} \right)^{\alpha_j} \tag{4.2.6}$$

This is the generalised Srivastava (1971) type estimator for multi-auxiliary information. One can easily observe that if $\alpha_j = 1 \forall \ j = 1, 2, ..., p$ in equation (4.2.6), then, this method of imputation becomes

$$y_{.ir_{mult}} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r \left[n \prod_{j=1}^{p} \left(\frac{\bar{x}_{n_j}}{\bar{x}_{r_j}} \right) - r \right] \frac{\sum_{j=1}^{p} x_{ij}}{\sum_{i \epsilon \bar{A}} \sum_{j=1}^{p} x_{ij}} & if \quad i \in \bar{A} \end{cases} \tag{4.2.7}$$

and the point estimator becomes

$$t_{mult} = \bar{y}_r \prod_{j=1}^{p} \left(\frac{\bar{x}_{n_j}}{\bar{x}_{r_j}} \right) \tag{4.2.8}$$

This is the popular multivariate ratio-type estimator, which was proposed by Olkin (1958). The multivariate product type estimator can be easily derived by choosing $\alpha_j = -1 \forall \ j = 1, 2, ..., p$.

For the sake of comprehensiveness we have divided the imputation procedures under three categories on the basis of availability of bivariate auxiliary information. Further, we have incorporated the bivariate versions of Singh and Deo (2003) imputation methods, although these imputation methods were not include in Singh and Deo (2003).

Strategy I: When $\bar{X}$ and corresponding estimates $\bar{x}_n$ and $\bar{z}_n$ are used.

Strategy II: When $\bar{X}$, $\bar{x}_r$ and $\bar{z}_r$ are used.

Strategy III: When $\bar{x}_r$, $\bar{x}_n$ and $\bar{z}_r$ are used.

On the similar lines the bivariate version of the Srivastava ratio type imputation methods using bivariate auxiliary information under three different strategies are given by

$$y_{ir_1} = \begin{cases} y_i & if \quad i \in A \\ \frac{1}{n-r} \left[n\bar{y}_r \left(\frac{\bar{X}}{\bar{x}_n} \right)^{\alpha_1} \left(\frac{\bar{Z}}{\bar{z}_n} \right)^{\alpha_2} - r\bar{y}_r \right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{ir_3} = \begin{cases} y_i & if \quad i \in A \\ \frac{1}{n-r} \left[n\bar{y}_r \left(\frac{\bar{X}}{\bar{x}_r} \right)^{\alpha_3} \left(\frac{\bar{Z}}{\bar{z}_r} \right)^{\alpha_4} - r\bar{y}_r \right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{ir_5} = \begin{cases} y_i & if \quad i \in A \\ \frac{1}{n-r} \left[n\bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r} \right)^{\alpha_5} \left(\frac{\bar{z}_n}{\bar{z}_r} \right)^{\alpha_6} - r\bar{y}_r \right] & if \quad i \in \bar{A} \end{cases}$$

which results the ratio type estimators are given by

$$t_{r_1} = \bar{y}_r \left(\frac{\bar{X}}{\bar{x}_n}\right)^{\alpha_1} \left(\frac{\bar{Z}}{\bar{z}_n}\right)^{\alpha_2} \tag{4.2.9}$$

$$t_{r_3} = \bar{y}_r \left(\frac{\bar{X}}{\bar{x}_r}\right)^{\alpha_3} \left(\frac{\bar{Z}}{\bar{z}_r}\right)^{\alpha_4} \tag{4.2.10}$$

$$t_{r_5} = \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\alpha_5} \left(\frac{\bar{z}_n}{\bar{z}_r}\right)^{\alpha_6} \tag{4.2.11}$$

respectively

Singh (2009) proposed a method of imputation of missing data given by

$$y_i = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r \left[\dfrac{(n-r)\bar{x}_n + \theta r(\bar{x}_n - \bar{x}_r)}{\theta \bar{x}_r + (1-\theta)\bar{x}_n}\right] \dfrac{x_i}{\sum_{i \in \bar{A}} x_i} & if \quad i \in \bar{A} \end{cases} \tag{4.2.12}$$

which results in the form of Walsh (1970) estimator is given by

$$t_{wr} = \bar{y}_r \left(\frac{\bar{X}}{\theta \bar{x}_n + (1-\theta)\,\bar{x}_r}\right) \tag{4.2.13}$$

Singh (2009) also proposed an imputation method using multi-auxiliary informations

$$y_{imult} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r \left[\dfrac{(n \prod_{j=1}^{p} \bar{x}_{nj} - r \prod_{j=1}^{p} \{\theta_j \bar{x}_{rj} + (1-\theta)\bar{x}_{nj}\}}{\prod_{j=1}^{p} \{\theta_j \bar{x}_{rj} + (1-\theta)\bar{x}_{nj}\}}\right] \dfrac{\sum_{j=1}^{p} x_{ij}}{\sum_{i \in \bar{A}} \sum_{j=1}^{p} x_{ij}} & if \quad i \in \bar{A} \end{cases} \tag{4.2.14}$$

which results in the form of multivariate Walsh type estimator given by

$$t_{wr_{mult}} = \bar{y}_r \prod_j \left(\frac{\bar{x}_{n_j}}{\theta_j \bar{x}_{n_j} + (1-\theta_j)\,\bar{x}_{r_j}}\right) \tag{4.2.15}$$

Again, for the sake of comprehensiveness, we have included three Walsh's type imputation method under three different strategies on the lines of Singh (2009). The bivariate version of the Walsh's type imputation methods using bivariate auxiliary information under three different strategies are given by

$$y_{ir_2} = \begin{cases} y_i & if \quad i \in A \\ \frac{1}{n-r}\left[n\bar{y}_r\left(\frac{\bar{X}}{\theta_1\bar{x}_n+(1-\theta_1)\bar{X}}\right)\left(\frac{\bar{Z}}{\theta_2\bar{z}_n+(1-\theta_2)\bar{Z}}\right)-r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{ir_4} = \begin{cases} y_i & if \quad i \in A \\ \frac{1}{n-r}\left[n\bar{y}_r\left(\frac{\bar{X}}{\theta_3\bar{x}_r+(1-\theta_3)\bar{X}}\right)\left(\frac{\bar{Z}}{\theta_4\bar{z}_r+(1-\theta_4)\bar{Z}}\right)-r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{ir_6} = \begin{cases} y_i & if \quad i \in A \\ \frac{1}{n-r}\left[n\bar{y}_r\left(\frac{\bar{X}}{\theta_5\bar{x}_n+(1-\theta_5)\bar{x}_r}\right)\left(\frac{\bar{Z}}{\theta_6\bar{z}_n+(1-\theta_6)\bar{z}_r}\right)-r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

which results again Walsh type estimators are given by

$$t_{r_2} = \bar{y}_r\left(\frac{\bar{X}}{\theta_1\bar{x}_n+(1-\theta_1)\bar{X}}\right)\left(\frac{\bar{Z}}{\theta_2\bar{z}_n+(1-\theta_2)\bar{Z}}\right) \tag{4.2.16}$$

$$t_{r_4} = \bar{y}_r\left(\frac{\bar{X}}{\theta_3\bar{x}_r+(1-\theta_3)\bar{X}}\right)\left(\frac{\bar{Z}}{\theta_4\bar{z}_r+(1-\theta_4)\bar{Z}}\right) \tag{4.2.17}$$

$$t_{r_6} = \bar{y}_r\left(\frac{\bar{X}}{\theta_5\bar{x}_n+(1-\theta_5)\bar{x}_r}\right)\left(\frac{\bar{Z}}{\theta_6\bar{z}_n+(1-\theta_6)\bar{z}_r}\right) \tag{4.2.18}$$

respectively.

Diana and Perri (2010) provided an important result which stated that any imputation procedure, using same amount of information, can provide no further improvement than the regression imputation up to the first order of approximation. Therefore, it is imperitive that we include a generalization of Diana and Perri (2010) imputation methods using bivariate auxiliary information under three different strategies given by

$$y_{.i1} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r + \frac{n}{n-r}\left[b_1\left(\bar{X}-\bar{x}_n\right)+b_2\left(\bar{Z}-\bar{z}_n\right)\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{.i2} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r + \frac{n}{n-r}\left[b_3\left(\bar{X}-\bar{x}_r\right)+b_4\left(\bar{Z}-\bar{z}_r\right)\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{.i3} = \begin{cases} y_i & if \quad i \in A \\ \bar{y}_r + \frac{n}{n-r}\left[b_5\left(x_i - \bar{x}_r\right) + b_6\left(z_i - \bar{z}_r\right)\right] & if \quad i \in \bar{A} \end{cases}$$

which result the regression type estimators are given by

$$t_1 = \bar{y}_r + b_1\left(\bar{X} - \bar{x}_n\right) + b_2\left(\bar{Z} - \bar{z}_n\right) \tag{4.2.19}$$

$$t_2 = \bar{y}_r + b_3\left(\bar{X} - \bar{x}_r\right) + b_4\left(\bar{Z} - \bar{z}_r\right) \tag{4.2.20}$$

$$t_3 = \bar{y}_r + b_5\left(\bar{x}_n - \bar{x}_r\right) + b_6\left(\bar{z}_n - \bar{z}_r\right) \tag{4.2.21}$$

respectively

Let $\mathbf{X} = (x_{ij})_{n\times p}, (i = 1, 2, ..., n;\ j = 1, 2, ..., p)$ be the $n \times p$ data matrix of the p- auxiliary variables associated with study variable y. It is assumed that complete information is available on the auxiliary vector $\tilde{\mathbf{x}}$, but responses for the study variable are missing. On the basis of the multi-auxiliary information, imputation strategies are divided into three categories

Strategy I: When $\bar{X}$ and corresponding estimates $\bar{x}_{n_1}, \bar{x}_{n_2}, ..., x_{n_k}$ are used

Strategy II: When $\bar{X}$ and corresponding estimates $\bar{x}_{r_1}, \bar{x}_{r_2}, ..., \bar{x}_{r_k}$ are used.

Strategy III: When $\bar{x}_{n_1}, \bar{x}_{n_2}, ..., x_{n_k}$and $\bar{x}_{r_1}, \bar{x}_{r_2}, ..., \bar{x}_{r_k}$ are used.

The generalised Diana and Perri type imputation methods in presence of multiauxiliary information under three different strategies are given below

$$y_{.i_{1mult}} = \begin{cases} \frac{ny_i}{r} + \sum\limits_j b_j\left(\bar{X} - x_i\right) & if \quad i \in A \\ \sum\limits_j b_j\left(\bar{X} - x_i\right) & if \quad i \in \bar{A} \end{cases} \tag{4.2.22}$$

$$y_{.i_{2mult}} = \begin{cases} \frac{ny_i}{r} - \sum\limits_j b_j \frac{nx_i}{r} & if \quad i \in A \\ \sum\limits_j b_j \frac{n\bar{X}}{n-r} & if \quad i \in \bar{A} \end{cases} \tag{4.2.23}$$

$$y_{.i_{3mult}} = \begin{cases} \frac{ny_i}{r} - \sum\limits_j b_j \frac{nx_i}{r} & if \quad i \in A \\ \sum\limits_j b_j \frac{n\bar{x}_n}{n-r} & if \quad i \in \bar{A} \end{cases} \tag{4.2.24}$$

leading to the following regression type estimators

$$t_{1mult} = \bar{y}_r + \sum_j b_j \left(\bar{X} - \bar{x}_{n_j}\right) \tag{4.2.25}$$

$$t_{2mult} = \bar{y}_r + \sum_j b_j \left(\bar{X} - \bar{x}_{r_j}\right) \tag{4.2.26}$$

$$t_{3mult} = \bar{y}_r + \sum_j b_j \left(\bar{x}_{n_j} - \bar{x}_{r_j}\right) \tag{4.2.27}$$

respectively.

It is noteworthy to mention here that as a regression type imputation methods provides maximum gain in efficiency and any other conventional adoptional can at maximum achive the efficiency of the regression type imputation methods. Therefore, inorder to save space, we have omitted the Srivastava type and Walsh type versions of multivariate ratio type imputation methods under all three strategies. Moreover, these versions under three different strategies and using multiauxiliary information can be easily obtained as special case when $\gamma_i = 1, i = 1, 2, ..., 6$ in section 4.2.
having the MSEs are given by

$$Var\left(t_m\right) = f_r S_y^2 \tag{4.2.28}$$

$$MSE\left(t_r\right) = MSE\left(t_{wr}\right) = f_n S_y^2 + f_{rn}\left(S_y^2 + R^2 S_x^2 - 2RS_{xy}\right) \tag{4.2.29}$$

$$\min .MSE\left(t_{wr}\right) = f_r S_y^2 - f_{rn}\frac{S_{xy}^2}{S_x^2} \tag{4.2.30}$$

$$Var\left(t_{r_{mult}}\right) = Var\left(t_{wr_{mult}}\right) = S_y^2\left\{f_n + f_{rn}\left(1 - R^2_{y,x_1,x_2,\ldots,x_n}\right)\right\} \tag{4.2.31}$$

$$\begin{aligned} \min .MSE\left(t_{r1_{mult}}\right) &= \min .MSE\left(t_{r2_{mult}}\right) = \min .MSE\left(t_{1_{mult}}\right) \\ &= S_y^2\left\{f_{rn} + f_n\left(1 - R^2_{y.x_1,x_2,\ldots,x_n}\right)\right\} \end{aligned} \tag{4.2.32}$$

$$\begin{aligned}\min.MSE\left(t_{r3_{mult}}\right) &= \min.MSE\left(t_{r4_{mult}}\right) = \min.MSE\left(t_{2_{mult}}\right)\\ &= S_y^2\left\{f_r\left(1-R^2_{y.x_1,x_2,\ldots,x_n}\right)\right\} \qquad (4.2.33)\end{aligned}$$

$$\begin{aligned}\min.MSE\left(t_{r5_{mult}}\right) &= \min.MSE\left(t_{r6_{mult}}\right) = \min.MSE\left(t_{3_{mult}}\right)\\ &= S_y^2\left\{f_n + f_{rn}\left(1-R^2_{y.x_1,x_2,\ldots,x_n}\right)\right\} \qquad (4.2.34)\end{aligned}$$

where $f_n = \left(\frac{1}{n}-\frac{1}{N}\right)$, $f_r = \left(\frac{1}{r}-\frac{1}{N}\right)$ & $f_{rn} = \left(\frac{1}{r}-\frac{1}{n}\right)$ and $C_y^2 = S_y^2/\bar{Y}^2$, $C_x^2 = S_x^2/\bar{X}^2$,

$\rho_{yx} = S_{xy}/S_xS_y$, S_y^2, S_x^2 and S_{xy} have their usual meanings.
In this chapter, we propose some Searls type variations of these imputation methods, apart from comparision of these imputation procedures.

4.3 Proposed imputation methods

In this section, we have proposed some new imputation methods in presence of bivariate auxiliary information and in presence of multi-auxiliary information. These propose imputation methods are further subdivided into three sub categories (i) Imputation methods based on Walsh's ideology (ii) Imputation methods based on Srivastava's ideology (iii) Imputation methods based on regression methods. We have invoked Searls's philosophy to improve upon these imputation methods under both bivariate auxiliary information and multi-auxiliary information setup. It is interesting to note that by using the Searls's philosophy in ratio type imputation methods as descibed in category (i) and (ii), the calculation of MSE required the second order term of sampling error. Therefore, the charactersing scalars may not be solved analyticaly and hence, an alternative solution for optimality is provided. These imputation methods are futher categories under three strategies precribed in previous section.

4.3.1 Based on bivariate auxiliary information

The imputation methods under *strategy I* are proposed as follows

$$y_{.i} = \begin{cases} \tau_1 y_i & if \quad i \in A \\ \tau_1 \bar{y}_r + \frac{n}{n-r}\left[B_1\left(\bar{X} - \bar{x}_n\right) + B_2\left(\bar{Z} - \bar{z}_n\right)\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_i = \begin{cases} \upsilon_1 y_i & if \quad i \in A \\ \dfrac{\upsilon_1}{n-r}\left[n\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)^{K_1}\left(\frac{\bar{Z}}{\bar{z}_n}\right)^{K_2} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_i = \begin{cases} \upsilon_2 y_i & if \quad i \in A \\ \frac{\upsilon_2}{n-r}\left[n\bar{y}_r\left(\frac{\bar{X}}{\delta_1\bar{x}_n + (1-\delta_1)\bar{X}}\right)\left(\frac{\bar{Z}}{\delta_2\bar{z}_n + (1-\delta_2)\bar{Z}}\right) - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

and the resultant regression type, Srivastava type and Walsh type estimators are given by

$$\begin{aligned} T_{d_1} &= \tau_1\bar{y}_r + B_1\left(\bar{X} - \bar{x}_n\right) + B_2\left(\bar{Z} - \bar{z}_n\right) \\ T_{r_1} &= \upsilon_1\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)^{K_1}\left(\frac{\bar{Z}}{\bar{z}_n}\right)^{K_2} \\ T_{r_2} &= \upsilon_2\bar{y}_r\left(\frac{\bar{X}}{\delta_1\bar{x}_n + (1-\delta_1)\bar{X}}\right)\left(\frac{\bar{Z}}{\delta_2\bar{z}_n + (1-\delta_2)\bar{Z}}\right) \end{aligned}$$

respectively.

The imputation methods under *strategy II* are proposed as follows

$$y_{.i} = \begin{cases} \tau_2 y_i & if \quad i \in A \\ \tau_2 \bar{y}_r + \frac{n}{n-r}\left[B_3\left(\bar{X} - \bar{x}_r\right) + B_4\left(\bar{Z} - \bar{z}_r\right)\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_i = \begin{cases} \upsilon_3 y_i & if \quad i \in A \\ \dfrac{\upsilon_3}{n-r}\left[n\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^{K_3}\left(\frac{\bar{Z}}{\bar{z}_r}\right)^{K_4} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_i = \begin{cases} \upsilon_4 y_i & if \quad i \in A \\ \frac{\upsilon_4}{n-r}\left[n\bar{y}_r\left(\frac{\bar{X}}{\delta_3\bar{x}_r + (1-\delta_3)\bar{X}}\right)\left(\frac{\bar{Z}}{\delta_4\bar{z}_r + (1-\delta_4)\bar{Z}}\right) - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

and the resultant regression type, Srivastava type and Walsh type estimators are given by

$$T_{d_2} = \tau_2\bar{y}_r + B_3\left(\bar{X}-\bar{x}_r\right) + B_4\left(\bar{Z}-\bar{z}_r\right)$$

$$T_{r_3} = \upsilon_3\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^{K_3}\left(\frac{\bar{Z}}{\bar{z}_r}\right)^{K_4}$$

$$T_{r_4} = \upsilon_4\bar{y}_r\left(\frac{\bar{X}}{\delta_3\bar{x}_r+(1-\delta_3)\bar{X}}\right)\left(\frac{\bar{Z}}{\delta_4\bar{z}_r+(1-\delta_4)\bar{Z}}\right)$$

The imputation methods under *strategy III* are proposed as follows

$$y_{.i} = \begin{cases} \tau_3 y_i & if \quad i\in A \\ \tau_3\bar{y}_r + \frac{n}{n-r}\left[B_5\left(x_i-\bar{x}_r\right)+B_6\left(z_i-\bar{z}_r\right)\right] & if \quad i\in\bar{A} \end{cases}$$

$$y_i = \begin{cases} \upsilon_5 y_i & if \quad i\in A \\ \dfrac{\upsilon_5}{n-r}\left[n\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{K_5}\left(\frac{\bar{z}_n}{\bar{z}_r}\right)^{K_6}-r\bar{y}_r\right] & if \quad i\in\bar{A} \end{cases}$$

$$y_i = \begin{cases} \upsilon_6 y_i & if \quad i\in A \\ \frac{\upsilon_6}{n-r}\left[n\bar{y}_r\left(\frac{\bar{X}}{\delta_5\bar{x}_n+(1-\delta_5)\bar{x}_r}\right)\left(\frac{\bar{Z}}{\delta_6\bar{z}_n+(1-\delta_6)\bar{z}_r}\right)-r\bar{y}_r\right] & if \quad i\in\bar{A} \end{cases}$$

and the resultant regression type, Srivastava type and Walsh type estimators are given by

$$T_{d_3} = \tau_3\bar{y}_r + B_5\left(\bar{x}_n-\bar{x}_r\right) + B_6\left(\bar{z}_n-\bar{z}_r\right)$$

$$T_{r_5} = \upsilon_5\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{K_5}\left(\frac{\bar{z}_n}{\bar{z}_r}\right)^{K_6}$$

$$T_{r_6} = \upsilon_6\bar{y}_r\left(\frac{\bar{X}}{\delta_5\bar{x}_n+(1-\delta_5)\bar{x}_r}\right)\left(\frac{\bar{Z}}{\delta_6\bar{z}_n+(1-\delta_6)\bar{z}_r}\right)$$

Imputation methods in the presence of multi-variate auxiliary information $x_1, x_2, ..., x_k$.

4.3.2 Based on multi-auxiliary information

The imputation methods under *strategy I* are proposed as follows

$$y_{.i_{mult}} = \begin{cases} \tau_1 y_i & if \quad i \in A \\ \tau_1 \bar{y}_r + \frac{n}{n-r}\left[\sum_j B_j\left(\bar{X} - \bar{x}_{n_j}\right)\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{.i_{mult}} = \begin{cases} \upsilon_1 y_i & if \quad i \in A \\ \dfrac{\upsilon_1}{n-r}\left[n\bar{y}_r \prod_j \left(\frac{\bar{X}}{\bar{x}_{n_j}}\right)^{K_j} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{.i_{mult}} = \begin{cases} \upsilon_2 y_i & if \quad i \in A \\ \dfrac{\upsilon_2}{n-r}\left[n\bar{y}_r \prod_j \left(\frac{\bar{X}}{\delta_j \bar{x}_{n_j} + (1-\delta_j)\overline{X}}\right) - r\bar{y}_r\right] & if \quad i \in \overline{A} \end{cases}$$

and the resultant regression type, Srivastava type and Walsh type estimators using multi-auxiliary information are given by

$$\begin{aligned} T_{d1_{mult}} &= \tau_1 \bar{y}_r + \sum_j B_j\left(\bar{X} - \bar{x}_{n_j}\right) \\ T_{r1_{mult}} &= \upsilon_1 \bar{y}_r \prod_j \left(\frac{\overline{X}}{\overline{x}_{n_j}}\right)^{K_j} \\ T_{r2_{mult}} &= \upsilon_2 \bar{y}_r \prod_j \left(\frac{\bar{X}}{\delta_j \bar{x}_{n_j} + (1-\delta_j)\bar{X}}\right) \end{aligned}$$

The imputation methods under *strategy II* are proposed as follows

$$y_{.i_{mult}} = \begin{cases} \tau_2 y_i & if \quad i \in A \\ \tau_2 \bar{y}_r + \frac{n}{n-r}\left[\sum_j B_j\left(\bar{X} - \bar{x}_{r_j}\right)\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{.i_{mult}} = \begin{cases} \upsilon_3 y_i & if \quad i \in A \\ \dfrac{\upsilon_3}{n-r}\left[n\bar{y}_r \prod_j \left(\frac{\bar{X}}{\bar{x}_{r_j}}\right)^{K_j} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{.i_{mult}} = \begin{cases} \upsilon_4 y_i & if \quad i \in A \\ \dfrac{\upsilon_4}{n-r}\left[n\bar{y}_r \prod_j \left(\frac{\bar{X}}{\delta_j \bar{x}_{r_j} + (1-\delta_j)\bar{X}}\right) - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

and the resultant regression type, Srivastava type and Walsh type estimators using multi-auxiliary information are given by

$$T_{d2_{mult}} = \tau_2 \bar{y}_r + \sum_j B_j \left(\bar{X} - \bar{x}_{r_j}\right)$$

$$T_{r3_{mult}} = \upsilon_3 \bar{y}_r \prod_j \left(\frac{\bar{X}}{\bar{x}_{r_j}}\right)^{K_j}$$

$$T_{r4_{mult}} = \upsilon_4 \bar{y}_r \prod_j \left(\frac{\bar{X}}{\delta_j \bar{x}_{r_j} + (1-\delta_j)\bar{X}}\right)$$

The imputation methods under *strategy III* are proposed as follows

$$y_{.i_{mult}} = \begin{cases} \tau_3 y_i & if \quad i \in A \\ \tau_3 \bar{y}_r + \frac{n}{n-r}\left[\sum_j B_j \left(x_i - \bar{x}_{r_j}\right)\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{.i_{mult}} = \begin{cases} \upsilon_5 y_i & if \quad i \in A \\ \dfrac{\upsilon_5}{n-r}\left[n\bar{y}_r \prod_j \left(\frac{\bar{x}_{n_j}}{\bar{x}_{r_j}}\right)^{K_j} - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

$$y_{.i_{mult}} = \begin{cases} \upsilon_6 y_i & if \quad i \in A \\ \dfrac{\upsilon_6}{n-r}\left[n\bar{y}_r \prod_j \left(\frac{\bar{x}_{n_j}}{\delta_j \bar{x}_{n_j} + (1-\delta_j)\bar{x}_{r_j}}\right) - r\bar{y}_r\right] & if \quad i \in \bar{A} \end{cases}$$

and the resultant regression type, Srivastava type and Walsh's type estimators using multi-auxiliary information are given by

$$T_{d3_{mult}} = \tau_3 \bar{y}_r + \sum_j B_j \left(\bar{x}_{n_j} - \bar{x}_{r_j}\right)$$

$$T_{r5_{mult}} = \upsilon_5 \bar{y}_r \prod_j \left(\frac{\bar{x}_{n_j}}{\bar{x}_{r_j}} \right)^{K_j}$$

$$T_{r6_{mult}} = \upsilon_6 \bar{y}_r \prod_j \left(\frac{\bar{x}_{n_j}}{\delta_j \bar{x}_{n_j} + (1-\delta_j)\, \bar{x}_{r_j}} \right)$$

Theorem 4.3.1. *The bias and minimum MSE of the proposed ratio type estimators T_{r_j}, $(j = 1, 2, ..., 6)$ are given by*

$$Bias\left(T_{r_j}\right) = \bar{Y}\left(\upsilon_i - 1\right) + \upsilon_i Bias\left(t_{r_j}\right) \tag{4.3.1}$$

and

$$\min MSE\left(T_{r_j}\right) = \bar{Y}^2 \left(1 - \frac{B_j^2}{A_j}\right) \tag{4.3.2}$$

$$\min MSE\left(T_{i_{mult}}\right) = \bar{Y}^2 \left(1 - \frac{B_j^2}{A_j}\right)$$

where $Bias\left(T_{r_j}\right)$ is the first order bias with parameter υ_j and $Bias\left(t_{r_j}\right)$ is the first order bias of the conventional counterpart with $\upsilon_i = 1$.

Proof. The MSE of T_{r_j}, $(j = 1, 2, ..., 6)$is given by

$$MSE\left(T_{r_1}\right) = \bar{Y}^2 \left[1 + \upsilon_1^2 \left\{1 + \left(\tfrac{1}{r} - \tfrac{1}{N}\right) C_y^2 + \left(\tfrac{1}{n} - \tfrac{1}{N}\right) 2K_1^2 C_x^2 + 2\delta_1^2 C_z^2 + K_1 \left(C_x^2\right.\right.\right.$$

$$\left. -4\rho_{yx} C_y C_x\right) + \delta_1 \left(C_z^2 - 4\rho_{yz} C_y C_z\right) + 4K_1 \delta_1 \rho_{zx} C_z C_x \} - 2\tau_4 \left\{1 + \left(\tfrac{1}{n} - \tfrac{1}{N}\right) \left\{ \ \tfrac{K_1^2}{2} C_x^2\right.\right.$$

$$\left.\left.\left. + \frac{\delta_1^2}{2} C_z^2 + \frac{K_1}{2}\left(C_x^2 - 2\rho_{yx} C_y C_x\right) + \frac{\delta_1}{2}\left(C_z^2 - 2\rho_{yz} C_y C_z\right) + K_1 \delta_1 \rho_{zx} C_z C_x \right\}\right\}\right]$$

which can be expressed as

$$MSE\left(T_{r_1}\right) = \bar{Y}^2 \left[1 + \upsilon_1^2 A_1 - 2\upsilon_1 B_1\right]$$

For optimum value of υ_1 differentiating the $MSE\left(T_{r_1}\right)$ with respect $\upsilon_1$to and equating to zero we get

$$\upsilon_{1opt} = \frac{B_1}{A_1}$$

substituting the optimum value of υ_1 in $MSE\left(T_{r_1}\right)$ we get minimum MSE

$$MSE\left(T_{r_1}\right) = \bar{Y}^2\left(1 - \frac{B_1^2}{A_1}\right)$$

The derivation of other estimators $T_j\,(j = 1, 2, ..., 6)$ can be done on similar lines. In general, we have

$$MSE(T_{r_j}) = \bar{Y}^2\left[1 + \upsilon_j^2 A_j - 2\upsilon_j B_j\right]$$

The optimum values of scalars involved are tabulated below for ready reference:
$\upsilon_{jopt} = \dfrac{B_j}{A_j}$; $(i = 1, 2, ..., 6)$

$$A_1 = 1 + f_r C_y^2 + f_n \left\{ \begin{array}{c} 2K_1^2 C_x^2 + 2K_2^2 C_z^2 + K_1\left(C_x^2 - 4\rho_{yx} C_y C_x\right) + K_2\left(C_z^2 - 4\rho_{yz} C_y C_z\right) \\ +4K_1 K_2 \rho_{zx} C_z C_x \end{array} \right\}$$

$$B_1 = 1 + f_n \left\{ \frac{K_1^2}{2} C_x^2 + \frac{K_2^2}{2} C_z^2 + \frac{K_1}{2}\left(C_x^2 - 2\rho_{yx} C_y C_x\right) + \frac{K_2}{2}\left(C_z^2 - 2\rho_{yz} C_y C_z\right) + K_1 K_2 \rho_{zx} C_z C_x \right\}$$

$$A_2 = 1 + f_r C_y^2 + f_n \left\{3\delta_1^2 C_x^2 + 3\delta_2^2 C_z^2 - 4\delta_1 \rho_{yx} C_y C_x - 4\delta_2 \rho_{yz} C_y C_z + 4\delta_1 \delta_2 \rho_{zx} C_z C_x\right\}$$

$$B_2 = 1 + f_n \left\{\delta_1^2 C_x^2 + \delta_2^2 C_z^2 - \delta_1 \rho_{yx} C_y C_x - \delta_2 \rho_{yz} C_y C_z + \delta_1 \delta_2 \rho_{zx} C_z C_x\right\}$$

$$A_{1mult} = 1 + f_r C_y^2 + f_n \left\{ 2\sum_{j=1}^{p} K_j^2 C_j^2 + \sum_{j=1}^{p} K_j \left(C_{x_j}^2 - 4\rho_{yx_j} C_y C_{x_j}\right) + 4\sum\sum_{i>j}^{p} K_i K_j \rho_{x_i x_j} C_{x_i} C_{x_j} \right\}$$

$$B_{1mult} = 1 + f_n \left\{ \frac{1}{2}\sum_{j=1}^{p} K_j^2 C_{x_j}^2 + \frac{1}{2}\sum_{j=1}^{p} K_j \left(C_{x_j}^2 - 2\rho_{yx_j} C_y C_{x_j}\right) + \sum\sum_{i>j}^{p} K_i K_j \rho_{x_i x_j} C_{x_i} C_{x_j} \right\}$$

$$A_{2mult} = 1 + f_r C_y^2 + f_n \left\{ 3 \sum_{j=1}^{p} \delta_j^2 C_{x_j}^2 - 4 \sum_{j=1}^{p} \delta_j \rho_{yx_j} C_{y_j} C_{x_j} + 4 \sum \sum_{i>j}^{p} \delta_i \delta_j \rho_{x_i x_j} C_{x_i} C_{x_j} \right\}$$

$$B_{2mult} = 1 + f_n \left\{ \sum_{j=1}^{p} \delta_j^2 C_{x_j}^2 - \sum_{j=1}^{p} \delta_j \rho_{yx_j} C_{y_j} C_{x_j} + \sum \sum_{i>j}^{p} \delta_i \delta_j \rho_{x_i x_j} C_{x_i} C_{x_j} \right\}$$

$$A_3 = 1 + f_r \left\{ \begin{array}{c} C_y^2 + 2K_3^2 C_x^2 + 2K_4^2 C_z^2 + K_3 \left(C_x^2 - 4\rho_{yx} C_y C_x\right) + K_4 \left(C_z^2 - 4\rho_{yz} C_y C_z\right) + \\ 4K_3 K_4 \rho_{zx} C_z C_x \end{array} \right\}$$

$$B_3 = 1 + f_r \left\{ \frac{K_3^2}{2} C_x^2 + \frac{K_4^2}{2} C_z^2 + \frac{K_3}{2} \left(C_x^2 - 2\rho_{yx} C_y C_x\right) + \frac{K_4}{2} \left(C_z^2 - 2\rho_{yz} C_y C_z\right) + K_3 K_4 \rho_{zx} C_z C_x \right\}$$

$$A_4 = 1 + f_r \left\{ C_y^2 + 3\delta_3^2 C_x^2 + 3\delta_4^2 C_z^2 - 4\delta_3 \rho_{yx} C_y C_x - 4\delta_4 \rho_{yz} C_y C_z + 4\delta_3 \delta_4 \rho_{zx} C_z C_x \right\}$$

$$B_4 = 1 + f_r \left\{ \delta_3^2 C_x^2 + \delta_4^2 C_z^2 - \delta_3 \rho_{yx} C_y C_x - \delta_4 \rho_{yz} C_y C_z + \delta_3 \delta_4 \rho_{zx} C_z C_x \right\}$$

$$A_{3mult} = 1 + f_r \left\{ C_y^2 + 2 \sum_{j=1}^{p} K_j^2 C_j^2 + \sum_{j=1}^{p} K_j \left(C_{x_j}^2 - 4\rho_{yx_j} C_y C_{x_j} \right) + 4 \sum \sum_{i>j}^{p} K_i K_j \rho_{x_i x_j} C_{x_i} C_{x_j} \right\}$$

$$B_{3mult} = 1 + f_r \left\{ \frac{1}{2} \sum_{j=1}^{p} K_j^2 C_{x_j}^2 + \frac{1}{2} \sum_{j=1}^{p} K_j \left(C_{x_j}^2 - 2\rho_{yx_j} C_y C_{x_j} \right) + \sum \sum_{i>j}^{p} K_i K_j \rho_{x_i x_j} C_{x_i} C_{x_j} \right\}$$

$$A_{4mult} = 1 + f_r \left\{ C_y^2 + 3 \sum_{j=1}^{p} \delta_j^2 C_{x_j}^2 - 4 \sum_{j=1}^{p} \delta_j \rho_{yx_j} C_{y_j} C_{x_j} + 4 \sum \sum_{i>j}^{p} \delta_i \delta_j \rho_{x_i x_j} C_{x_i} C_{x_j} \right\}$$

$$B_{4mult} = 1 + f_r \left\{ \sum_{j=1}^{p} \delta_j^2 C_{x_j}^2 - \sum_{j=1}^{p} \delta_j \rho_{yx_j} C_{y_j} C_{x_j} + \sum \sum_{i>j}^{p} \delta_i \delta_j \rho_{x_i x_j} C_{x_i} C_{x_j} \right\}$$

$$A_5 = 1 + f_r C_y^2 + f_{rn} \left\{ \begin{array}{c} 2K_5^2 C_x^2 + 2K_6^2 C_z^2 + K_5 \left(C_x^2 - 4\rho_{yx} C_y C_x\right) + K_6 \left(C_z^2 - 4\rho_{yz} C_y C_z\right) + \\ 4K_5 K_6 \rho_{zx} C_z C_x \end{array} \right\}$$

$$B_5 = 1 + f_{rn} \left\{ \begin{array}{c} \frac{K_5^2}{2} C_x^2 + \frac{K_6^2}{2} C_z^2 + \frac{K_5}{2} \left(C_x^2 - 2\rho_{yx} C_y C_x\right) + \frac{K_6}{2} \left(C_z^2 - 2\rho_{yz} C_y C_z\right) + \\ K_5 K_6 \rho_{zx} C_z C_x \end{array} \right\}$$

$$A_6 = 1 + f_r C_y^2 + f_{rn}\left\{3\delta_5^2 C_x^2 + 3\delta_6^2 C_z^2 - 4\delta_5 \rho_{yx} C_y C_x - 4\delta_6 \rho_{yz} C_y C_z + 4\delta_5 \delta_6 \rho_{zx} C_z C_x\right\}$$

$$B_6 = 1 + f_{rn}\left\{\delta_5^2 C_x^2 + \delta_6^2 C_z^2 - \delta_5 \rho_{yx} C_y C_x - \delta_6 \rho_{yz} C_y C_z + \delta_5 \delta_6 \rho_{zx} C_z C_x\right\}$$

$$A_{5mult} = 1 + f_r C_y^2 + f_{rn}\left\{2\sum_{j=1}^{p} K_j^2 C_j^2 + \sum_{j=1}^{p} K_j\left(C_{x_j}^2 - 4\rho_{yx_j} C_y C_{x_j}\right) + 4\sum\sum_{i>j}^{p} K_i K_j \rho_{x_i x_j} C_{x_i} C_{x_j}\right\}$$

$$B_{5mult} = 1 + f_{rn}\left\{\frac{1}{2}\sum_{j=1}^{p} K_j^2 C_{x_j}^2 + \frac{1}{2}\sum_{j=1}^{p} K_j\left(C_{x_j}^2 - 2\rho_{yx_j} C_y C_{x_j}\right) + \sum\sum_{i>j}^{p} K_i K_j \rho_{x_i x_j} C_{x_i} C_{x_j}\right\}$$

$$A_{6mult} = 1 + f_r C_y^2 + f_{rn}\left\{3\sum_{j=1}^{p} \delta_j^2 C_{x_j}^2 - 4\sum_{j=1}^{p} \delta_j \rho_{yx_j} C_{y_j} C_{x_j} + 4\sum\sum_{i>j}^{p} \delta_i \delta_j \rho_{x_i x_j} C_{x_i} C_{x_j}\right\}$$

$$B_{6mult} = 1 + f_{rn}\left\{\sum_{j=1}^{p} \delta_j^2 C_{x_j}^2 - \sum_{j=1}^{p} \delta_j \rho_{yx_j} C_{y_j} C_{x_j} + \sum\sum_{i>j}^{p} \delta_i \delta_j \rho_{x_i x_j} C_{x_i} C_{x_j}\right\}$$

■

It is important to mention here that simultaneous optimization w.r.t, υ_i, K_i and δ_i of the expression of MSE is not possible and we use optimum value of $K_i = K_{iopt}$ and $\delta_i = \delta_{iopt}$ when $\upsilon_i = 1$ and use this within $\upsilon_i = \upsilon_{iopt}$ to obtain (4.3.2).

Theorem 4.3.2. *The bias and minimum MSE of the proposed difference type estimators* $(k = 1, 2, 3)$ *are given by*

$$Bias\left(T_{d_k}\right) = \bar{Y}\left(\tau_k - 1\right) \tag{4.3.3}$$

and

$$\min MSE\left(T_{d_k}\right) = \frac{\bar{Y}^2 MSE\left(t_{d_k}\right)}{\bar{Y}^2 + MSE\left(t_{d_k}\right)} \tag{4.3.4}$$

$$minMSE\left(T_{dk_{mult}}\right) = \frac{\bar{Y}^2 MSE\left(t_{dk_{mult}}\right)}{\bar{Y}^2 + MSE\left(t_{dk_{mult}}\right)} \tag{4.3.5}$$

where $MSE\left(T_k\right)$ *is the first order MSE with parameter* τ_k *and* $MSE\left(t_k\right)$ *is the first order MSE of the conventional counterpart with* $\tau_k = 1$.

Proof. The MSE of T_{d_k} and $T_{dk_{mult}}(k = 1, 2, 3)$ is given by

$$\min.MSE\left(T_{d_k}\right)=\frac{\bar{Y}^2MSE\left(t_{d_k}\right)}{\bar{Y}^2+MSE\left(t_{d_k}\right)} \text{ and } \min.MSE\left(T_{dk_{mult}}\right)=\frac{\bar{Y}^2MSE\left(t_{dk_{mult}}\right)}{\bar{Y}^2+MSE\left(t_{dk_{mult}}\right)}$$

$$\tau_1=\frac{\bar{Y}^2}{\bar{Y}^2+S_y^2\left\{\left(\frac{1}{r}-\frac{1}{n}\right)+\left(\frac{1}{n}-\frac{1}{N}\right)\left(1-R_{y.zx}^2\right)\right\}},\ B_1=\tau_1\frac{S_y\left(\rho_{yx}-\rho_{yz}\rho_{zx}\right)}{S_x\left(1-\rho_{zx}^2\right)},$$

$$B_2=\tau_1\frac{S_y\left(\rho_{yz}-\rho_{yx}\rho_{zx}\right)}{S_x\left(1-\rho_{zx}^2\right)}$$

$$\tau_2=\frac{\bar{Y}^2}{\bar{Y}^2+S_y^2\left\{\left(\frac{1}{r}-\frac{1}{N}\right)\left(1-R_{y.zx}^2\right)\right\}},\ B_3=\tau_2\frac{S_y\left(\rho_{yx}-\rho_{yz}\rho_{zx}\right)}{S_x\left(1-\rho_{zx}^2\right)},$$

$$B_4=\tau_2\frac{S_y\left(\rho_{yz}-\rho_{yx}\rho_{zx}\right)}{S_x\left(1-\rho_{zx}^2\right)}$$

$$\tau_3=\frac{\bar{Y}^2}{\bar{Y}^2+S_y^2\left\{\left(\frac{1}{n}-\frac{1}{N}\right)+\left(\frac{1}{r}-\frac{1}{n}\right)\left(1-R_{y.zx}^2\right)\right\}},\ B_5=\tau_3\frac{S_y\left(\rho_{yx}-\rho_{yz}\rho_{zx}\right)}{S_x\left(1-\rho_{zx}^2\right)},$$

$$B_6=\tau_3\frac{S_y\left(\rho_{yz}-\rho_{yx}\rho_{zx}\right)}{S_x\left(1-\rho_{zx}^2\right)}$$

$$\tau_{1mult}=\frac{\bar{Y}^2}{\bar{Y}^2+S_y^2\left\{\left(\frac{1}{r}-\frac{1}{n}\right)+\left(\frac{1}{n}-\frac{1}{N}\right)\left(1-R_{y.x_1,x_2,...,x_p}^2\right)\right\}},\ B_j=\tau_{1mult}\frac{S_y\left(\rho_{yxj}-\rho_{yx_i}\rho_{x_ix_j}\right)}{S_{x_j}\left(1-\rho_{x_ix_j}^2\right)}$$
$(i\neq j=1,2,...,p)$

$$\tau_{2mult}=\frac{\bar{Y}^2}{\bar{Y}^2+S_y^2\left\{\left(\frac{1}{r}-\frac{1}{N}\right)\left(1-R_{y.x_1,x_2,...,x_p}^2\right)\right\}},\ B_j=\tau_{2mult}\frac{S_y\left(\rho_{yxj}-\rho_{yx_i}\rho_{x_ix_j}\right)}{S_{x_j}\left(1-\rho_{x_ix_j}^2\right)}$$
$(i\neq j=1,2,...,p)$

$$\tau_{3mult}=\frac{\bar{Y}^2}{\bar{Y}^2+S_y^2\left\{\left(\frac{1}{n}-\frac{1}{N}\right)+\left(\frac{1}{r}-\frac{1}{n}\right)\left(1-R_{y.x_1,x_2,...,x_p}^2\right)\right\}},\ B_j=\tau_{3mult}\frac{S_y\left(\rho_{yxj}-\rho_{yx_i}\rho_{x_ix_j}\right)}{S_{x_j}\left(1-\rho_{x_ix_j}^2\right)}$$
$(i\neq j=1,2,...,p)$

■

It is interesting to note that simultaneous optimization w.r.t. the characterizing scalars τ_k and B_k of the expression of MSE is possible for proposed difference type estimators but not for proposed ratio type estimators.

Theorem 4.3.3. *The proposed ratio type imputation methods* $(j = 1, 2, ..., 6)$ *are better than proposed difference type imputation methods* $(k = 1, 2, 3)$ *iff*

$$\frac{B_j^2}{A_j} > \tau_{kopt} \tag{4.3.6}$$

and vice versa. Otherwise both are equally efficient in case of equality in (4.3.6).

Proof. It may be easily observed from (4.3.4) that the MSE of proposed difference type estimators $(k = 1, 2, 3)$ are given by

$$\min MSE\left(T_k\right) = \bar{Y}^2\left(1 - \tau_{kopt}\right) \tag{4.3.7}$$

Comparing (4.3.7) with (4.3.2), we have the theorem. ■

The only way ascertain (4.3.6) whether this holds in practice or not is through an empirical study.

4.4 Empirical study

We have conducted an empirical study on four populations described below, along with their key parameters.

1. The relative comparison among the estimators is given using a real data set. Data considered from Singh (1969,p.377) a detailed can be seen Singh (1965).

   y : No. of females employed.

   x : No. of females in service.

   z : No. of educated females.

   N=61,n=20, r=8, $\bar{Y}$=7.46, $\bar{X}$=5.31, $\bar{Z}$=179,S_y^2=28.0818, S_x^2=16.1761,S_z^2=2028.1953,

   ρ_{xy}=0.7737,ρ_{yz}=-0.2070,ρ_{zx}=-0.0033.

2. Data considered from the book “Advanced sampling theory with applications” by Sarjinder Singh (p.1114).

   Apples, commercial crop: season average price in \$ per pound, by states, 1994-1996.

   Descriptive parameters of the population-

   Y=Season average price per pound during 1996

   X=Season average price per pound during 1995

   Z=Season average price per pound during 1994

   N=36, n=18, r=8, S_y^2=0.006458, S_x^2=0.005654,S_z^2=0.004017, $\bar{Y}$=0.2033, $\bar{X}$=0.1856, $\bar{Z}$=0.1708, C_y^2=0.1563, C_x^2=0.1641, C_z^2=0.13757,ρ_{xy}=0.8775,ρ_{yz}=0.8577,ρ_{zx}=0.8788.

3. The data for the illustration has been taken from ICMR, the Department of Paediatrics, BHU, during 1983-1984 have been taken under study. The data describe a group of 95 school children, where the target variable y, is the weught (in kg) of the children while x and z have been considered the two auxiliary chracter for the target variable y. The following values were obtained:

   N=95, n=35, r=10, $\bar{Y}$=115.9526, $\bar{X}$=19.4968, $\bar{Z}$=55.8611, C_y=0.05146, C_x=0.15613, C_z=0.0586, ρ_{xy}=0.713, ρ_{yz}=0.620, ρ_{zx}=0.846.

4. The data describe a group of 95 school children of Varanasi where the target variable Y, is the weight (in kg) of the children while X and Z have been considered the two auxiliary character for the target variable Y:

   N=95, n=35, r=10, $\bar{Y}$=115.9526, $\bar{X}$=19.4968, $\bar{Z}$=51.1726, C_y=0.05146, C_x=0.15613, C_z=0.03006, ρ_{xy}=0.713, ρ_{yz}=0.374,ρ_{zx}=0.328.

   In the given Table, we give the Mean square error (MSE) and percent relative efficiency (PRE) for the estimators under consideration with respect to the estimator $\bar{y}_r$. We have also compared the proposed estimators with the MSE of the estimators suggested by for the same data set.

Table 4.1: MSE and PRE of the estimators

	Population 1		*Population* 2	
Estimator	*MSE*	*PRE*	*MSE*	*PRE*
$t = \bar{y}_r$	3.0498	100	0.000628	100
Strategy I				
$t_1 = t_{r_1} = t_{r_2}$	2.4454	124.7139	0.000484	129.7693
T_{d_1}	2.3425	130.1942	0.000478	131.2884
T_{r_1}	2.3382	**130.4325**	0.000478	**131.3107**
T_{r_2}	2.3425	130.1920	0.000479	131.0503
Strategy II				
$t_2 = t_{r_3} = t_{r_4}$	1.0967	278.0950	0.000124	507.3699
T_{d_2}	1.0755	283.5753	0.000123	508.8890
T_{r_3}	1.0686	**285.4086**	0.000123	**509.7027**
T_{r_4}	1.0755	283.5594	0.000124	507.2209
Strategy III				
$t_3 = t_{r_5} = t_{r_6}$	1.7010	179.2905	0.000268	234.4686
T_{d_3}	1.6506	184.7707	0.000266	235.9877
T_{r_5}	1.6436	**185.5538**	0.000266	**236.2089**
T_{r_6}	1.6506	184.7637	0.000267	235.1316

Table 4.2: MSE and PRE of the estimators

	Population 3		*Population* 4	
Estimator	*MSE*	*PRE*	*MSE*	*PRE*
$t = \bar{y}_r$	3.1856	100	3.1856	100
Strategy I				
$t_1 = t_{r_1} = t_{r_2}$	2.8583	111.4490	2.8448	111.9779
T_{d_1}	2.8577	111.4727	2.8442	112.0016
T_{r_1}	2.8575	**111.4818**	2.8440	**112.0107**
T_{r_2}	2.8577	111.4723	2.8442	112.0012
Strategy II				
$t_2 = t_{r_3} = t_{r_4}$	1.5629	203.8163	1.4960	212.9356
T_{d_2}	1.5628	203.8400	1.4958	212.9593
T_{r_3}	1.5617	**203.9764**	1.4948	**213.1046**
T_{r_4}	1.5628	203.8369	1.4959	212.9560
Strategy III				
$t_3 = t_{r_5} = t_{r_6}$	1.8902	168.5299	1.8368	173.4329
T_{d_3}	1.8899	168.5536	1.8365	173.4566
T_{r_5}	1.8891	**168.6301**	1.8357	**173.5360**
T_{r_6}	1.8900	168.5516	1.8365	173.4544

The table 4.1 clearly exhibit superiority of Srivastava type Searls imputation methods T_{r_1}, T_{r_3}, T_{r_5} over both Walsh type Searls imputation methods T_{r_2}, T_{r_4}, T_{r_6} as well as over regression type Searls imputation methods. Also, the regression type Searls imputation methods T_{d_1}, T_{d_2} and T_{d_3} are found to be better than the Walsh type Searls imputation methods T_{r_2}, T_{r_4} and T_{r_6}. The proposed imputation methods provide a considerable gain over the conventional imputation methods in all three strategy i.e. T_{d_1}, T_{r_1} and T_{r_2} are better than t_1, t_{r_1} and t_{r_2} under strategy I. Similarly, T_{d_2}, T_{r_3} and T_{r_4} are better than t_2, t_{r_3} and t_{r_4} under strategy II and T_{d_3}, T_{r_5} and T_{r_6} are better than t_3, t_{r_5} and t_{r_6} under strategy III.

Also, similar conclusions can be drawn from tables 4.2.

4.5 Conclusion

The proposed regression type imputation methods T_{d_1}, T_{d_2} and T_{d_3} always perform better than the corresponding regression (difference) type imputation methods t_1, t_2 and t_3 as provided by Theorem 4.3.2. This fact is also reflected in the computational results in Table 4.1 and Table 4.2. This is an important result and contradicts the conclusion of Diana and Perri (2010) that no further improvement upon regression imputation methods is possible.

The proposed ratio type imputation methods T_{r_1},...,T_{r_6} generally perform better than the conventional regression imputation methods. Moreover, the Srivastava type imputation methods T_{r_1}, T_{r_3} and T_{r_5} always perform better than regression type imputation methods T_{d_1}, T_{d_2} and T_{d_3} respectively, which is in consonance with Theorem 4.3.3. This result has important significance due to the fact that under conventional setup no ratio type imputation methods can ever surpass the efficiency of the regression type imputation methods but under Searls approch this benchmark can be reversed as demonstrated theoretically as well as empirically.

All the important proposed imputation methods including Ahmed et al. (2006), Singh and Deo (2003), Singh (2009) and Diana and Perri (2010) are special case of the proposed imputation methods for suitable choice of characterising scalar. Further, the proposed methods of imputation provide better results in comparision to all these earlier work. This fact is clearly demonstrated by computational study.

Chapter 5

On efficient estimation of population mean under non-response

5.1 Introduction

We know that the problem of non-response was first studied by Hansen and Hurvitz (1946) related to mailed questionnaire. Non-response adversely affects the estimate of the population characteristics in survey sampling. Many authors dealt the non-response and suggested different type of method or estimators to estimate the population characteristics under non-response.

Non-response not only adversely affects the estimates but also leads to biased estimates with inflated variances. Cochran (1977) and Rao (1983) suggested the use of the ratio method of estimation for the population mean $\bar{Y}$ of the study variate y with sub-sampling from non-respondents. Khare and Srivastava (1993) suggested an estimation procedure of population mean using auxiliary character in presence of non-response, Khare and Srivastava (1995) proposed to study the conventional and alternative two phase sampling ratio, product and regression estimators in presence of non-response. Khare and Srivastava (1997) proposed transformed ratio type estimators for the population mean in the presence of non-response. Okafor and Lee (2000) proposed a double sampling scheme for ratio and regression estimation with sub sampling the non-respondent are also dealing with non-response problem. Some other authors considered the problem of non-response in estimation of population parameter i.e. population mean, population total, population ratio and product, and population variance etc.

Under the problem of non-response, some authors suggested that by using the auxiliary information, we can improve the efficiency of the estimators, see Khare and Srivastava (1993). Singh and Kumar (2008) proposed a general class of estimators of population mean in survey sampling using auxiliary information with sub sampling the non-respondent. Also, Singh and Kumar and Kozak (2010) suggested a number of estimators for estimating population mean under non-response. An important finding of all these papers was that the difference or the corresponding regression type estimators were found to be best in terms of MSE and any ratio type estimator can at best attains the MSE of regression (difference) estimator. In this chapter, we have proposed some improvement over these best estimators proposed by various authors in their earlier works.

5.2 Conventional estimators under non-respone

An unbiased estimator for the population mean $\bar{Y}$ is

$$\bar{y}^* = w_1\bar{y}_1 + w_2\bar{y}_{2,} \tag{5.2.1}$$

where $w_1 = n_1/n$ and $w_2 = n_2/n$.

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

$$Var\left(\bar{y}^*\right) = PS_y^2 + QS_{y_2}^2 \tag{5.2.2}$$

where $P = \left(\frac{1}{n} - \frac{1}{N}\right)$ and $Q = \frac{W_2(k-1)}{n}$.

Let $x_i(i = 1, 2, ..., N)$ denote an auxiliary variate correlated with study variate $y_i(i = 1, 2, ..., N)$. The population variances of x and y are denoted by S_x^2 and S_y^2, and the population covariance by S_{xy}. The population correlation coefficient is $\rho = S_{xy}/S_x^2\ S_y^2$.

Cochran (1977) and Rao (1983) suggested ratio and regression type estimators under non-response.

In this section, we have divided the conventional estimators under seven strategies using auxiliary information for single phase and two phase sampling.

Strategy I: When $\bar{y}^*$, $\bar{x}^*$ and $\bar{X}$ are used.

Strategy II: When $\bar{y}^*$, $\bar{x}$ and $\bar{X}$ are used.

Strategy III: When $\bar{y}^*$, $\bar{x}^*$ and $\bar{x}$ are used.

Strategy IV: When $\bar{y}^*$, $\bar{x}$, $\bar{x}^*$ and $\bar{X}$ are used.

Strategy V: When $\bar{y}^*$, $\bar{x}^*$ and $\bar{x}'$ are used.

Strategy VI: When $\bar{y}^*$, $\bar{x}$ and $\bar{x}'$ are used.

Strategy VII: When $\bar{y}^*$, $\bar{x}^*$, $\bar{x}$ and $\bar{x}'$ are used.

Strategy I

The non-response occurs on both the study variable y and auxiliary variable x, and the population mean $\bar{X}$ of the auxiliary variable is known. The ratio and regression type estimators are

$$t_{R_1} = \bar{y}^* \left(\frac{\bar{X}}{\bar{x}^*}\right)^{\beta_1} \tag{5.2.3}$$

$$t_{lr_1} = \bar{y}^* + b^* \left(\bar{X} - \bar{x}^*\right) \tag{5.2.4}$$

where $b^* = s^*_{xy}/s^{*^2}_x$ and the estimates s^*_{xy} and $s^{*^2}_x$ are based on the available data under the given sampling design.

To the first degree of approximation, the variances of estimators t_{R_1} and t_{lr_1} are given by

$$\min .Var\left(t_{R_1}\right) = PS^2_y + QS^2_{y_2} - \left\{\left(PS_{xy} + QS_{xy_2}\right)^2 / \left(PS^2_y + QS^2_{y_2}\right)\right\} \tag{5.2.5}$$

$$Var\left(t_{lr_1}\right) = PS^2_y\left(1-\rho^2\right) + Q\left(S^2_{y_2} + \beta^2 S^2_{x_2} - 2\beta S_{xy_2}\right) \tag{5.2.6}$$

where $R = \bar{Y}/\bar{X}$, $\beta = S_{xy}/S^2_x$ and $\beta_{1opt} = \left(PS_{xy} + QS_{xy_2}\right)/\left(PS^2_y + QS^2_{y_2}\right)$.

Strategy II

The non-response occurs on the study variable y, and information on the auxiliary variable x is available from all the sample units along the population mean $\bar{X}$ of the auxiliary variable is known. The ratio and regression type estimators are

$$t_{R_2} = \bar{y}^* \left(\frac{\bar{X}}{\bar{x}} \right)^{\beta_2} \tag{5.2.7}$$

$$t_{lr_2} = \bar{y}^* + b\left(\bar{X} - \bar{x}\right) \tag{5.2.8}$$

where $b = s^*_{xy}/s^2_x$

To the first degree of approximation, the variances of the estimators t_{R_2} and t_{lr_2} are given by

$$\min . Var\left(t_{R_2}\right) = Var\left(t_{lr_2}\right) = PS^2_y\left(1 - \rho^2\right) + QS^2_{y_2} \tag{5.2.9}$$

where $\beta_2 = S_{xy_2} \left/ S^2_{x_2}\right.$ and $\beta_{2opt} = S_{xy} \left/ S^2_x\right.$.

Strategy III

The non-response occurs on the study variable y, and the information on the auxiliary variable x is obtained from all the sample units, but the population mean $\bar{X}$ of the auxiliary variable is not known. The ratio and regression type estimators are

$$t_{R_3} = \bar{y}^* \left(\frac{\bar{x}}{\bar{x}^*} \right)^{\beta_3} \tag{5.2.10}$$

$$t_{lr_3} = \bar{y}^* + b_{(2r)}\left(\bar{x} - \bar{x}^*\right) \tag{5.2.11}$$

where $b_{(2r)} = s_{xy_{(2r)}}/s^2_{x_{(2r)}}$

To the first degree of approximation, the variances of the estimators t_{R_3} and t_3 are given by

$$\min . Var\left(t_{R_3}\right) = Var\left(t_{lr_3}\right) = PS^2_y + QS^2_{y_2}\left(1 - \rho^2_2\right) \tag{5.2.12}$$

where $\beta_{3opt} = S_{xy_2} \left/ S^2_{x_2}\right.$.

Strategy IV

Singh and Kumar (2010) suggested the estimators are given below

$$t_{R_4} = \bar{y}^* \left(\frac{\bar{X}}{\bar{x}^*} \right)^{\beta_4} \left(\frac{\bar{X}}{\bar{x}} \right)^{\beta_5} \tag{5.2.13}$$

$$t_{lr_4} = \bar{y}^* + d_1 \left(\bar{x} - \bar{x}^*\right) + d_2 \left(\bar{X} - \bar{x}\right) \tag{5.2.14}$$

To the first degree of approximation, the variances of the estimators t_{R_4} and t_{lr_4} are given by

$$\min .Var\left(t_{lr_4}\right) = \min .Var\left(t_{R_4}\right) = PS_y^2\left(1-\rho^2\right) + QS_{y_2}^2\left(1-\rho_2^2\right) \tag{5.2.15}$$

where optimum values of β_4, β_5, d_1 and d_2 are given by $\beta_{4opt} = \beta_{(2)}/R$, $\beta_{5opt} = \beta - \beta_{(2)}/R$, $d_{1opt} = \beta_{(2)}$ and $d_{2opt} = \beta$.

Strategy V

Similarly, Okafor and Lee (2000) proposed a double sampling scheme for ratio estimation with sub sampling the non-respondent are also dealing with non-response problem.

$$t_{R_5} = \bar{y}^* \left(\frac{\bar{x}^{'}}{\bar{x}^*}\right)^{\beta_6} \tag{5.2.16}$$

$$t_{lr_5} = \bar{y}^* + b^* \left(\bar{x}^{'} - \bar{x}^*\right) \tag{5.2.17}$$

To the first degree of approximation, the variances of the estimators t_{R_5} and t_5 are given by

$$\min .Var\left(t_{R_5}\right) = Var\left(t_{lr_5}\right) = TS_y^2 + SS_y^2\left(1-\rho^2\right) + Q\left\{S_{y_2}^2 + \beta S_{x_2}^2\left(\beta - 2\beta_{(2)}\right)\right\} \tag{5.2.18}$$

where $\beta_{(6)opt} = \left\{(P-T)\,S_{xy} + QS_{xy_2}\right\}/PS_x^2 + QS_{x_2}^2$, $S = \left(\frac{1}{n} - \frac{1}{n'}\right)$ and $T = \left(\frac{1}{n'} - \frac{1}{N}\right)$.

Strategy VI

$$t_{R_6} = \bar{y}^* \left(\frac{\bar{x}^{'}}{\bar{x}}\right)^{\beta_7} \tag{5.2.19}$$

$$t_{lr_6} = \bar{y}^* + b^{**} \left(\bar{x}^{'} - \bar{x}\right) \tag{5.2.20}$$

To the first degree of approximation, the variances of the estimators t_{R_6} and t_{lr_6} are given by

$$\min .Var\left(t_{R_6}\right) = Var\left(t_{lr_6}\right) = TS_y^2 + SS_y^2\left(1-\rho^2\right) + QS_{y_2}^2 \tag{5.2.21}$$

where $\beta_{(7)opt} = S_{xy}/S_x^2$.

Strategy VII

Singh and Kumar (2008) proposed estimators for the population mean $\bar{Y}$ by using double sampling scheme under non-response.

$$t_{R_7} = \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}^*}\right)^{\beta_8} \left(\frac{\bar{x}'}{\bar{x}}\right)^{\beta_9} \tag{5.2.22}$$

$$t_{lr_7} = \bar{y}^* + d_3\left(\bar{x} - \bar{x}^*\right) + d_4\left(\bar{x}' - \bar{x}\right) \tag{5.2.23}$$

To the first degree of approximation, the variances of the above estimators are given by

$$\min .Var\left(t_{R_7}\right) = Var\left(t_{lr_7}\right) = TS_y^2 + SS_y^2\left(1-\rho^2\right) + QS_{y_2}^2\left(1-\rho_2^2\right) \tag{5.2.24}$$

where optimum values of β_8, β_9, d_3 and d_4 are given by $\beta_{8opt} = \beta_{(2)}/R$, $\beta_{9opt} = \beta/R$, $d_{3opt} = \beta_{(2)}$ and $d_{4opt} = \beta$.

In this chapter, we propose some new ratio type estimators. It has been observed by us that recently there has been a keen interest in improving the estimators by using Searls technique. Here, we have proposed a number of new ratio type estimators motivated by Cochran (1977), Rao (1983, 1986, 1987), Khare and Srivastava (1995, 1997), Okafor and Lee (2000) and Singh and Kumar (2008, 2010) under the one phase and two phase sampling by using seven different strategy under non-response.

5.3 Proposed estimators

In this section, we propose the ratio and regression (difference) estimators by using Searls type transformation (STT) under different strategies. The Searls type transformation can be defined as

$$T = \alpha\bar{y}^* \tag{5.3.1}$$

where α is a suitable chosen constant.

The mean square error of the above estimators is given by

$$\min.MSE\left(T\right)=\frac{\bar{Y}^{2}MSE\left(t\right)}{\bar{Y}^{2}+MSE\left(t\right)} \tag{5.3.2}$$

In above equation (5.3.2) shown that the STD estimators are better than conventional estimators. The estimator parallel to strategies given in the previous section are proposed using the STT as follows.

Strategy I

We propose the ratio and regression (difference) type estimation of population mean given by

$$T_{R_1}=\alpha_1\bar{y}^{*}\left(\frac{\bar{X}}{\bar{x}^{*}}\right)^{\beta_1} \tag{5.3.3}$$

$$T_{lr_1}=\alpha_1\bar{y}^{*}+\beta_1\left(\bar{X}-\bar{x}^{*}\right) \tag{5.3.4}$$

Strategy II

We propose the ratio and regression (difference) type estimation of population mean given by

$$T_{R_2}=\alpha_2\bar{y}^{*}\left(\frac{\bar{X}}{\bar{x}}\right)^{\beta_2} \tag{5.3.5}$$

$$T_{lr_2}=\alpha_2\bar{y}^{*}+\beta_2\left(\bar{X}-\bar{x}\right) \tag{5.3.6}$$

Strategy III

We propose the ratio and regression (difference) type estimation of population mean given by

$$T_{R_3}=\alpha_3\bar{y}^{*}\left(\frac{\bar{x}}{\bar{x}^{*}}\right)^{\beta_3} \tag{5.3.7}$$

$$T_{lr_3}=\alpha_3\bar{y}^{*}+\beta_3\left(\bar{x}^{*}-\bar{x}\right) \tag{5.3.8}$$

Similarly another estimator obtained by chaining

Strategy IV

We propose the ratio and regression (difference) type estimation of population mean given by

$$T_{R_4} = \alpha_4 \bar{y}^* \left(\frac{\bar{X}}{\bar{x}^*}\right)^{\beta_4} \left(\frac{\bar{X}}{\bar{x}}\right)^{\beta_5} \tag{5.3.9}$$

$$T_{lr_4} = \alpha_4 \bar{y}^* + \beta_4 \left(\bar{x}^* - \bar{x}\right) + \gamma_4 \left(\bar{X} - \bar{x}\right) \tag{5.3.10}$$

It is noteworthy to mention that the estimators given by (5.3.9) and (5.3.9) are obtain by chaining (5.3.3), (5.3.5) and (5.3.4), (5.3.6) respectively. Also, under the two phase sampling scheme, we propose the following Searls type regression (difference) estimators under the following strategies.

Strategy V

We propose the ratio and regression (difference) type estimation of population mean given by

$$T_{R_5} = \alpha_5 \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}^*}\right)^{\beta_6} \tag{5.3.11}$$

$$T_{lr_5} = \alpha_5 \bar{y}^* + \beta_5 \left(\bar{x}' - \bar{x}^*\right) \tag{5.3.12}$$

Strategy VI

We propose the ratio and regression (difference) type estimation of population meangiven by

$$T_{R_6} = \alpha_6 \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}}\right)^{\beta_7} \tag{5.3.13}$$

$$T_{lr_6} = \alpha_6 \bar{y}^* + \beta_6 \left(\bar{x}' - \bar{x}\right) \tag{5.3.14}$$

Strategy VII

We propose the ratio and regression (difference) type estimation of population meangiven by

$$T_{R_7} = \alpha_7 \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}^*}\right)^{\beta_8} \left(\frac{\bar{x}'}{\bar{x}}\right)^{\beta_9} \tag{5.3.15}$$

$$T_{lr_7} = \alpha_7 \bar{y}^* + \beta_7 \left(\bar{x} - \bar{x}^*\right) + \gamma_7 \left(\bar{x}' - \bar{x}\right) \tag{5.3.16}$$

where α_i, β_i and γ_j are suitably chosen constant. It is important to note that all these estimators T_i are generalizations over t_i and we get t_i if we put $\alpha_i = 1$ in $T_i\,(i = 1, 2, ...7)$. In this study, we have focussed our attention over improvement in conventional difference and regression type estimators, which are the BLUEs.

Theorem 5.3.1. *The bias and minimum MSE of the STD estimator T_i $(i = 1, 2, ...7)$ is given by*

$$Bias\left(T_i\right) = \overline{Y}\left(\alpha_i - 1\right) \tag{5.3.17}$$

and

$$\min MSE_{\alpha_i}\left(T_i\right) = \frac{\overline{Y}^2 MSE\left(t_i\right)}{\overline{Y}^2 + MSE\left(t_i\right)} \tag{5.3.18}$$

where $MSE_{\alpha_i}\left(T_i\right)$ is the first order MSE with parameter α_i and $MSE\left(t_i\right)$ is the first order MSE of the non-Searls counterpart. i.e. $\alpha_i = 1$.

Proof. $MSE\left(T_1\right) = \left(\alpha_1 - 1\right)^2 \bar{Y}^2 + \alpha_1^2\left(AS_y^2 + BS_{y_2}^2\right) + \beta_1^2\left(AS_x^2 + BS_{x_2}^2\right) - 2\alpha\beta\left(AS_{xy} + BS_{xy_2}\right)$

for optimum value of α_1 and β_1, differentiating above equation partially with respect to α_1 and β_1, we get

$$\alpha_1 = \frac{\bar{Y}^2}{\left\{\bar{Y}^2 + AS_y^2 + BS_{y_2}^2 - \dfrac{\left(AS_{xy} + BS_{xy_2}\right)^2}{\left(AS_x^2 + BS_{x_2}^2\right)}\right\}},$$

$$\beta_1 = \frac{\bar{Y}^2\left(AS_{xy} + BS_{xy_2}\right)}{\left(AS_x^2 + BS_{x_2}^2\right)\left\{\bar{Y}^2 + AS_y^2 + BS_{y_2}^2 - \dfrac{\left(AS_{xy} + BS_{xy_2}\right)^2}{\left(AS_x^2 + BS_{x_2}^2\right)}\right\}}$$

after putting these values in $MSE\left(T_1\right)$, we get $\min .MSE\left(T_1\right)$

$$= \frac{\bar{Y}^2\left[\left\{AS_y^2 + BS_{y_2}^2\right\} - \frac{\left\{AS_{xy} + BS_{xy_2}\right\}^2}{\left\{AS_x^2 + BS_{x_2}^2\right\}}\right]}{\bar{Y}^2 + \left[\left\{AS_y^2 + BS_{y_2}^2\right\} - \frac{\left\{AS_{xy}+BS_{xy_2}\right\}^2}{\left\{AS_x^2+BS_{x_2}^2\right\}}\right]}$$

$$\min .MSE\left(T_1\right) = \frac{\bar{Y}^2\left[MSE\left(t_1\right)\right]}{\left[\bar{Y}^2 + MSE\left(t_1\right)\right]}$$

where $A = \left(\frac{1}{n} - \frac{1}{N}\right), B = \frac{W_2\left(k-1\right)}{n}$

similarly, we get the min.MSE of all these estimators as

$$\min .MSE\left(T_i\right) = \frac{\bar{Y}^2\left[MSE\left(t_i\right)\right]}{\left[\bar{Y}^2 + MSE\left(t_i\right)\right]}$$

The optimum values of scalars for different estimators involved are given as

$$\alpha = \frac{\bar{Y}^2}{\bar{Y}^2 + AS_y^2 + BS_{y_2}^2}$$

$$\alpha_2 = \frac{\bar{Y}^2}{\left\{\bar{Y}^2 + AS_y^2(1-\rho^2) + BS_{y_2}^2\right\}}, \beta_2 = \frac{\bar{Y}^2 S_{xy}}{S_x^2\left\{\bar{Y}^2 + AS_y^2(1-\rho^2) + BS_{y_2}^2\right\}}$$

$$\alpha_3 = \frac{\bar{Y}^2}{\left\{\bar{Y}^2 + AS_y^2 + BS_{y_2}^2(1-\rho_2^2)\right\}}, \beta_3 = \frac{\bar{Y}^2 S_{xy_2}}{S_{x_2}^2\left\{\bar{Y}^2 + AS_y^2 + BS_{y_2}^2(1-\rho_2^2)\right\}}$$

$$\alpha_4 = \frac{\bar{Y}^2}{\left\{\bar{Y}^2 + AS_y^2(1-\rho^2) + BS_{y_2}^2(1-\rho_2^2)\right\}}, \beta_4 = \frac{\bar{Y}^2 S_{xy_2}}{S_{x_2}^2\left\{\bar{Y}^2+AS_y^2(1-\rho^2)+BS_{y_2}^2(1-\rho_2^2)\right\}},$$

$$\gamma_4 = \frac{\bar{Y}^2 S_{xy}}{S_x^2\left\{\bar{Y}^2 + AS_y^2(1-\rho^2) + BS_{y_2}^2(1-\rho_2^2)\right\}}$$

$$\alpha_5 = \frac{\bar{Y}^2}{\left\{\bar{Y}^2 + AS_y^2 + BS_{y_2}^2 - \frac{\left(CS_{xy} + BS_{xy_2}\right)^2}{\left(CS_x^2 + BS_{x_2}^2\right)}\right\}},$$

$$\beta_5 = \frac{\bar{Y}^2\left(CS_{xy} + BS_{xy_2}\right)}{\left(CS_x^2 + BS_{x_2}^2\right)\left\{\bar{Y}^2 + AS_y^2 + BS_{y_2}^2 - \frac{\left(CS_{xy} + BS_{xy_2}\right)^2}{\left(CS_x^2 + BS_{x_2}^2\right)}\right\}}$$

$$\alpha_6 = \frac{\bar{Y}^2}{\left\{\bar{Y}^2 + DS_y^2 + CS_y^2(1-\rho^2) + BS_{y2}^2\right\}}, \beta_2 = \frac{\bar{Y}^2 S_{xy}}{S_x^2\left\{\bar{Y}^2 + DS_y^2 + CS_y^2(1-\rho^2) + BS_{y2}^2\right\}}$$

$$\alpha_7 = \frac{\bar{Y}^2}{\left\{\bar{Y}^2 + DS_y^2 + CS_y^2(1-\rho^2) + BS_{y2}^2(1-\rho_2^2)\right\}},$$
$$\beta_7 = \frac{\bar{Y}^2 S_{xy2}}{S_{x2}^2\left\{\bar{Y}^2 + DS_y^2 + CS_y^2(1-\rho^2) + BS_{y2}^2(1-\rho_2^2)\right\}},$$

$$\gamma_7 = \frac{\bar{Y}^2 S_{xy}}{S_x^2\left\{\bar{Y}^2 + DS_y^2 + CS_y^2(1-\rho^2) + BS_{y2}^2(1-\rho_2^2)\right\}}.$$ ■

Theorem 5.3.2. *Under the optimum values of scalars,the STD estimators T_i have always lesser MSE than the conventional difference or regression type estimators t_i i.e.*

$$\min .MSE\left(T_i\right) < \min MSE\left(t_i\right) \;\; (i = 1, 2, ..., 7) \tag{5.3.19}$$

Proof. Using (5.3.19), proof is obvious. ■

Theorem 5.3.3. *The bias and minimum MSE of the new ratio type estimator T_{s_j} $(j = 1, 2, ...7)$ is given by*

$$Bias\left(T_{s_j}\right) = \overline{Y}\left(\alpha_j - 1\right) \tag{5.3.20}$$

and

$$\min MSE_{\alpha_j}\left(T_{s_j}\right) = \bar{Y}^2\left(1 - \frac{B_j^2}{A_j}\right) \tag{5.3.21}$$

Proof. The MSE of T_{s_i} $(i = 1, 2, ..., 7)$ is given by

$$\min MSE_{\alpha_i}\left(T_{s_i}\right) = \bar{Y}^2\left(1 - \frac{B_i^2}{A_i}\right) \tag{5.3.22}$$

The optimum values of scalars for different estimators involved are given as

$$A_1 = \bar{Y}^2 + P\left\{S_y^2 + \beta_1 R^2 S_x^2\left(2\beta_1 + 1\right) - 4\beta_1 R S_{xy}\right\} + Q\left\{S_{y2}^2 + \beta_1 R^2 S_{x2}^2\left(2\beta_1 + 1\right) - 4\beta_1 R S_{xy2}\right\}$$

$$B_1 = \bar{Y}^2 + P\left\{\tfrac{\beta_1}{2} R^2 S_x^2\left(\beta_1 + 1\right) - \beta_1 R S_{xy}\right\} + Q\left\{\tfrac{\beta_1}{2} R^2 S_{x2}^2\left(\beta_1 + 1\right) - 2\beta_1 R S_{xy2}\right\}$$

$$A_2 = \bar{Y}^2 + P\left\{S_y^2 + \beta_2 R^2 S_x^2\left(2\beta_2+1\right) - 4\beta_2 R S_{xy}\right\} + QS_{y_2}^2$$

$$B_2 = \bar{Y}^2 + P\left\{\tfrac{\beta_2}{2} R^2 S_x^2\left(\beta_2+1\right) - \beta_2 R S_{xy}\right\}$$

$$A_3 = \bar{Y}^2 + PS_y^2 + Q\left\{S_{y_2}^2 + \beta_3 R^2 S_{x_2}^2\left(2\beta_3+1\right) - 4\beta_3 R S_{xy_2}\right\}$$

$$B_3 = \bar{Y}^2 + Q\left\{\tfrac{\beta_3}{2} R^2 S_{x_2}^2\left(\beta_3+1\right) - \beta_3 R S_{xy_2}\right\}$$

$$\begin{aligned} A_4 = {} & \bar{Y}^2 + P\left\{S_y^2 + \beta_4 R^2 S_x^2\left(2\beta_4+1\right) + \beta_5 R^2 S_x^2\left(2\beta_5+1\right) - 4\beta_4 R S_{xy} - 4\beta_5 R S_{xy} + 4\beta_4\beta_5 R^2 S_x^2\right\} \\ & + Q\left\{S_{y_2}^2 + \beta_4 R^2 S_{x_2}^2\left(2\beta_4+1\right) - 4\beta_4 R S_{xy_2}\right\} \end{aligned}$$

$$\begin{aligned} B_4 = {} & \bar{Y}^2 + P\left\{\tfrac{\beta_4}{2} R^2 S_x^2\left(\beta_4+1\right) + \tfrac{\beta_5}{2} R^2 S_x^2\left(\beta_5+1\right) - \beta_4 R S_{xy} - \beta_5 R S_{xy} + \beta_4\beta_5 R^2 S_x^2\right\} \\ & + Q\left\{\tfrac{\beta_4}{2} R^2 S_{x_2}^2\left(\beta_4+1\right) - \beta_4 R S_{xy_2}\right\} \end{aligned}$$

$$\begin{aligned} A_5 = {} & \bar{Y}^2 + P\left(S_y^2 + \beta_6 R^2 S_x^2\left(2\beta_6+1\right) - 4\beta_6 R S_{xy}\right) + Q\left(S_{y_2}^2 + \beta_6 R^2 S_{x_2}^2\left(2\beta_6+1\right) - 4\beta_6 R S_{xy_2}\right) \\ & - T\left(\beta_6 R^2 S_x^2\left(2\beta_6+1\right) - 4\beta_6 R S_{xy}\right) \end{aligned}$$

$$\begin{aligned} B_5 = {} & \bar{Y}^2 + P\left\{\tfrac{\beta_6}{2} R^2 S_x^2\left(\beta_6+1\right) - \beta_6 R S_{xy}\right\} + Q\left\{\tfrac{\beta_6}{2} R^2 S_{x_2}^2\left(\beta_6+1\right) - \beta_6 R S_{xy_2}\right\} \\ & - T\left\{\tfrac{\beta_6}{2} R^2 S_x^2\left(\beta_6+1\right) - \beta_6 R S_{xy}\right\} \end{aligned}$$

$$A_6 = \bar{Y}^2 + P\left(S_y^2 + \beta_7 R^2 S_x^2\left(2\beta_7+1\right) - 4\beta_7 R S_{xy}\right) + QS_{y_2}^2 - T\left(\beta_7 R^2 S_x^2\left(2\beta_7+1\right) - 4\beta_7 R S_{xy}\right)$$

$$B_6 = \bar{Y}^2 + P\left\{\tfrac{\beta_7}{2} R^2 S_x^2\left(\beta_7+1\right) - \beta_7 R S_{xy}\right\} - T\left\{\tfrac{\beta_7}{2} R^2 S_x^2\left(\beta_7+1\right) - \beta_7 R S_{xy}\right\}$$

$$\begin{aligned} A_7 = {} & \bar{Y}^2 + P\left(S_y^2 + \beta_9 R^2 S_x^2\left(2\beta_9+1\right) - 4\beta_9 R S_{xy}\right) + Q\left(S_{y_2}^2 + \beta_8 R^2 S_{x_2}^2\left(2\beta_8+1\right) - 4\beta_8 R S_{xy_2}\right) \\ & - T\left(\beta_9 R^2 S_x^2\left(2\beta_9+1\right) - 4\beta_9 R S_{xy}\right) \end{aligned}$$

$$\begin{aligned} B_7 = {} & \bar{Y}^2 + P\left\{\tfrac{\beta_9}{2} R^2 S_x^2\left(\beta_9+1\right) - \beta_9 R S_{xy}\right\} + Q\left\{\tfrac{\beta_8}{2} R^2 S_{x_2}^2\left(\beta_8+1\right) - \beta_8 R S_{xy_2}\right\} \\ & - T\left\{\tfrac{\beta_9}{2} R^2 S_x^2\left(\beta_9+1\right) - \beta_9 R S_{xy}\right\} \end{aligned}$$

■

It is interesting to note that simultaneous optimization with respect to the characterizing scalars γ_i and δ_i of the expression 5.3.18 of MSE is possible for regression (difference) type estimators. But simultaneous optimization with respect to the characterizing scalars α_i and β_i of the expression 5.3.21 not possible for ratio type estimators.

Theorem 5.3.4. *The proposed ratio type estimators T_{s_j}, $(j = 1, 2, ..., 7)$ are better than difference type estimators T_k, $(k = 1, 2, ..., 7)$ iff*

$$\frac{B_j^2}{A_j} > \gamma_{kopt} \tag{5.3.23}$$

and vice versa. Otherwise both are equally efficient in case of equality in (5.3.23).

Proof. It may be easily observed from (5.3.18) that the MSE of the difference type estimators T_i, $(i = 1, 2, ..., 7)$ are given by

$$\min MSE\left(T_i\right) = \bar{Y}^2\left(1 - \gamma_{i_{opt}}\right) \tag{5.3.24}$$

Comparing (5.3.24) with (5.3.21), we have the theorem. ■

The only way ascertain (5.3.23) whether this holds in practice or not is through a computational study.

Theorem 5.3.5. *The proposed ratio type estimators T_{s_j}, $j = 1, 2, ..., 7$ are better than the conventional ratio type estimators t_{R_j}, $j = 1, 2, ..., 7$ iff*

$$\frac{B_j^2}{A_j} > \left\{1 - \frac{min.MSE(t_{R_j})}{\bar{Y}^2}\right\} \tag{5.3.25}$$

and vice versa. Otherwise both are equally efficient in case of equality in (5.3.25).

Proof. It may be easily observed from (5.3.21) and (5.3.18), we have the theorem. ■

Theorem 5.3.6. *The proposed ratio type estimators T_{s_j}, $j = 1, 2, ..., 7$ are better than the* *conventional regression type estimators t_i, $i = 1, 2, ..., 7$ iff*

$$\frac{B_j^2}{A_j} > \left\{1 - \frac{min.MSE(t_i)}{\bar{Y}^2}\right\} \tag{5.3.26}$$

and vice versa. Otherwise both are equally efficient in case of equality in (5.3.26).

Proof. It may be easily observed from (5.3.21) and (5.3.18), we have the theorem. ∎

5.4 Empirical study

In order to have a better understanding about the efficiency of the proposed estimators we have conducted a comprehensive empirical study on three populations and compared the proposed estimators with the existing estimators.

The percentage relative efficiency (PRE) with respect to $\bar{y}^*$ is calculated as

$$PRE = \frac{Var(\bar{y}^*)}{min.MSE(T_{s_j})} \times 100, j = 1, 2, ..., 7. \tag{5.4.1}$$

1. The first population considered from Srivastava $(1993, p.50)$ consist a list of 70 villages in a Tehsil of India along with their population in 1981 and cultivated area (in acres) in the same year is taken in to consideration. Here the cultivated area (in acres) is taken as main study character and the population of village is taken as auxiliary character.The parameters of the population are as follows:
   $N = 70$,$n' = 40$,$n = 25$,$\bar{Y} = 981.29$,$\bar{X} = 1755.53$,$S_y = 613.66$, $S_x = 1406.13$, $\bar{Y}_2 = 597.29$, $\bar{X}_2 = 1100.24$, $S_{y_2} = 244.11$,$S_{x_2} = 631.51$, $\rho = 0.778$, $\rho_2 = 0.445$, $R = 0.5589$ $\beta = 0.3395$, $\beta_2 = 0.1720$, $W_2 = 0.20$.

2. For the population of 96 villages of rural areas under polish station, Singur, District Hooghly from district cencus Handbook (1981), published by government of India, the data on yhe number of cultivators y, as a study chracter and the population of villages, as an auxiliary chracter x have been taken. The non-response rate in the population is considered to be 25 percent. The values of the parameters of the population are given as follows:

$N = 96$,$n' = 65$,$n = 25$,$\bar{Y} = 185.22$,$\bar{X} = 1807.23$, $S_y = 195.03$,$S_x = 1921.77$, $S_{y2} = 97.82$, $S_{x_2} = 1068.44$, $\rho = 0.904$, $\rho_2 = 0.895$, $R = 0.1025$, $W_2 = 0.25$.

3. Third population considered from Srivastava ($1993, p.50$).

   The data belongs to the data on physical growth of upper-socio-economic group of 95 school children of Varanasi under an ICMR study.The first 25% (i.e.24 children) units have been considered as non-response units.

   The values of the parameters related to the study variate y (the weight in kg) and the auxiliary variate x (the chest circumference in cm) have been given below:

   $N = 95$,$n' = 70$,$n = 35$,$\bar{Y} = 19.497$,$\bar{X} = 55.8611$, $S_y = 3.0435$,$S_x = 3.2735$, $S_{y_2} = 2.3552$, $S_{x_2} = 2.5137$, $\rho = 0.8460$, $\rho_2 = 0.7290$, $R = 0.3490$, $\beta = 0.7865$, $\beta_2 = 0.6829$, $W_2 = 0.25$ $N_2 = 24$, $N_1 = 71$.

Table 5.1: MSE and PRE of the estimators

Estimator	k=2	k=3	k=4	k=5
$\bar{y}^*$	10160.17(100)	10636.88(100)	11113.60(100)	11590.32(100)
T_s	10054.08(101.05)	10520.67(101.10)	10986.79 (101.15)	11452.47(101.20)
Strategy I				
t_{R_1}	4288.77(236.90)	4745.94(224.13)	5195.20 (213.92)	5637.74(205.58)
T_1	4269.75(237.95)	4722.66(225.23)	5167.32 (215.07)	5604.92(206.78)
T_{s_1}	4247.51**(239.20)**	4696.51**(226.48)**	5137.05**(216.34)**	5570.35**(208.07)**
Strategy II				
t_{R_2}	4298.93(236.34)	4775.65(222.73)	5252.36 (211.59)	5729.08(202.31)
T_2	4279.82(237.39)	4752.08(223.83)	5223.87 (212.74)	5695.19(203.51)
T_{s_2}	4259.43**(238.53)**	4729.86**(224.89)**	5199.82**(213.73)**	5669.32**(204.44)**
Strategy III				
t_{R_3}	10065.76(100.94)	10448.08(101.81)	10830.39(102.61)	11212.71(103.37)
T_3	9961.63(101.99)	10335.93(102.91)	10709.93(103.77)	11083.65(104.57)
T_{s_3}	9959.45**(102.01)**	10331.38**(102.96)**	10702.83**(103.84)**	11073.81**(104.66)**
Strategy IV				
t_{R_4}	4204.53(241.65)	4586.84(231.90)	4969.16 (223.65)	5351.47(216.58)
T_4	4186.25(242.70)	4565.09(233.00)	4961.16 (224.80)	5321.89(217.78)
T_{s_4}	4164.88**(243.95)**	4540.74**(234.25)**	4916.11**(226.06)**	5291.00**(219.06)**
Strategy V				
t_{R_5}	6727.54(151.02)	7182.84(148.09)	7638.14 (145.50)	8093.44(143.21)
T_5	6680.87(152.07)	7123.29(149.32)	7554.66 (147.11)	7977.40(145.29)
T_{s_5}	6662.06**(152.51)**	7104.86**(149.71)**	7546.58**(147.26)**	7987.23**(145.11)**
Strategy VI				
t_{R_6}	6741.11(150.72)	7217.83(147.37)	7694.55 (144.43)	8171.26(141.84)
T_6	6694.24(151.77)	7164.13(148.47)	7633.55(145.58)	8102.50(143.04)
T_{s_6}	6677.77**(152.15)**	7146.59**(148.84)**	7614.95**(145.94)**	8082.86**(143.39)**
Strategy VII				
t_{R_7}	6646.71(152.86)	7029.02(151.33)	7411.34(149.95)	7793.66(148.71)
T_7	6601.14(153.91)	6978.08(152.43)	7354.73(151.10)	7731.08(149.91)
T_{s_7}	6583.19**(154.33)**	6957.40**(152.88)**	7331.14**(151.59)**	7704.39**(150.44)**

Table 5.2: MSE and PRE of the estimators

Estimator	k=2	k=3	k=4	k=5
$\bar{y}^*$	1220.94(100)	1316.63(100)	1412.31(100)	1508.00(100)
T_s	1178.98(103.56)	1267.96(103.84)	1356.47(104.12	1444.51(104.40)
Strategy I				
t_{R_1}	225.71(540.93)	245.57(536.15)	265.30(532.34)	284.93(529.25)
T_1	224.23(544.49)	243.83(539.99)	263.26(536.46)	282.58(533.65)
T_{s_1}	223.29(**546.80**)	242.63(**542.65**)	261.79 (**539.48**)	280.80(**537.03**)
Strategy II				
t_{R_2}	301.36(405.14)	397.05(331.61)	492.74(286.62)	588.43(256.28)
T_2	298.74(408.69)	392.51(335.44)	485.76(290.74)	578.50(260.67)
T_{s_2}	297.73(**410.08**)	391.21(**336.55**)	484.18 (**291.69**)	576.64(**261.51**)
Strategy III				
t_{R_3}	1144.29(106.70)	1163.33(113.18)	1182.37(119.45)	1201.41(125.52)
T_3	1107.36(110.26)	1125.18(117.01)	1142.98(123.56)	1160.76(129.91)
T_{s_3}	1106.75(**110.32**)	1123.95(**117.14**)	1141.10(**123.77**)	1158.21(**130.20**)
Strategy IV				
t_{R_4}	224.72(312.90)	243.76(291.08)	262.79(277.11)	281.84(267.40)
T_4	223.25(546.88)	242.04(543.98)	260.80(541.53)	279.54(539.46)
T_{s_4}	222.32(**549.18**)	240.86(**546.63**)	259.36(**544.54**)	277.81(**542.82**)
Strategy V				
t_{R_5}	380.12(226.28)	400.02(223.52)	419.93(221.57)	439.83(220.11)
T_5	375.95(324.76)	395.33(333.05)	414.54(340.69)	433.63(347.76)
T_{s_5}	374.68(**325.86**)	393.68(**334.44**)	412.61(**342.28**)	431.48(**349.49**)
Strategy VI				
t_{R_6}	455.79(184.51)	551.48(162.44)	647.16(149.51)	742.85(141.01)
T_6	449.81(271.43)	542.75(242.58)	635.18(222.35)	727.11(207.40)
T_{s_6}	448.60(**272.17**)	541.30(**243.23**)	633.50(**222.94**)	725.19(**207.94**)
Strategy VII				
t_{R_7}	379.14(227.20)	398.18(224.87)	417.22(223.21)	436.26(221.97)
T_7	374.99(325.59)	393.61(334.50)	412.21(342.62)	430.78(350.06)
T_{s_7}	373.74(**326.68**)	392.05(**335.83**)	410.31(**344.20**)	428.53(**351.90**)

Table 5.3: MSE and PRE of the estimators

Estimator	k=2	k=3	k=4	k=5
$\bar{y}^*$	0.2067(100)	0.2464(100)	0.2860(100)	0.3256(100)
T_s	0.2066(100.054)	0.2462(100.064)	0.2857(100.075)	0.3253(100.085)
Strategy I				
t_{R_1}	0.0664(311.05)	0.0853(288.83)	0.1040(274.84)	0.1227(265.22)
T_1	0.0664**(311.104)**	0.0852**(288.893)**	0.1040**(274.913)**	0.1227**(265.307)**
T_{s_1}	0.0664(311.06)	0.0853(288.84)	0.1040 (274.86)	0.1227(265.25)
Strategy II				
t_{R_2}	0.0871(237.29)	0.1267(194.38)	0.1663(171.90)	0.2060(158.07)
T_2	0.0871**(237.342)**	0.1267**(194.441)**	0.1663**(171.977)**	0.2058**(158.158)**
T_{s_2}	0.0871(237.31)	0.1267(194.41)	0.1663 (171.95)	0.2059(158.13)
Strategy III				
t_{R_3}	0.1857(111.34)	0.2043(120.61)	0.2228(128.35)	0.2414(134.89)
T_3	0.1856**(111.392)**	0.2041**(120.680)**	0.2227**(128.421)**	0.2412**(134.974)**
T_{s_3}	0.1856(111.39)	0.2042(120.67)	0.2227 (128.41)	0.2412(134.96)
Strategy IV				
t_{R_4}	0.0660(312.89)	0.0846(291.08)	0.1032(277.11)	0.1217(267.40)
T_4	0.0660**(312.950)**	0.0846**(291.144)**	0.1031**(277.186)**	0.1217**(267.486)**
T_{s_4}	0.0660(312.91)	0.0846(291.09)	0.1032 (277.13)	0.1217(267.43)
Strategy V				
t_{R_5}	0.0913(226.28)	0.1102(223.52)	0.1290(221.57)	0.1479(220.11)
T_5	0.0913**(226.331)**	0.1101**(223.692)**	0.1288**(221.952)**	0.1475**(220.722)**
T_{s_5}	0.0913(226.30)	0.1102(223.55)	0.1290 (221.60)	0.1479(220.15)
Strategy VI				
t_{R_6}	0.1120184.51)	0.1516(162.44)	0.1913(149.51)	0.2309(141.01)
T_6	0.1120(184.568)	0.1516(162.503)	0.1912**(149.581)**	0.2307**(141.098)**
T_{s_6}	0.1120184.54)	0.1516(162.48)	0.1912(149.56)	0.2308(141.08)
Strategy VII				
t_{R_7}	0.0910(227.20)	0.1095(224.87)	0.1281(223.21)	0.1467(221.97)
T_7	0.0909**(227.259)**	0.1095**(224.934)**	0.1280 **(223.286)**	0.1466**(222.057)**
T_{s_7}	0.0909(227.23)	0.1095(224.89)	0.12811 (223.24)	0.14667(222.01)

From perusal of above results it is observed that the new ratio type estimators proposed T_{s_i} are always better than the conventional ratio type counterparts t_{R_i}. Hence, we conclude that all proposed new ratio type estimators have higher efficiency in comparision to the conventional ratio type estimators. A comparision of STD (or regression) estimators $T_i(i = 1, 2, ..., 7)$ with new ratio type estimators $T_{s_i}(i = 1, 2, ..., 7)$, we observe that the new ratio type estimators are always better than the estimators proposed by Bhushan and Pandey regression type estimators under optimality condition (5.3.23). Therefore, in population I and II proposed ratio type estimators T_{s_j} are better than Bhushan and Pandey regression type estimators T_i, as (5.3.23) is satisfied. While for population III the case is reversed.

5.5 Conclusion

Diana and Perri (2013) in their study opined that the regression and difference type estimator provide best possible improvement and cannot be improved upon. In this paper, we have proposed various Searls type difference estimators, which provide an improvement over the corresponding regression estimators in their respective strategies. The result are proved both theoretically as well as empirically. The results of conventional difference or regression type estimators are a special case of the respective Searls type estimators for $\alpha_i = 1$ respectively.

Chapter 6

An efficient classes of estimators using two auxiliary variables under non-response

6.1 Introduction

The survey statistician very often make use of the available auxiliary information to improve the precision of the estimates. In situations where the population mean of the auxiliary variable is known and the non-response is present have been dealt by various authors including Cochran (1977), Rao (1986, 1987), Khare and Srivastava (1993). However, in situations when the population mean of the auxiliary variable is unknown , the sample mean $\bar{x}'$ obtained from a large first phase sample of size n' drawn from N units by simple random sampling without replacement is used as suggested by Khare and Srivastava (1997), Okafor and Lee (2000), Tabasum and Khan (2004) and recently by Singh and kumar (2010) among others. It is assumed that all the first phase smple units supplied the auxiliary information, then, a second phase sample of size n $(n < n')$ is drawn from the n' by SRSWOR and study variable y is measured on it. At the second phase from the sample of size n, let n_1 units respond and n_2 units refuse to respond. Now, we use Hansen and Hurwitz sampling strategy to sub-sample r units from n_2 non-responding units and enumerated by direct interview such that $r = n_2/k, k > 1$.

In this chapter, we consider the deterministic setup of non-response exactly on the similar lines as that of Okafor and Lee (2000) and assume that the whole population (denoted by Ω) is divided in to two groups, one is that of the responding group of N_1 units (denoted by Ω_1) and the other non-responding group of N_2 units (denoted by Ω_2). Let the first and second phase samples be denoted by u and u' respectively, and let $u_1 = u \cap \Omega_1$ and $u_2 = u \cap \Omega_2$. the sub sample from u_2 is denoted by $u_{(2)}$.

6.2 Existing estimators

The conventional ratio and regression typr estimators for the population mean when there is non-response on the study variable as well as on the auxiliary variable are given as

$$\bar{y}_{R(1)} = \bar{y}^* \frac{\bar{x}'}{\bar{x}^*} \tag{6.2.1}$$

$$\bar{y}_{Reg_{(1)}} = \bar{y}^* + b^*_{yx}\left(\bar{x}' - \bar{x}^*\right) \tag{6.2.2}$$

the MSE's of these estimators are given by

$$\begin{aligned} MSE\left(\bar{y}_{R(1)}\right) &= \bar{Y}^2\left[\left(f_n - f_n^{'}\right)\left\{C_y^2 + \left(1 - 2K_{yx}\right)C_x^2\right\} + g_n\left\{C_{y(2)}^2 + \left(1 - 2K_{yx(2)}\right)C_{x(2)}^2\right\}\right. \\ &\quad \left. + f_n^{'}C_y^2\right] \end{aligned} \tag{6.2.3}$$

$$\begin{aligned} MSE\left(\bar{y}_{Reg_{(1)}}\right) &= \bar{Y}^2\left[\left(f_n - f_n^{'}\right)\left(1 - \rho_{yx}^2\right)C_y^2 + g_n\left\{C_{y(2)}^2 + K_{yx}\left(K_{yx} - 2K_{yx(2)}\right)C_{x(2)}^2\right\}\right. \\ &\quad \left. + f_n^{'}C_y^2\right] \end{aligned} \tag{6.2.4}$$

and the conventional ratio and regression type estimators for population mean when there is non-response on the study variable alone are given as

$$\bar{y}_{R(2)} = \bar{y}^* \frac{\bar{x}'}{\bar{x}} \tag{6.2.5}$$

$$\bar{y}_{Reg(2)} = \bar{y}^* + b_{yx}^{**}(\bar{x}' - \bar{x}) \tag{6.2.6}$$

the MSE's of these estimators are given by

$$MSE\left(\bar{y}_{R(2)}\right) = \bar{Y}^2\left[\left(f_n - f_n^{'}\right)\left\{C_y^2 + (1 - 2K_{yx})C_x^2\right\} + g_n C_{y(2)}^2 + f_n^{'}C_y^2\right] \tag{6.2.7}$$

$$MSE\left(\bar{y}_{Reg(1)}\right) = \bar{Y}^2\left[\left(f_n - f_n^{'}\right)\left(1 - \rho_{yx}^2\right)C_y^2 + g_n C_{y(2)}^2 + f_n^{'}C_y^2\right] \tag{6.2.8}$$

Singh and Kumar (2010b) proposed the following estimators using information based on two auxiliary variables x and z for estimating the population mean of the study variable y under different situations is given by

$$\bar{y}_{(SK)}^1 = \left\{\bar{y}^* + b_{yx}^*(\bar{x}' - \bar{x}^*)\right\}\frac{\bar{Z}}{\left\{\bar{Z} + \eta_1\left(\bar{z}^* - \bar{Z}\right)\right\}} \tag{6.2.9}$$

$$\bar{y}_{(SK)}^2 = \left\{\bar{y}^* + b_{yx}^*(\bar{x}' - \bar{x}^*)\right\}\frac{\bar{z}'}{\left\{\bar{z}' + \eta_2\left(\bar{z}^* - \bar{z}'\right)\right\}} \tag{6.2.10}$$

$$\bar{y}_{(SK)}^3 = \left\{\bar{y}^* + b_{yx}^{**}(\bar{x}' - \bar{x})\right\}\frac{\bar{Z}}{\left\{\bar{Z} + \eta_3\left(\bar{z} - \bar{Z}\right)\right\}} \tag{6.2.11}$$

$$\bar{y}_{(SK)}^4 = \left\{\bar{y}^* + b_{yx}^{**}(\bar{x}' - \bar{x})\right\}\frac{\bar{z}'}{\left\{\bar{z}' + \eta_4\left(\bar{z} - \bar{z}'\right)\right\}} \tag{6.2.12}$$

where

$\bar{y}^* = \frac{n_1}{n}\bar{y}_{(1)} + \frac{n_2}{n}\bar{y}_{(2)}^*$, $\bar{x}^* = \frac{n_1}{n}\bar{x}_{(1)} + \frac{n_2}{n}\bar{x}_{(2)}^*$, $\bar{z}^* = \frac{n_1}{n}\bar{z}_{(1)} + \frac{n_2}{n}\bar{z}_{(2)}^*$, $\bar{x}' = \frac{1}{n'}\sum(x_i)$, $\bar{z}' = \frac{1}{n'}\sum(z_i)$, $\bar{x} = \frac{1}{n}\sum(x_i)$ and $\bar{z} = \frac{1}{n}\sum(z_i)$ with $\left(\bar{y}_{(1)}, \bar{x}_{(1)}, \bar{z}_{(1)}\right)$ and $\left(\bar{y}_{(2)}^*, \bar{x}_{(2)}^*, \bar{z}_{(2)}^*\right)$ being the sample means based on n_1 units and sub-sample means based on r units of the variates (y, x, z) respectively and η_1, η_2, η_3 and η_4 are suitably chosen constants.

The MSE's of these estimators are given by

$$\begin{aligned} MSE\left(\bar{y}_{SK}^1\right) &= \bar{Y}^2\Big[f_n^{'}\left\{C_y^2 + \eta_1(\eta_1 - 2K_{yz})C_z^2\right\} + \left(f_n - f_n^{'}\right)\left\{\left(1 - \rho_{yx}^2\right)C_y^2 + \right. \\ &\quad \left.\eta_1(\eta_1 - 2A^*)C_z^2\right\} + g_n\left\{C_{y(2)}^2 + K_{yx}\left(K_{yx} - 2K_{yx(2)}\right)C_{x(2)}^2 + \right. \\ &\quad \left.\eta_1(\eta_1 - 2B^*)C_{z(2)}^2\right\}\Big] \end{aligned} \tag{6.2.13}$$

$$MSE\left(\bar{y}^2_{SK}\right) = \bar{Y}^2\left[f'_n C_y^2 + \left(f_n - f'_n\right)\left\{\left(1-\rho^2_{yx}\right)C_y^2 + \eta_2\left(\eta_2 - 2A^*\right)C_z^2\right\} + g_n\left\{C^2_{y(2)} + K_{yx}\left(K_{yx} - 2K_{yx(2)}\right)C^2_{x(2)} + \eta_2\left(\eta_2 - 2B^*\right)C^2_{z(2)}\right\}\right] \tag{6.2.14}$$

$$MSE\left(\bar{y}^3_{SK}\right) = \bar{Y}^2\left[f'_n\left\{C_y^2 + \eta_3\left(\eta_3 - 2K_{yz}\right)C_z^2\right\} + \left(f_n - f'_n\right)\left\{\left(1-\rho^2_{yx}\right)C_y^2 + \eta_3\left(\eta_3 - 2A^*\right)C_z^2\right\} + g_n C^2_{y(2)}\right] \tag{6.2.15}$$

$$MSE\left(\bar{y}^4_{SK}\right) = \bar{Y}^2\left[f'_n C_y^2 + \left(f_n - f'_n\right)\left\{\left(1-\rho^2_{yx}\right)C_y^2 + \eta_4\left(\eta_4 - 2A^*\right)C_z^2\right\} + g_n C^2_{y(2)}\right] \tag{6.2.16}$$

where

$C_x^2 = S_x^2/\bar{X}^2$	$C^2_{x(2)} = S^2_{x(2)}/\bar{X}^2$	$C_y^2 = S_y^2/\bar{Y}^2$
$C^2_{y(2)} = S^2_{y(2)}/\bar{Y}^2$	$\rho_{yx} = S_{yx}/S_yS_x$	$\rho_{yx(2)} = S_{yx(2)}/S_{y(2)}S_{x(2)}$
$\rho_{xz} = S_{xz}/S_xS_z$	$\rho_{xz(2)} = S_{xz(2)}/S_{x(2)}S_{z(2)}$	$\rho_{yz} = S_{yz}/S_yS_z$
$\rho_{yz(2)} = S_{yz(2)}/S_{y(2)}S_{z(2)}$	$K_{yx} = \rho_{yx}C_y/C_x$	$K_{yx(2)} = \rho_{yx(2)}C_{y(2)}/C_{x(2)}$
$K_{yz} = \rho_{yz}C_y/C_z$	$K_{yz(2)} = \rho_{yz(2)}C_{y(2)}/C_{z(2)}$	$K_{xz} = \rho_{xz}C_x/C_z$
$K_{xz(2)} = \rho_{xz(2)}C_{x(2)}/C_{z(2)}$	$A^* = K_{yz} - K_{yx}K_{xz}$	$B^* = K_{yz(2)} - K_{yx}K_{xz(2)}$

Also, the minimum MSE's of the above stated estimators are given by

$$\min .MSE\left(\bar{y}^1_{SK}\right) = MSE\left(\bar{y}_{Reg_{(1)}}\right) - \bar{Y}^2\left[\frac{B_1^2}{A_1}\right] \tag{6.2.17}$$

$$\min .MSE\left(\bar{y}^2_{SK}\right) = MSE\left(\bar{y}_{Reg_{(1)}}\right) - \bar{Y}^2\left[\frac{B_2^2}{A_1}\right] \tag{6.2.18}$$

$$\min .MSE\left(\bar{y}^3_{SK}\right) = MSE\left(\bar{y}_{Reg_{(2)}}\right) - \bar{Y}^2\left[\frac{B_3^2}{A_2}\right] \tag{6.2.19}$$

$$\min .MSE\left(\bar{y}^4_{SK}\right) = MSE\left(\bar{y}_{Reg(2)}\right) - \bar{Y}^2 f' A^{*2}C_z^2 \tag{6.2.20}$$

where

$B_1 = f'_n K_{yz}C_z^2 + \left(f_n - f'_n\right)A^*C_z^2 + g_nB^*C^2_{z(2)}$, $A_1 = f_nC_z^2 + g_nC^2_{z(2)}$, $B_2 = \left(f_n - f'_n\right)A^*C_z^2 + g_nB^*C^2_{z(2)}$, $B_3 = f'_nK_{yz} + \left(f_n - f'_n\right)A^*$ and $A_2 = f_n$.

Further, Shabbir and Khan (2013) proposed the following estimators using information based on two auxiliary variables x and z for estimating the population mean of the study variable y under the different situations is given by

$$\bar{y}^5_{(SK)} = \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}^*}\right)^{\alpha_1} \left(\frac{\bar{Z}}{\bar{z}^*}\right)^{\alpha_2} + b^*_{yx}\left(\bar{x}' - \bar{x}^*\right) + b^*_{yz}\left(\bar{Z} - \bar{z}^*\right) \tag{6.2.21}$$

$$\bar{y}^6_{(SK)} = \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}^*}\right)^{\alpha_3} \left(\frac{\bar{z}}{\bar{z}^*}\right)^{\alpha_4} + b^*_{yx}\left(\bar{x}' - \bar{x}^*\right) + b^*_{yz}\left(\bar{z}' - \bar{z}^*\right) \tag{6.2.22}$$

$$\bar{y}^7_{(SK)} = \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}}\right)^{\alpha_5} \left(\frac{\bar{Z}}{\bar{z}}\right)^{\alpha_6} + b^{**}_{yx}\left(\bar{x}' - \bar{x}\right) + b^{**}_{yz}\left(\bar{Z} - \bar{z}\right) \tag{6.2.23}$$

$$\bar{y}^8_{(SK)} = \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}}\right)^{\alpha_7} \left(\frac{\bar{z}'}{\bar{z}}\right)^{\alpha_8} + b^{**}_{yx}\left(\bar{x}' - \bar{x}\right) + b^{**}_{yz}\left(\bar{z}' - \bar{z}\right) \tag{6.2.24}$$

where α_1, α_2, α_3, α_4, α_5, α_6, α_7 and α_8 are the characterizing scalars to be chosen suitably and $b^*_{yx} = s^*_{yx}/s^{*2}_x$, $b^*_{yz} = s^*_{yz}/s^{*2}_z$, $b^{**}_{yx} = s^*_{yx}/s^2_x$, $b^{**}_{yz} = s^{**}_{yz}/s^2_z$ are simple regression coefficients.

The MSE's of these estimators are given by

$$\begin{aligned}
MSE\left(\bar{y}^5_{SK}\right) &= \bar{Y}^2\left[f_n C^2_y + g_n C^2_{y(2)} + \left(\alpha_1 + K_{yx}\right)^2 m_1 + \left(\alpha_2 + K_{yz}\right)^2 m_2 - 2\left(\alpha_1 + K_{yx}\right) m_3\right. \\
&\quad \left. -2\left(\alpha_2 + K_{yz}\right) m_4 + 2\left(\alpha_1 + K_{yx}\right)\left(\alpha_2 + K_{yz}\right) m_5\right] && (6.2.25) \\
MSE\left(\bar{y}^6_{SK}\right) &= \bar{Y}^2\left[f_n C^2_y + g_n C^2_{y(2)} + \left(\alpha_3 + K_{yx}\right)^2 m_1 + \left(\alpha_4 + K_{yz}\right)^2 m'_2 - 2\left(\alpha_3 + K_{yx}\right) m_3\right. \\
&\quad \left. -2\left(\alpha_4 + K_{yz}\right) m'_4 + 2\left(\alpha_3 + K_{yx}\right)\left(\alpha_4 + K_{yz}\right) m_5\right] && (6.2.26) \\
MSE\left(\bar{y}^7_{SK}\right) &= \bar{Y}^2\left[f'_n C^2_y + g_n C^2_{y(2)} + \left\{\left(f_n - f'_n\right) C^2_y + \left(\alpha_5 + K_{yx}\right)^2 C^2_x - 2\left(\alpha_5 + K_{yx}\right) \times\right.\right. \\
&\quad \rho_{yx} C_y C_x + 2\left(\alpha_5 + K_{yx}\right)\left(\alpha_6 + K_{yz}\right)\rho_{xz} C_x C_z + f_n\left\{\left(\alpha_6 + K_{yz}\right)^2 C^2_z - \right. \\
&\quad \left.\left. 2\left(\alpha_6 + K_{yz}\right)\rho_{yz} C_y C_z\right\}\right] && (6.2.27) \\
MSE\left(\bar{y}^8_{SK}\right) &= \bar{Y}^2\left[f'_n C^2_y + g_n C^2_{y(2)} + \left(f_n - f'_n\right)\left\{C^2_y + \left(\alpha_7 + K_{yx}\right)^2 C^2_x - 2\left(\alpha_7 + K_{yx}\right) \times\right.\right. \\
&\quad \rho_{yx} C_y C_x + \left(\alpha_8 + K_{yz}\right)^2 C^2_z - 2\left(\alpha_8 + K_{yz}\right)\rho_{yz} C_y C_z + 2\left(\alpha_7 + K_{yx}\right) \times \\
&\quad \left(\alpha_8 + K_{yz}\right)\rho_{xz} C_x C_z + f_n\left\{\left(\alpha_6 + K_{yz}\right)^2 C^2_z - 2\left(\alpha_6 + K_{yz}\right)\rho_{yz} C_y C_z\right\}\Big] && (6.2.28)
\end{aligned}$$

where

$$m_1 = (f_n - f'_n)\, C_x^2 + g_n C_{x(2)}^2 \qquad m_2 = f_n C_z^2 + g_n C_{z(2)}^2$$

$$m_3 = (f_n - f'_n)\, \rho_{yx} C_y C_x + g_n \rho_{yx(2)} C_{y(2)} C_{x(2)} \qquad m_4 = f_n \rho_{yz} C_y C_z + g_n \rho_{yz(2)} C_{y(2)} C_{z(2)}$$

$$m_5 = (f_n - f'_n)\, \rho_{xz} C_x C_z + g_n \rho_{xz(2)} C_{x(2)} C_{z(2)} \qquad m'_2 = (f_n - f'_n)\, C_z^2 + g_n C_{z(2)}^2$$

$$m'_4 = (f_n - f'_n)\, \rho_{yz} C_y C_z + g_n \rho_{yz(2)} C_{y(2)} C_{z(2)}$$

Also, the minimum MSE's of the above stated estimators are given by

$$\min .MSE\left(\bar{y}_{SK}^5\right) = MSE\left(\bar{y}^*\right) - \bar{Y}^2 \left\{ \frac{m_2 m_3^2 + m_1 m_4^2 - 2 m_3 m_4 m_5}{m_1 m_2 - m_5^2} \right\} \tag{6.2.29}$$

$$\min .MSE\left(\bar{y}_{SK}^6\right) = MSE\left(\bar{y}^*\right) - \bar{Y}^2 \left\{ \frac{m'_2 m_3^2 + m_1 m_4'^2 - 2 m_3 m_4 m_5}{m_1 m'_2 - m_5^2} \right\} \tag{6.2.30}$$

$$\min .MSE\left(\bar{y}_{SK}^7\right) = MSE\left(\bar{y}^*\right) - \bar{Y}^2 \left[\frac{\left\{ f_n \rho_{yz}^2 + f'_n \left(\rho_{yx}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}\right)\right\}}{\left\{ f_n - (f_n - f'_n)\, \rho_{yx}^2 \right\}} \right] \tag{6.2.31}$$

$$\min .MSE\left(\bar{y}_{SK}^8\right) = MSE\left(\bar{y}^*\right) - \bar{Y}^2 f'_n C_y^2 \left\{ \frac{\rho_{yx}^2 + \rho_{yz}^2 - 2\rho_{yx}\rho_{yz}\rho_{xz}}{\left(1 - \rho_{xz}^2\right)} \right\} \tag{6.2.32}$$

On the similar lines, Bhushan and Naqvi (2015) proposed some generalised classes of estimators in presence of non-response using two auxiliary information, given by

$$t_{(BN)}^1 = f_1\left(\bar{y}^*, \bar{x}^*, \bar{x}', u\right) \tag{6.2.33}$$

$$t_{(BN)}^2 = f_2\left(\bar{y}^*, \bar{x}^*, \bar{x}', \bar{z}^*, \bar{z}'\right) \tag{6.2.34}$$

$$t_{(BN)}^3 = f_3\left(\bar{y}^*, \bar{x}^*, \bar{x}', v\right) \tag{6.2.35}$$

$$t_{(BN)}^4 = f_4\left(\bar{y}^*, \bar{x}, \bar{x}', \bar{z}, \bar{z}'\right) \tag{6.2.36}$$

where $u = \bar{z}^*/\bar{Z}$ and $v = \bar{z}/\bar{Z}$ and these estimators satisfying the following conditions $f_1\left(\bar{Y}, \bar{X}, \bar{X}, 1\right) = \bar{Y}$, $f_1^0 = 1$, $f_1^1 = -f_1^2$, $f_2\left(\bar{Y}, \bar{X}, \bar{X}, \bar{Z}, \bar{Z}\right) = \bar{Y}$, $f_2^0 = 1$, $f_2^1 = -f_2^2$, $f_2^3 = -f_2^4$, $f_3\left(\bar{Y}, \bar{X}, \bar{X}, 1\right) = \bar{Y}$, $f_3^0 = 1$, $f_3^1 = -f_3^2$, $f_4\left(\bar{Y}, \bar{X}, \bar{X}, \bar{Z}, \bar{Z}\right) = \bar{Y}$, $f_4^0 = 1$, $f_4^1 = -f_4^2$ and $f_4^3 = -f_4^4$.

and the MSE's and minimum MSE's are given by

$$MSE\left(t^{1}_{(BN)}\right) = MSE\left(\bar{y}^{*}\right) + \eta_{6}\left(f^{3}_{1}\right)^{2} + 2\bar{Y}\eta_{7}f^{3}_{1} + \bar{X}^{2}\eta_{3}\left(f^{1}_{1}\right)^{2} + 2\bar{Y}\bar{X}\eta_{8}f^{1}_{1} + 2\bar{X}\eta_{9}f^{1}_{1}f^{3}_{1} \tag{6.2.37}$$

$$MSE\left(t^{2}_{(BN)}\right) = MSE\left(\bar{y}^{*}\right) + \bar{X}^{2}\eta_{3}\left(f^{1}_{2}\right)^{2} + \bar{Z}^{2}\eta_{10}\left(f^{3}_{2}\right)^{2} + 2\bar{X}\bar{Z}f^{1}_{2}f^{3}_{2} + 2\bar{Y}\bar{X}\eta_{8}f^{1}_{2} + 2\bar{Y}\bar{Z}\eta_{11}f^{3}_{2} \tag{6.2.38}$$

$$MSE\left(t^{3}_{(BN)}\right) = MSE\left(\bar{y}^{*}\right) + \bar{X}^{2}\eta^{'}_{3}\left(f^{1}_{3}\right)^{2} + \eta_{4}\left(f^{3}_{3}\right)^{2} + 2\bar{X}\eta^{'}_{9}f^{1}_{3}f^{3}_{3} + 2\bar{Y}\bar{X}\eta^{'}_{8}f^{1}_{3} + 2\bar{Y}\eta_{5}f^{3}_{3} \tag{6.2.39}$$

$$MSE\left(t^{4}_{(BN)}\right) = MSE\left(\bar{y}^{*}\right) + \bar{X}^{2}\eta^{'}_{3}\left(f^{1}_{4}\right)^{2} + \bar{Z}^{2}\eta_{10}\left(f^{3}_{4}\right)^{2} + 2\bar{X}\bar{Z}\eta^{'}_{9}f^{1}_{4}f^{3}_{4} + 2\bar{Y}\bar{X}\eta^{'}_{8}f^{1}_{4} + 2\bar{Y}\bar{Z}\eta^{'}_{11}f^{3}_{4} \tag{6.2.40}$$

where

$$\begin{aligned}
m_{6} &= f_{n}C^{2}_{z}, & m_{7} &= f_{n}\rho_{yz}C_{y}C_{z}, \\
m^{'}_{1} &= \left(f_{n} - f^{\prime}_{n}\right)C^{2}_{x}, & m^{'}_{3} &= \left(f_{n} - f^{\prime}_{n}\right)\rho_{yx}C_{y}C_{x}, \\
m^{'}_{5} &= \left(f_{n} - f^{\prime}_{n}\right)\rho_{xz}C_{x}C_{z}, & m^{'}_{6} &= \left(f_{n} - f^{\prime}_{n}\right)C^{2}_{z}, \\
m^{'}_{7} &= \left(f_{n} - f^{\prime}_{n}\right)\rho_{yz}C_{y}C_{z} & & .
\end{aligned}$$

The minimum MSE's of the above stated estimators are given by

$$\min.MSE\left(t^{1}_{(BN)}\right) = MSE\left(\bar{y}^{*}\right) - \bar{Y}^{2}\left\{\frac{\left(2m_{3}m_{4}m_{5} - m_{2}m^{2}_{3} - m_{1}m^{2}_{4}\right)}{\left(m^{2}_{5} - m_{1}m_{2}\right)}\right\} \tag{6.2.41}$$

$$\min.MSE\left(t^{2}_{(BN)}\right) = MSE\left(\bar{y}^{*}\right) - \bar{Y}^{2}\left\{\frac{\left(2m_{3}m^{'}_{4}m_{5} - m^{\prime}_{2}m^{2}_{3} - m_{1}m^{\prime 2}_{4}\right)}{\left(m^{2}_{5} - m_{1}m^{\prime}_{2}\right)}\right\} \tag{6.2.42}$$

$$\min.MSE\left(t^{3}_{(BN)}\right) = MSE\left(\bar{y}^{*}\right) - \bar{Y}^{2}\left\{\frac{\left(2m_{7}m^{'}_{3}m^{'}_{5} - m^{\prime}_{1}m^{2}_{7} - m_{6}m^{\prime 2}_{3}\right)}{\left(m^{\prime 2}_{5} - m^{'}_{1}m_{6}\right)}\right\} \tag{6.2.43}$$

$$\min.MSE\left(t^{4}_{(BN)}\right) = MSE\left(\bar{y}^{*}\right) - \bar{Y}^{2}\left\{\frac{\left(2m^{'}_{3}m^{'}_{5}m^{'}_{7} - m^{\prime}_{1}m^{\prime 2}_{7} - m^{'}_{6}m^{\prime 2}_{3}\right)}{\left(m^{\prime 2}_{5} - m^{'}_{1}m^{'}_{6}\right)}\right\} \tag{6.2.44}$$

6.3 Proposed estimators

Let

$$t = \bar{y}^* \tag{6.3.1}$$

$$MSE\left(\bar{y}^*\right) = \bar{Y}^2\left\{f_n C_y^2 + g_n C_{y(2)}^2\right\} \tag{6.3.2}$$

Searls (1964) proposed a technique which is very useful to improve the efficiency of the any estimator.

Let us consider the Searls type estimators under non-response given by

$$T = \alpha t \tag{6.3.3}$$

so that

$$MSE\left(T\right) = \frac{\bar{Y}^2 MSE\left(t\right)}{\bar{Y}^2 + MSE\left(t\right)} \tag{6.3.4}$$

It can be easily observed that the minimum MSE of T is always lesser than t.

On the lines of the Singh and Kumar (2010b), Shabbir and Khan (2013) and Bhushan and Naqvi (2014) by using the Searls type transformation, we propose the following efficient estimators using information based on two auxiliary variables for estimating the population mean of the study variable in presence of non-response under the strategies given below.

Strategy I: In this strategy, we assume that non-response occurs on the study variable as well as on the auxiliary variable and the population mean $\bar{X}$ of first auxiliary variable x is unknown however the population mean $\bar{Z}$ of the second auxiliary variable z is known. The proposed estimator for estimating the population mean is

$$T_{(BP)}^1 = \gamma_1\bar{y}^* + \theta_1\left(\bar{x}^* - \bar{x}^{'}\right) + \omega_1\left(\bar{z}^* - \bar{Z}\right) \tag{6.3.5}$$

Strategy II: In this strategy, we assume that non-response occurs on the study variable as well as on the auxiliary variable and the population mean $\bar{X}$ of first auxiliary variable x and the population mean $\bar{Z}$ of the second auxiliary variable z both are unknown. The proposed estimator for estimating the population mean is given by

$$T^2_{(BP)} = \gamma_2 \bar{y}^* + \theta_2 \left(\bar{x}^* - \bar{x}^{'}\right) + \omega_2 \left(\bar{z}^* - \bar{z}^{'}\right) \tag{6.3.6}$$

Strategy III: In this strategy, we assume that non-response occurs on the study variable only and the population mean $\bar{X}$ of first auxiliary variable x is unknown however the population mean $\bar{Z}$ of the second auxiliary variable z is known. The proposed estimator for estimating the population mean is

$$T^3_{(BP)} = \gamma_3 \bar{y}^* + \theta_3 \left(\bar{x} - \bar{x}^{'}\right) + \omega_3 \left(\bar{z} - \bar{Z}\right) \tag{6.3.7}$$

Strategy IV: In this strategy, we assume that non-response occurs on the study variable only and the population mean of both the auxiliary variables are unknown. The proposed estimator under this strategy is

$$T^4_{(BP)} = \gamma_4 \bar{y}^* + \theta_4 \left(\bar{x} - \bar{x}^{'}\right) + \omega_4 \left(\bar{z} - \bar{z}^{'}\right) \tag{6.3.8}$$

Theorem 6.3.1. *The bias and minimum mean square error of the proposed estimators T^k_{BP}, $k = 1, 2, 3, 4$ are given by*

$$Bias\left(T^k_{(BP)}\right) = \left(\gamma_k - 1\right)\bar{Y} \tag{6.3.9}$$

$$\min .MSE\left(T^k_{(BP)}\right) = \frac{\bar{Y}^2 \min .MSE\left(t^k_{(BN)}\right)}{\bar{Y}^2 + \min .MSE\left(t^k_{(BN)}\right)} \tag{6.3.10}$$

Proof. Now, for the bias amd MSE of $T^1_{(BP)}$, we have

$T^1_{BP} = \gamma_1 \bar{Y}\left(1 + \varepsilon^*_0\right) + \theta_1 \left\{\bar{X}\left(1 + \varepsilon^*_1\right) - \bar{X}\left(1 + \varepsilon^{\prime}_1\right)\right\} + \omega_1 \left\{\bar{Z}\left(1 + \varepsilon^*_2\right) - \bar{Z}\right\}$

Thus,

$T^1_{(BP)} - \bar{Y} = \left(\gamma_1 - 1\right)\bar{Y} + \gamma_1 \bar{Y}\varepsilon^*_0 + \theta_1 \bar{X}\left(\varepsilon^*_1 - \varepsilon^{\prime}_1\right) + \omega_1 \bar{Z}\varepsilon^*_2$

for bias taking expectation on both sides, we get

$Bias\left(T^1_{(BP)}\right) = \left(\gamma_1 - 1\right)\bar{Y}$

for MSE, squaring and taking expectation on both sides of the above equation, we get

$MSE\left(T^1_{(BP)}\right) = (\gamma_1-1)^2\bar{Y}^2 + \gamma_1^2\bar{Y}^2\left\{f_nC_y^2 + g_nC^2_{y(2)}\right\} + \theta_1^2\bar{X}^2\left\{\left(f_n - f_n^{'}\right)C_x^2 + g_nC^2_{x(2)}\right\} + \omega_1^2\bar{Z}^2\left\{f_nC_z^2 + g_nC^2_{z(2)}\right\} + 2\gamma_1\theta_1\bar{X}\bar{Y}\left\{\left(f_n - f_n^{'}\right)\rho_{yx}C_yC_x + g_n\rho_{yx(2)}C_{y(2)}C_{x(2)}\right\} + 2\gamma_1\omega_1 \bar{Y}\bar{Z}\left\{f_n\rho_{yz}C_yC_z + g_n\rho_{yz(2)}C_{y(2)}C_{z(2)}\right\} + 2\theta_1\omega_1\bar{X}\bar{Z}\left\{\left(f_n - f_n^{'}\right)\rho_{xz}C_xC_z + g_n\rho_{xz(2)}C_{x(2)}C_{z(2)}\right\}$
$MSE\left(T^1_{(BP)}\right) = \bar{Y}^2 + \gamma_1^2\bar{Y}^2 + \gamma_1^2Var\left(\bar{y}^*\right) - 2\gamma_1\bar{Y}^2 + \theta_1^2\bar{X}^2m_1 + \omega_1^2\bar{Z}^2m_2 + 2\gamma_1\theta_1\bar{X}\bar{Y}m_3 + 2\gamma_1\omega_1\bar{Y}\bar{Z}m_4 + 2\theta_1\omega_1\bar{X}\bar{Z}m_5$

Differentiating $MSE\left(T^1_{(BP)}\right)$, for optimum value of the γ_1, θ_1 and ω_1, partially with respect to γ_1, θ_1, ω_1 and equating to zero. we get the optimum value of γ_1, θ_1 and ω_1 are given by

$$\frac{\partial MSE\left(T^1_{(BP)}\right)}{\partial\gamma_1} = 0$$

$$\frac{\partial MSE\left(T^1_{(BP)}\right)}{\partial\theta_1} = 0$$

$$\frac{\partial MSE\left(T^1_{(BP)}\right)}{\partial\omega_1} = 0$$

so that

$$\gamma_1 = \frac{\bar{Y}^2 - \theta_1\bar{X}\bar{Y}m_3 - \omega_1\bar{Y}\bar{Z}m_4}{\bar{Y}^2 + A}$$

$$\theta_1 = \frac{-\gamma_1\bar{Y}m_3 - \omega_1\bar{Z}m_5}{\bar{X}m_1}$$

$$\omega_1 = \frac{-\gamma_1\bar{Y}m_4 - \theta_1\bar{X}m_5}{\bar{Z}m_2}$$

solving these equations, we get

$$\gamma_{1opt} = \frac{\bar{Y}^2}{\bar{Y}^2 + A - \bar{Y}^2\left\{\dfrac{m_2m_3^2 + m_1m_4^2 - 2m_3m_4m_5}{m_1m_2 - m_5^2}\right\}}$$

$$\theta_{1opt} = \gamma_{1opt}\left(\frac{\bar{Y}}{\bar{X}}\right)\left\{\frac{m_4m_5 - m_2m_3}{m_1m_2 - m_5^2}\right\}$$

$$\omega_{1opt} = \gamma_{1opt}\left(\frac{\bar{Y}}{\bar{Z}}\right)\left\{\frac{m_3m_5 - m_1m_4}{m_1m_2 - m_5^2}\right\}$$

By using these optimum values of γ_1, θ_1 and ω_1 in $MSE\left(T^1_{(BP)}\right)$, we get the minimum mean square error as

$$\min .MSE\left(T^1_{(BP)}\right)=\frac{\bar{Y}^2\left[A-\bar{Y}^2\left\{\frac{m_2m_3^2+m_1m_4^2-2m_3m_4m_5}{m_1m_2-m_5^2}\right\}\right]}{\bar{Y}^2+\left[A-\bar{Y}^2\left\{\frac{m_2m_3^2+m_1m_4^2-2m_3m_4m_5}{m_1m_2-m_5^2}\right\}\right]}$$

where $A=MSE\left(\bar{y}^*\right)$

$$\min .MSE\left(T^1_{(BP)}\right)=\frac{\bar{Y}^2\min .MSE\left(t^1_{(BN)}\right)}{\bar{Y}^2+\min .MSE\left(t^1_{(BN)}\right)}$$

The derivation of $MSE\left(T^i_{(BP)}\right)$, $i=1,2,3.$ can be easily done on similar lines so that, the optimum values of characterizing scalars are

$$\gamma_{2opt}=\frac{\bar{Y}^2}{\bar{Y}^2+A-\bar{Y}^2\left\{\frac{m_2'm_3^2+m_1m_4'^2-2m_3m_4'm_5}{m_1m_2'-m_5^2}\right\}}$$

$$\theta_{2opt}=\gamma_{2opt}\left(\frac{\bar{Y}}{\bar{X}}\right)\left\{\frac{m_4'm_5-m_2'm_3}{m_1m_2'-m_5^2}\right\}$$

$$\omega_{2opt}=\gamma_{2opt}\left(\frac{\bar{Y}}{\bar{Z}}\right)\left\{\frac{m_3m_5-m_1m_4'}{m_1m_2'-m_5^2}\right\}$$

$$\gamma_{3opt}=\frac{\bar{Y}^2}{\bar{Y}^2+A-\bar{Y}^2\left\{\frac{m_1'm_7^2+m_6m_3'^2-2m_7m_3'm_5'}{m_1'm_6-m_5'^2}\right\}}$$

$$\theta_{3opt}=\gamma_{3opt}\left(\frac{\bar{Y}}{\bar{X}}\right)\left\{\frac{m_5'm_7-m_3'm_6}{m_1'm_6-m_5'^2}\right\}$$ ■

$$\omega_{3opt}=\gamma_{3opt}\left(\frac{\bar{Y}}{\bar{Z}}\right)\left\{\frac{m_3'm_5-m_1'm_7}{m_1'm_6-m_5'^2}\right\}$$

$$\gamma_{4opt}=\frac{\bar{Y}^2}{\bar{Y}^2+A--\bar{Y}^2\left\{\frac{m_1'm_7'^2+m_6'm_3'^2-2m_3'm_5'm_7'}{m_1'm_6'-m_5'^2}\right\}}$$

$$\theta_{4opt}=\gamma_{4opt}\left(\frac{\bar{Y}}{\bar{X}}\right)\left\{\frac{m_5'm_7'-m_3'm_6'}{m_1'm_6'-m_5'^2}\right\}$$

$$\omega_{4opt}=\gamma_{4opt}\left(\frac{\bar{Y}}{\bar{Z}}\right)\left\{\frac{m_3'm_5'-m_1'm_7'}{m_1'm_6'-m_5'^2}\right\}$$

6.4 An empirical study

We have conducted an empirical study on 4 population described below along with their key parameters.

1. Data considered from the book "Advanced sampling theory with applications" by Sarjinder Singh (p.1114).

   Fish caught: Estimated number of fish caught by marine. Descriptive parameters of the population-

   Y=Estimated number of fish caught during 1995

   X=Estimated number of fish caught during 1993

   Z=Estimated number of fish caught during 1992

   N=69, n'=55, n=20, $\bar{Y} = 4514.899$, $\bar{Z} = 4230.174$, $\bar{X} = 4591.072$, $C_y = 1.3509$, $C_z = 1.3164$, $C_x = 1.3755$, $C_{y(2)} = 1.3850$, $C_{z(2)} = 1.3640$, $C_{x(2)} = 1.4160$, $\rho_{yx} = 0.9564$, $\rho_{yz} = 0.9538$, $\rho_{xz} = 0.9632$, $\rho_{yx(2)} = 0.9674$, $\rho_{yz(2)} = 0.9668$, $\rho_{xz(2)} = 0.9699$.

2. The present data belong to the data on physical growth og upper socio-economic group of 95 school going children of Varansi under an ICMR study, Department of Pediatrics, BHU during 1983-1984 has been taken under study, (Khare and Sinha (2007)). The first 25% (i.e. 24 children) units have been considered as non-response units. The value of parameters related to the study character y (the weight of children in kg) and the auxiliary character x (the height of the children in cm) and additional auxiliary chracter z (chest circumference of the children in cm)n have been given as follows:

   $\bar{Y} = 19.4968$, $\bar{Z} = 55.8611$, $\bar{X} = 51.1726$, $C_y = 0.15613$, $C_z = 0.05860$, $C_x = 0.03006$, $C_{y(2)} = 0.12075$, $C_{z(2)} = 0.05402$, $C_{x(2)} = 0.02478$, $\rho_{yx} = 0.328$, $\rho_{yz} = 0.846$, $\rho_{xz} = 0.297$, $\rho_{yx(2)} = 0.477$, $\rho_{yz(2)} = 0.729$, $\rho_{xz(2)} = 0.570$.

The percentage relative efficiency (PRE) of the proposed estimator is calculated by

$$PRE = \frac{Var(\bar{y}^*)}{min.MSE(T_{BP_i})} \times 100$$

Table 6.1: MSE and PRE of the estimators

Estimators	1/2	1/3	1/4
$\bar{y}^*$	1915872(100)	2510898(100)	3105923(100)
$\bar{y}_{R(1)}$	282681.3(677.749)	322643.4(778.227)	362605.6(856.557)
$\bar{y}_{R(2)}$	837745(228.694)	1432771(175.248)	2027797(153.167)
$\bar{y}_{Reg(1)}$	276385.4(693.188)	314578.8(798.177)	352772.3(880.433)
$\bar{y}_{Reg(2)}$	833217.8(229.936)	1428243.7(175.803)	2023269.5(153.510)
Strategy I			
$\bar{y}^1_{(SK)}$	296847.3(845.855)	335450.7(925.896)	0.1239(231.661)
$\bar{y}^5_{(SK)}$	141773.2(1351.363)	173677.5(1445.724)	205166.8(1513.853)
$t^1_{(BN)}$	141773.2(1351.363)	173677.5(1445.724)	205166.8(1513.853)
$T^1_{(BP)}$	140794(**1360.762**)	172210.2(**1458.042**)	203122.4(**1529.09**)
Strategy II			
$\bar{y}^2_{(SK)}$	274526.2(697.882)	311980.6(804.825)	349432.1(888.849)
$\bar{y}^6_{(SK)}$	3812295(50.255)	3964322(63.337)	4071188(76.290)
$t^2_{(BN)}$	250802.9(763.895)	280813(894.1529)	310819.9(999.268)
$T^2_{(BP)}$	247754.6(**773.294**)	276997.1(**906.470**)	306151.7(**1014.505**)
Strategy III			
$\bar{y}^3_{(SK)}$	811473.8(236.098)	1406500(178.521)	2001526(155.178)
$\bar{y}^7_{(SK)}$	704277.1(272.0338)	1299303(193.2496)	1894329(163.959)
$t^3_{(BN)}$	704277.1(272.0338)	1299303(193.2496)	1894329(163.959)
$T^3_{(BP)}$	680757(**281.4325**)	1221448(**205.5674**)	173256(**179.196**)
Strategy IV			
$\bar{y}^4_{(SK)}$	831960.3(230.284)	1426986(175.958)	2022012(153.606)
$\bar{y}^8_{(SK)}$	815811.2(234.843)	1410837(177.972)	2005863(154.842)
$t^4_{(BN)}$	815811.2(234.843)	1410837(177.972)	2005863(154.842)
$T^4_{(BP)}$	784417.6(**244.241**)	1319511(**190.29**)	1826164(**170.079**)

Table 6.2: MSE and PRE of the estimators

Estimators	1/2	1/3	1/4
$\bar{y}^*$	0.207214(100)	0.24722(100)	0.287225(100)
$\bar{y}_{R(1)}$	0.1893(109.489)	0.2231(110.805)	0.2569(104.289)
$\bar{y}_{R(2)}$	0.1954(106.045)	0.2354(105.018)	0.2754(104.289)
$\bar{y}_{Reg(1)}$	0.1845 (112.210)	0.2161(114.415)	0.2476(115.992)
$\bar{y}_{Reg(2)}$	0.1930(107.380)	0.2330(106.113)	0.2730(105.217)
Strategy I			
$\bar{y}^1_{(SK)}$	0.0789(262.596)	0.1019(242.622)	0.1239(231.661)
$\bar{y}^5_{(SK)}$	0.0679(304.874)	0.0880(280.615)	0.1076(266.708)
$t^1_{(BN)}$	0.0679(304.874)	0.0881(280.615)	0.1018(266.708)
$T^1_{(BP)}$	0.679(**304.929**)	0.0888(**280.681**)	0.1077(**266.784**)
Strategy II			
$\bar{y}^2_{(SK)}$	0.1155(179.405)	0.1364(181.270)	0.1571(182.857)
$\bar{y}^6_{(SK)}$	0.1014(204.173)	0.1230(200.841)	0.1437(199.851)
$t^2_{(BN)}$	0.0926(223.688)	0.1124(219.768)	0.1318(217.844)
$T^2_{(BP)}$	0.0926(**223.742**)	0.1125(**219.832**)	0.1318(**217.919**)
Strategy III			
$\bar{y}^3_{(SK)}$	0.0941(220.152)	0.1341(184.315)	0.1741(164.944)
$\bar{y}^7_{(SK)}$	0.0867(238.995)	0.1267(195.110)	0.1667(172.286)
$t^3_{(BN)}$	0.0519(399.098)	0.0919(268.933)	0.1319(217.708)
$T^3_{(BP)}$	0.0519(**399.153**)	0.0919(**268.998**)	0.1319(**217.783**)
Strategy IV			
$\bar{y}^4_{(SK)}$	0.1118(174.432)	0.1588(155.681)	0.1988(144.476)
$\bar{y}^8_{(SK)}$	0.1116(185.646)	0.1516(163.049)	0.1916(149.886)
$t^4_{(BN)}$	0.1118(185.647)	0.1516(163.049)	0.1917(149.887)
$T^4_{(BP)}$	0.1116(**185.701**)	0.1517(**163.114**)	0.1915(**149.962**)

6.5 Conclusion

This chapter shows that the proposed estimators using the Searls (1964) philosophy in presence of bivariate auxiliary information perform better than the estimators suggested by Singh and Kumar (2010b), Shabbir and Khan (2013) and Bhushan and Naqvi (2015) in terms of percentage relative efficiency (PRE). This fact has been also supported through an empirical study. Further, the proposed estimators are important owing to the fact that they provide an improvement over the regression estimators, which are BLUEs. The result of this chapter is quite illuminating, both theoretically and empirically.

Chapter 7

Improved estimators for estimating population mean presence of random non-response

7.1 Introduction

The negative impact of non-response of the estimators of population mean is well known. Numerous authors have suggested a various estimators for estimating population mean under the random non-responses for various situations. The suitable use of the auxiliary information is known to benefit the estimation procedure. The estimation of population mean using auxiliary information has drawn the attention of various important authors and some of them are given below. Srivastava and Jhajj (1981) suggested a generalized class of estimators of the population mean in survey sampling using auxiliary information which incorporated many important estimators as its special case.

The problem of estimating the mean of a finite population using auxiliary information has been discussed by several authors, including, Singh and Tracy (2001) where they suggested an estimator of population mean in presence of random non-response. Similarly, under the problem of random non-response some authors suggested that by using the auxiliary information, we can improve the efficiency of the estimators, see Singh and Joarder

(1998) suggested an estimator of finite population variance using random non-response in survey sampling, Singh et al. (2000) proposed a regression type estimators for random non-response in survey sampling, Singh and Singh (1979, 1985) suggested an estimators by using double sampling and unequal sampling under random non-response, Singh et al. (2012) and Singh et al. (2007) suggested a families of estimators for maen ,ratio and product of finite population and finite population variance under random non-response and Singh and Tracy (2001) suggested an estimator of population mean in presence of random non-response.

7.2 Distribution of random non-response, notations and expectations

Let $\Omega = (1, 2, ..., N)$ denote a population of N units from which a simple random sample of size n is drawn without replacement. If $r\,(r = 0, 1, 2, ..., (n-2))$ denotes the number of sampling units on which information could not be obtained due to a random non-response, then the remaining $(n-r)$ units can be treated as a simple random sample from U. It is assumed that r is less than $(n-1)$ i.e. $0 \leq r \leq (n-2)$. Singh and Joarder (1998) assumed that r has the following discrete distribution as

$$P\left(r\right) = \frac{n-r}{(nq+2p)}\binom{n-2}{r}p^{r}q^{n-2-r} \tag{7.2.1}$$

where p is the probability of non-response, $q = 1-p$ and $\binom{n-2}{r}$ represents that total number ways to obtain r non-response out of a possible $(n-2)$. We define

For the variate y_i and x_i

$\bar{X} = N^{-1}\sum\limits_{i=1}^{N} x_i$: Population mean of the ith variate x_i.

$\bar{Y} = N^{-1}\sum\limits_{i=1}^{N} y_i$: Population mean of the ith variate y_i.

C_x, C_y: The population coefficient of variation (CV) of the variate y and x.

$\rho_{il} = S_{yx}/S_yS_x$: The population correlation coefficient between the variates y and x.

$(n-1)\,S_{yx} = \sum\limits_{i=1}^{N}\left(y_i - \bar{Y}_i\right)\left(x_i - \bar{X}\right)$;$K_{xy} = \rho_{xy}\frac{C_y}{C_x}$

$K = (K_{12})$,$d = (\lambda_{12}C_y + \lambda_{12}C_x)$

$\lambda_{w_1,w_2,} = \mu_{w_1,w_2,} \diagup \left(\mu_{20}\right)^{\frac{w_1}{2}} \left(\mu_{02}\right)^{\frac{w_2}{2}}$
$(N-1)\,\mu_{w_1,w_2} = \sum\limits_{i=1}^{N} \left(y_i - \bar{Y}\right)^{w_1} \left(x_i - \bar{X}\right)^{w_2}$
where w_1 and w_2 are non-negative integers

Define the following terms
$B = C_x^2 K^2 + \dfrac{\left(C_x K \lambda_{03} - d\right)^2}{\left(\lambda_{04} - \lambda_{03}^2 - 1\right)}$ or $B = \left(C_x^2 K^2 + \Delta B_1^2\right)$
where $B_1 = \dfrac{\Delta_2}{\Delta}$, $\Delta_2 = \left(C_x K \lambda_{03} - d\right)$ and $\Delta = \left(\lambda_{04} - \lambda_{03}^2 - 1\right)$
For the estimates based on the sample are given by:
$\bar{y} = n^{-1} \sum\limits_{i=1}^{n} y_i,\ \bar{y}^* = (n-r)^{-1} \sum\limits_{i=1}^{n-r} y_i,\ s_x^2 = (n-1)^{-1} \sum\limits_{i=1}^{n} \left(x_i - \bar{x}\right)^2,$
$s_x^{*2} = (n-1)^{-1} \sum\limits_{i=1}^{n-r} \left(x_i - \bar{x}^*\right)^2$ are conditionally unbiased estimators of S_x^2 respectively.
where $\theta^* = \left(\dfrac{1}{nq+2p} - \dfrac{1}{N}\right)$ and $\theta = \left(\dfrac{1}{n} - \dfrac{1}{N}\right)$
Singh and Joarder (1998) obtained the following maximum likelihood estimator of p (probability of non- response),as

$$\hat{p} = \frac{(n-1+r) - \sqrt{(n-1+r)^2 - \frac{4rn(n-3)}{(n-2)}}}{2\,(n-3)} \tag{7.2.2}$$

If $r = 0$ then $\hat{p} = 0$, and, if $r = n-2$ then $\hat{p} = 1$; thus $\hat{p}$ is an admissible estimators of response probability p.

Let us define

$\bar{y}^* = \bar{Y}\left(1+\epsilon_0\right),\ \bar{x}^* = \bar{X}\left(1+\epsilon_1\right),\ \bar{x} = \bar{X}\left(1+\epsilon_2\right),\ s_x^{*2} = S_x^2\left(1+\epsilon_3\right)$ and $s_x^2 = S_x^2\left(1+\epsilon_4\right)$

Then under the model

$E\left(\varepsilon_i\right) = 0, (i = 0, 1, ..., 4)$

$E\left(\varepsilon_0^2\right) = \theta^* C_y^2,\ E\left(\varepsilon_1^2\right) = \theta^* C_x^2,\ E\left(\varepsilon_2^2\right) = \theta C_x^2,\ E\left(\varepsilon_3^2\right) = \theta^*\left(\lambda_{04} - 1\right),\ E\left(\varepsilon_4^2\right) = \theta\left(\lambda_{04} - 1\right),$ $E\left(\varepsilon_0\varepsilon_1\right) = \theta^* \rho_{yx} C_y C_x,\ E\left(\varepsilon_0\varepsilon_2\right) = \theta \rho_{yx} C_y C_x,\ E\left(\varepsilon_0\varepsilon_3\right) = \theta^* \lambda_{12} C_y, E\left(\varepsilon_0\varepsilon_4\right) = \theta \lambda_{12} C_y,$ $E\left(\varepsilon_1\varepsilon_2\right) = \theta C_x^2, E\left(\varepsilon_1\varepsilon_3\right) = \theta^* \lambda_{03} C_x,\ E\left(\varepsilon_1\varepsilon_4\right) = \theta \lambda_{03} C_x,\ E\left(\varepsilon_2\varepsilon_3\right) = \theta \lambda_{03} C_x,\ E\left(\varepsilon_2\varepsilon_4\right) =$ $\theta \lambda_{03} C_x$ and $E\left(\varepsilon_3\varepsilon_4\right) = \theta\left(\lambda_{04} - 1\right)$

7.3 Exiting estimators

Srivastava and Jhajj (1981) suggested a class of estimators of the population mean is defined by

$$t_{rn} = \bar{y}g\left(u, v\right) \tag{7.3.1}$$

where $g\left(u, v\right)$ is a function of $\left(u, v\right)$, where $u = \frac{\bar{x}}{\bar{X}}$ and $v = \frac{s_x^2}{S_x^2}$, such that $g\left(1, 1\right) = 1$ and satisfying certain regularity conditions.

To the first degree of approximation the bias and minimum mean square error of t is respectively, given below

$$\begin{aligned} Bias\left(t_{rn}\right) &= \frac{\theta\bar{Y}}{2}\left[2K_{yx}C_x^2 g_1\left(1,1\right) + 2\lambda_{12}g_2\left(1,1\right) + C_x^2 g_{11}\left(1,1\right) + 2\lambda_{03}C_x g_{12}\left(1,1\right)\right. \\ &\quad \left. + \left(\lambda_{04} - 1\right)g_{22}\left(1,1\right)\right] & (7.3.2) \\ MSE\left(t_{rn}\right) &= \theta S_y^2\left[1 - \rho_{yx}^2 - \left(\lambda_{03}\rho_{yx} - \lambda_{12}\right)^2/\Delta\right] & (7.3.3) \end{aligned}$$

Similarly, a class of estimator for mean under random non-response model. Singh et al. (2007) suggest a family of estimator under three different strategies. Which provided almost all the estimators proposed till that date are their special cases.

Strategy I: When random non-response for r units on study variable y and auxiliary variable x are present in the sample, and the population mean $\bar{X}$ and the variance S_x^2 of x are known.

We define a family of estimators of $\bar{Y}$ as

$$t_{rn_1} = \bar{y}^* f\left(u^\bullet, v^\bullet\right) \tag{7.3.4}$$

where $f\left(u^\bullet, v^\bullet\right)$ is a function of $\left(u^\bullet, v^\bullet\right)$, where $u^\bullet = \frac{\bar{x}^*}{\bar{X}}$ and $v^\bullet = \frac{s_x^{*2}}{S_x^2}$, such that $f\left(1, 1\right) = 1$ and satisfying the following regularity conditions:

1. Whatever the sample $\left(u^\bullet, v^\bullet\right)$ assumes values in a bounded, closed convex subset, S, of the two dimenstional real space containing the point $\left(1, 1\right)$.

2. In S the function $f\left(u^\bullet, v^\bullet\right)$ is continuous and bounded.

3. The first and second order partial derivetives of $f\left(u^\bullet, v^\bullet\right)$ exist as well as are continuous

and bounded in S.

To the first degree of approximation minimum MSE of t_{rn_1} is given by

$$mim.MSE\left(t_{rn_1}\right) = \theta^* S_y^2\left[1 - \rho_{yx}^2 - \left(\lambda_{03}\rho_{yx} - \lambda_{12}\right)^2/\Delta\right] \tag{7.3.5}$$

where $\bar{y}^*$, $t_{rn_2} = \bar{y}^* f_1\left(u^\bullet\right)$ and $t_{rn_3} = \bar{y}^* f_2\left(v^\bullet\right)$ are the members of the above family and the mean square error of $\bar{y}^*$, $t_2$and t_3 are given by

$$Var\left(\bar{y}^*\right) = \theta^* S_y^2 \tag{7.3.6}$$

$$\min.MSE\left(t_{rn_2}\right) = \theta^* S_y^2\left[1 - \rho_{yx}^2\right] \tag{7.3.7}$$

$$\min.MSE\left(t_{rn_3}\right) = \theta^* S_y^2\left[1 - \lambda_{12}^2/\left(\lambda_{04} - 1\right)\right] \tag{7.3.8}$$

Strategy II: We consider the situation when information on variable y can not be obtained for r units, while the population mean $\bar{X}$ and the variance S_x^2 of the auxiliary variable x are known.

We consider the following family of estimator of $\bar{Y}$ given by

$$t_{rn_4} = \bar{y}^* \phi\left(u, v\right) \tag{7.3.9}$$

where $\phi\left(u, v\right)$ is a function of (u, v), where $u = \frac{\bar{x}}{\bar{X}}$ and $v = \frac{s_x^2}{S_x^2}$, such that $\phi\left(1, 1\right) = 1$ and satisfying certain regularity conditions, as defined similarly as in strategy I.

To the first degree of approximation MSE of t_{rn_4} is given by

$$mim.MSE\left(t_{rn_4}\right) = \theta S_y^2\left[1 - \rho_{yx}^2 - \left(\lambda_{03}\rho_{yx} - \lambda_{12}\right)^2/\Delta\right] + \left(\theta^* - \theta\right) S_y^2 \tag{7.3.10}$$

where $t_{rn_5} = \bar{y}^* \phi_1\left(u\right)$ and $t_{rn_6} = \bar{y}^* \phi_2\left(v\right)$ are the members of the above family and the minimum MSE of t_{rn_5} and t_{rn_6} are given by

$$\min.MSE\left(t_{rn_5}\right) = \theta S_y^2\left(1 - \rho_{yx}^2\right) + \left(\theta^* - \theta\right) S_y^2 \tag{7.3.11}$$

$$\min.MSE\left(t_{rn_6}\right)=\theta S_y^2\left[1-\lambda_{12}^2/\left(\lambda_{04}-1\right)\right]+\left(\theta^*-\theta\right)S_y^2 \tag{7.3.12}$$

Strategy III: We consider the situation when information on study variable y can not be obtained for r units while information on the auxiliary variable x is obtained for all the sample units. but the population mean $\bar{X}$ and the variance S_x^2 of the auxiliary character x are not known.

Under these circumstances, Singh et al. (2007) defines the following class of estimator of $\bar{Y}$ as

$$t_{rn_7}=\bar{y}^*\psi\left(u^*,v^*\right) \tag{7.3.13}$$

where $\psi\left(u^*,v^*\right)$ is a function of $\left(u^*,v^*\right)$, where $u=\frac{\bar{x}^*}{\bar{x}}$ and $v=\frac{s_x^{*2}}{s_x^2}$, such that $\psi\left(1,1\right)=1$ and satisfying the regularity conditions, as defined similarly as in strategy I.

To the first degree of approximation MSE of t_{rn_7} is given by

$$\min.MSE\left(t_{rn_7}\right)=S_y^2\left[\theta^*-\left(\theta^*-\theta\right)\left\{\rho_{yx}^2+\left(\lambda_{03}\rho_{yx}-\lambda_{12}\right)^2/\Delta\right\}\right] \tag{7.3.14}$$

where $t_{rn_8}=\bar{y}^*\psi_1\left(u^*\right)$ and $t_{rn_9}=\bar{y}^*\psi_2\left(v^*\right)$ are the members of the above family and the minimum MSE of $t_{rn_8}$and t_{rn_9} are defined as below

$$\min.MSE\left(t_{rn_8}\right)=S_y^2\left[\theta^*-\left(\theta^*-\theta\right)\rho_{yx}^2\right] \tag{7.3.15}$$

$$\min.MSE\left(t_{rn_9}\right)=S_y^2\left[\theta^*-\left(\theta^*-\theta\right)\lambda_{12}^2/\left(\lambda_{04}-1\right)\right] \tag{7.3.16}$$

The minimum MSE of the conventional estimators t_{rn_i} $(i=1,2,...,9)$ can be written as

$$\min.MSE\left(T_{rn_i}\right)=\bar{Y}^2\left\{1-\left(1-\frac{\min.MSE\left(t_{rn_i}\right)}{\bar{Y}^2}\right)\right\} \tag{7.3.17}$$

7.4 Proposed estimators

In this section, we propose the difference and ratio type estimators by using Searls type transformation (STT) under different strategies. The Searls type transformation can be defined as

$$T = \alpha\bar{y}^* \tag{7.4.1}$$

where α is a suitable chosen constant.

The mean square error of the above estimators is given by

$$\min.MSE\left(T\right) = \frac{\bar{Y}^2 MSE\left(\bar{y}^*\right)}{\bar{Y}^2 + MSE\left(\bar{y}^*\right)} \tag{7.4.2}$$

In above equation (7.4.2) shown that the STD estimators are better than conventional estimators. The proposed difference and ratio type estimators parallel to strategies given in the previous section using Searls philosophy are defined below:

Strategy I

$$T_{rn_1} = \alpha_1\bar{y}^* + d_1\left(\bar{x}^* - \bar{X}\right) \tag{7.4.3}$$

$$T_{rn_2} = \alpha_2\bar{y}^* + d_2\left(s_x^{*2} - S_x^2\right) \tag{7.4.4}$$

$$T_{rn_3} = \alpha_3\bar{y}^* + d_3\left(\bar{x}^* - \bar{X}\right) + k_1\left(s_x^{*2} - S_x^2\right) \tag{7.4.5}$$

$$T_{m_1} = \gamma_1\bar{y}^*\left(\frac{\bar{X}}{\bar{x}^*}\right)^{\beta_1} \tag{7.4.6}$$

$$T_{m_2} = \gamma_2\bar{y}^*\left(\frac{S_x^2}{s_x^{*2}}\right)^{\beta_2} \tag{7.4.7}$$

$$T_{m_3} = \gamma_3\bar{y}^*\left(\frac{\bar{X}}{\bar{x}^*}\right)^{\beta_3}\left(\frac{S_x^2}{s_x^{*2}}\right)^{\eta_1} \tag{7.4.8}$$

Strategy II

$$T_{rn_4} = \alpha_4\bar{y}^* + d_4\left(\bar{x} - \bar{X}\right) \tag{7.4.9}$$

$$T_{rn_5} = \alpha_5\bar{y}^* + d_5\left(s_x^2 - S_x^2\right) \tag{7.4.10}$$

$$T_{rn_6} = \alpha_6\bar{y}^* + d_6\left(\bar{x} - \bar{X}\right) + k_2\left(s_x^2 - S_x^2\right) \tag{7.4.11}$$

$$T_{m_4} = \gamma_4\bar{y}^*\left(\frac{\bar{X}}{\bar{x}}\right)^{\beta_4} \tag{7.4.12}$$

$$T_{m_5} = \gamma_5\bar{y}^*\left(\frac{S_x^2}{s_x^2}\right)^{\beta_5} \tag{7.4.13}$$

$$T_{m_6} = \gamma_6 \bar{y}^* \left(\frac{\bar{X}}{\bar{x}}\right)^{\beta_6} \left(\frac{S_x^2}{s_x^2}\right)^{\eta_2} \tag{7.4.14}$$

Strategy III

$$T_{rn_7} = \alpha_7 \bar{y}^* + d_7 \left(\bar{x}^* - \bar{x}\right) \tag{7.4.15}$$

$$T_{rn_8} = \alpha_8 \bar{y}^* + d_8 \left(s_x^{*2} - s_x^2\right) \tag{7.4.16}$$

$$T_{rn_9} = \alpha_9 \bar{y}^* + d_9 \left(\bar{x}^* - \bar{x}\right) + k_3 \left(s_x^{*2} - s_x^2\right) \tag{7.4.17}$$

$$T_{m_7} = \gamma_7 \bar{y}^* \left(\frac{\bar{x}}{\bar{x}^*}\right)^{\beta_7} \tag{7.4.18}$$

$$T_{m_8} = \gamma_8 \bar{y}^* \left(\frac{s_x^2}{s_x^{*2}}\right)^{\beta_8} \tag{7.4.19}$$

$$T_{m_9} = \gamma_9 \bar{y}^* \left(\frac{\bar{x}}{\bar{x}^*}\right)^{\beta_9} \left(\frac{s_x^2}{s_x^{*2}}\right)^{\eta_3} \tag{7.4.20}$$

The biases and MSE's of the proposed estimators are given in subsequent theorems.

Theorem 7.4.1. *The bias and minimum MSE of the proposed ratio estimator T_{r_j}, $(j = 1, 2, ..., 9)$ are given by*

$$Bias_{\alpha_j}\left(T_{m_j}\right) = \overline{Y}\left(\gamma_j - 1\right) + \alpha_j Bias_1\left(t_{m_j}\right) \tag{7.4.21}$$

and

$$\min MSE\left(T_{m_j}\right) = \overline{Y}^2 \left(1 - \frac{B_j^2}{A_j}\right) \tag{7.4.22}$$

where $Bias_{\gamma_j}\left(T_{m_j}\right)$ is the first order bias with parameter γ_j and $Bias_1\left(t_{m_j}\right)$ is the first order bias of the conventional estimators counterpart with $\gamma_j = 1$.

Proof. The MSE of T_{m_j} $(j = 1, 2, ..., 9)$ is given by

$MSE(T_{m_j}) = \overline{Y}^2 \left[1 + \gamma_j^2 A_j - 2\gamma_j B_j\right]$

The optimum values of scalars involved are tabulated below for ready reference:

$\gamma_{jopt} = \dfrac{B_j}{A_j}$; $(j = 1, 2, ...9)$

substituting the optimum value of γ_j in $MSE\left(T_{m_j}\right)$ we get minimum MSE

$MSE\left(T_{r_j}\right) = \overline{Y}^2 \left(1 - \dfrac{B_j^2}{A_j}\right)$

where

$$A_1 = \left[1 + \theta^* C_y^2 + 2\beta_1^2 \theta^* C_x^2 + \beta_1 \theta^* \left(C_x^2 - 4\rho_{yx} C_y C_x\right)\right]$$

$$B_1 = \left[1 + \frac{\beta_1^2}{2} \theta^* C_x^2 + \frac{\beta_1}{2} \theta^* \left(C_x^2 - 2\rho_{yx} C_y C_x\right)\right]$$

$$A_2 = \left[1 + \theta^* C_y^2 + 2\beta_2^2 \theta^* \left(\lambda_{04} - 1\right) + \beta_2 \theta^* \left\{\left(\lambda_{04} - 1\right) - 4\lambda_{12} C_y\right\}\right]$$

$$B_2 = \left[1 + \frac{\beta_2^2}{2} \theta^* \left(\lambda_{04} - 1\right) + \frac{\beta_2}{2} \theta^* \left\{\left(\lambda_{04} - 1\right) - 2\lambda_{12} C_y\right\}\right]$$

$$\begin{aligned} A_3 = & \left[1 + \theta^* C_y^2 + 2\beta_3^2 \theta^* C_x^2 + 2\beta_4^2 \theta^* \left(\lambda_{04} - 1\right) + \beta_3 \theta^* \left(C_x^2 - 4\rho_{yx} C_y C_x\right)\right. \\ & \left. + \beta_4 \theta^* \left\{\left(\lambda_{04} - 1\right) - 4\lambda_{12} C_y\right\} + 4\beta_3 \beta_4 \theta^* \lambda_{03} C_x\right] \end{aligned}$$

$$\begin{aligned} B_3 = & \left[1 + \frac{\beta_3^2}{2} \theta^* C_x^2 + \frac{\beta_4^2}{2} \theta^* \left(\lambda_{04} - 1\right) + \frac{\beta_3}{2} \theta^* \left(C_x^2 - 2\rho_{yx} C_y C_x\right)\right. \\ & \left. + \frac{\beta_4}{2} \theta^* \left\{\left(\lambda_{04} - 1\right) - 2\lambda_{12} C_y\right\} + \beta_3 \beta_4 \theta^* \lambda_{03} C_x\right] \end{aligned}$$

$$A_4 = \left[1 + \theta^* C_y^2 + 2\beta_5^2 \theta C_x^2 + \beta_5 \theta \left(C_x^2 - 4\rho_{yx} C_y C_x\right)\right]$$

$$B_4 = \left[1 + \tfrac{\beta_5^2}{2} \theta C_x^2 + \tfrac{\beta_5}{2} \theta \left(C_x^2 - 2\rho_{yx} C_y C_x\right)\right]$$

$$A_5 = \left[1 + \theta^* C_y^2 + 2\beta_6^2 \theta \left(\lambda_{04} - 1\right) + \beta_6 \theta \left\{\left(\lambda_{04} - 1\right) - 4\lambda_{12} C_y\right\}\right]$$

$$B_5 = \left[1 + \tfrac{\beta_6^2}{2} \theta^* \left(\lambda_{04} - 1\right) + \tfrac{\beta_6}{2} \theta^* \left\{\left(\lambda_{04} - 1\right) - 2\lambda_{12} C_y\right\}\right]$$

$$\begin{aligned} A_6 = & \left[1 + \theta^* C_y^2 + 2\beta_7^2 \theta C_x^2 + 2\beta_8^2 \theta \left(\lambda_{04} - 1\right) + \beta_7 \theta \left(C_x^2 - 4\rho_{yx} C_y C_x\right)\right. \\ & \left. + \beta_8 \theta \left\{\left(\lambda_{04} - 1\right) - 4\lambda_{12} C_y\right\} + 4\beta_7 \beta_8 \theta \lambda_{03} C_x\right] \end{aligned}$$

$$\begin{aligned} B_6 = & \left[1 + \frac{\beta_7^2}{2} \theta C_x^2 + \frac{\beta_8^2}{2} \theta \left(\lambda_{04} - 1\right) + \frac{\beta_7}{2} \theta \left(C_x^2 - 2\rho_{yx} C_y C_x\right)\right. \\ & \left. + \frac{\beta_8}{2} \theta \left\{\left(\lambda_{04} - 1\right) - 2\lambda_{12} C_y\right\} + \beta_7 \beta_8 \theta \lambda_{03} C_x\right] \end{aligned}$$

$$A_7 = \left[1 + \theta^* C_y^2 + \left(\theta^* - \theta\right) \left\{2\beta_9^2 C_x^2 + \beta_9 \left(C_x^2 - 4\rho_{yx} C_y C_x\right)\right\}\right]$$

$$B_7=\left[1+(\theta^*-\theta)\left\{\frac{\beta_9^2}{2}C_x^2+\frac{\beta_9}{2}\left(C_x^2-2\rho_{yx}C_yC_x\right)\right\}\right]$$

$$A_8=\left[1+\theta^*C_y^2+(\theta^*-\theta)\left[2\beta_{10}^2\left(\lambda_{04}-1\right)+\beta_{10}\left\{\left(\lambda_{04}-1\right)-4\lambda_{12}C_y\right\}\right]\right]$$

$$B_8=\left[1+(\theta^*-\theta)\left[\frac{\beta_{10}^2}{2}\left(\lambda_{04}-1\right)+\frac{\beta_{10}}{2}\left\{\left(\lambda_{04}-1\right)-2\lambda_{12}C_y\right\}\right]\right]$$

■

$$A_9=\left[1+\theta^*C_y^2+(\theta^*-\theta)\left[2\beta_{11}^2C_x^2+2\beta_{12}^2\left(\lambda_{04}-1\right)+\beta_{11}\left(C_x^2-4\rho_{yx}C_yC_x\right)\right.\right.$$
$$\left.\left.+\beta_{12}\left\{\left(\lambda_{04}-1\right)-4\lambda_{12}C_y\right\}+4\beta_{11}\beta_{12}\lambda_{03}C_x\right]\right]$$

$$B_9=\left[1+(\theta^*-\theta)\left[\frac{\beta_{11}^2}{2}C_x^2+\frac{\beta_{12}^2}{2}\left(\lambda_{04}-1\right)+\frac{\beta_{11}}{2}\left(C_x^2-2\rho_{yx}C_yC_x\right)\right.\right.$$
$$\left.\left.+\frac{\beta_{12}}{2}\left\{\left(\lambda_{04}-1\right)-2\lambda_{12}C_y\right\}+\beta_{11}\beta_{12}\lambda_{03}C_x\right]\right]$$

It is noteworthy that the simultaneous optimization w.r.t. γ_j, β_j and η_l of the expression of MSE is not possible and we use optimum value of $\beta_j=\beta_{jopt}$ when $\gamma_j=1$ and use this within $\gamma_j=\gamma_{jopt}$ to obtain .

Theorem 7.4.2. *The bias and minimum MSE of the proposed difference estimator T_{d_i} is given by*

$$Bias\left(T_{rn_i}\right)=\overline{Y}\left(\alpha_i-1\right) \tag{7.4.23}$$

and

$$\min MSE_{\alpha_i}\left(T_{rn_i}\right)=\frac{\overline{Y}^2 MSE_1\left(t_i\right)}{\overline{Y}^2+MSE_1\left(t_i\right)} \tag{7.4.24}$$

where $MSE_{\alpha_i}\left(T_{rn_i}\right)$ is the first order MSE with parameter α_i and $MSE_1\left(t_{rn_i}\right)$ is the first order MSE of the conventional estimators counterpart with $\alpha_i=1$.

Proof. The MSE of T_{rn_1} is given by

$$MSE\left(T_{rn_1}\right)=\frac{\bar{Y}^2\theta^*S_y^2\left(1-\rho_{yx}^2\right)}{\left[\bar{Y}^2+\theta^*S_y^2\left(1-\rho_{yx}^2\right)\right]}$$
$$MSE\left(T_{rn_1}\right)=\frac{\bar{Y}^2\left[MSE\left(t_{rn1}^{(1)}\right)\right]}{\left[\bar{Y}^2+MSE\left(t_{rn_1}^{(1)}\right)\right]}$$

The optimum values of scalars involved are given below:

$$\alpha_1 = \frac{\bar{Y}^2}{\left[\bar{Y}^2 + \theta^* S_y^2 \left(1 - \rho_{yx}^2\right)\right]},\ d_1 = -\alpha \rho_{yx} \frac{S_y}{S_x}.$$

$$\alpha_2 = \frac{\bar{Y}^2}{\left[\bar{Y}^2 + \theta^* S_y^2 \left(1 - \lambda_{12}^2 / \left(\lambda_{04} - 1\right)\right)\right]},\ d_2 = -\alpha_2 \frac{\lambda_{12} S_y}{S_x^2 \left(\lambda_{04} - 1\right)}.$$

$$\alpha_3 = \frac{\bar{Y}^2}{\left[\bar{Y}^2 + \theta^* S_y^2 \left\{\left(1 - \rho_{yx}^2\right) - \left(\lambda_{03}\rho_{yx} - \lambda_{12}\right)^2 / \left((\lambda_{04} - 1 - \lambda_{03}^2\right)\right\}\right]},$$

$$d_3 = -\alpha_3 \frac{S_y}{S_x} \left\{\rho_{yx} + \frac{\lambda_{03}\left(\lambda_{03}\rho_{yx} - \lambda_{12}\right)^2}{\left(\lambda_{04} - 1 - \lambda_{03}^2\right)}\right\},\ k_1 = -\alpha_3 \frac{S_y}{S_x^2} \left\{\frac{\left(\lambda_{03}\rho_{yx} - \lambda_{12}\right)}{\left(\lambda_{04} - 1 - \lambda_{03}^2\right)}\right\}.$$

$$\alpha_4 = \frac{\bar{Y}^2}{\left[\bar{Y}^2 + S_y^2 \left(\theta^* - \theta \rho_{yx}^2\right)\right]},\ d_4 = -\alpha_4 \rho_{yx} \frac{S_y}{S_x}.$$

$$\alpha_5 = \frac{\bar{Y}^2}{\left[\bar{Y}^2 + S_y^2 \left(\theta^* - \theta \frac{\lambda_{12}^2}{\left(\lambda_{04} - 1\right)}\right)\right]},\ d_5 = -\alpha_5 \frac{\lambda_{12} S_y}{S_x^2 \left(\lambda_{04} - 1\right)}.$$

$$\alpha_6 = \frac{\bar{Y}^2}{\left[\bar{Y}^2 + S_y^2 \left\{\theta^* - \theta \left(\rho_{yx}^2 + \frac{\left(\lambda_{03}\rho_{yx} - \lambda_{12}\right)^2}{\left((\lambda_{04} - 1 - \lambda_{03}^2\right)}\right)\right\}\right]},\ d_6 = -\alpha_6 \frac{S_y}{S_x} \left\{\rho_{yx} + \frac{\lambda_{03}\left(\lambda_{03}\rho_{yx} - \lambda_{12}\right)^2}{\left(\lambda_{04} - 1 - \lambda_{03}^2\right)}\right\},$$

$$k_2 = -\alpha_6 \frac{S_y}{S_x^2} \left\{\frac{\left(\lambda_{03}\rho_{yx} - \lambda_{12}\right)}{\left(\lambda_{04} - 1 - \lambda_{03}^2\right)}\right\}$$

$$\alpha_7 = \frac{\bar{Y}^2}{\left[\bar{Y}^2 + \theta^* S_y^2 - \left(\theta^* - \theta\right) S_y^2 \rho_{yx}^2\right]},\ d_7 = -\alpha_7 \rho_{yx} \frac{S_y}{S_x}$$

$$\alpha_8 = \frac{\bar{Y}^2}{\left[\bar{Y}^2 + \theta^* S_y^2 - \left(\theta^* - \theta\right) \frac{\lambda_{12}^2 S_y^2}{\left(\lambda_{04} - 1\right)}\right]},\ d_8 = -\alpha_8 \frac{\lambda_{12} S_y}{S_x^2 \left(\lambda_{04} - 1\right)}$$

$$\alpha_9 = \frac{\bar{Y}^2}{\left[\bar{Y}^2 + S_y^2 \left\{\theta^* - \left(\theta^* - \theta\right) \left(\rho_{yx}^2 + \frac{\left(\lambda_{03}\rho_{yx} - \lambda_{12}\right)^2}{\left((\lambda_{04} - 1 - \lambda_{03}^2\right)}\right)\right\}\right]}$$

$$d_9 = -\alpha_9 \frac{S_y}{S_x} \left\{ \rho_{yx} + \frac{\lambda_{03} \left(\lambda_{03} \rho_{yx} - \lambda_{12} \right)^2}{\left(\lambda_{04} - 1 - \lambda_{03}^2 \right)} \right\}$$

■

$$k_3 = -\alpha_9 \frac{S_y}{S_x^2} \left\{ \frac{\left(\lambda_{03} \rho_{yx} - \lambda_{12} \right)}{\left(\lambda_{04} - 1 - \lambda_{03}^2 \right)} \right\}$$

Corollary 7.4.3. *The proposed difference estimators for population mean T_{rn_i} have always lesser MSE than the conventional estimators for population mean t_{rn_i} $(i = 1, 2, ...9)$, when $\alpha_i = 1$.*

Proof. It may be easily observed from comparision of (7.4.22) with (7.4.24). It is interesting to note that simultaneous optimization w.r.t. the characterizing scalars α_i ,d_i and k_j of the expression of MSE is possible for proposed difference estimator but not proposed ratio estimators. ■

Theorem 7.4.4. *The proposed difference estimation method is better than proposed ratio estimation methods $(i = 1, 2, ..., 9)$ iff*

$$\alpha_{iopt} > \frac{B_j^2}{A_j} \tag{7.4.25}$$

and vice versa. Otherwise both are equally efficient in case of equality in (7.4.25).

Proof. It may be easily observed from (7.4.24) that the MSE of proposed difference estimators T_{rn_i} is given by

$$\min MSE \left(T_{rn_i} \right) = \overline{Y}^2 \left(1 - \alpha_{iopt} \right) \tag{7.4.26}$$

Comparing (7.4.26)with (7.4.22), we have the theorem. ■

Theorem 7.4.5. *The proposed ratio estimation method is better than the conventional estimators t_{m_i} , $(i = j = 1, 2, ..., 9)$ iff*

$$\frac{B_j^2}{A_j} > \left\{ 1 - \frac{\min .MSE \left(t_i \right)}{\bar{Y}^2} \right\} \tag{7.4.27}$$

and vice versa. Otherwise both are equally efficient in case of equality in (7.4.27)

Proof. It may be easily observed from (7.4.22) that the MSE of proposed ratio estimators T_{m_j} is given by

$$\min MSE\left(T_{m_j}\right) = \overline{Y}^2\left(1 - \frac{B_j^2}{A_j}\right) \tag{7.4.28}$$

Comparing (7.4.28)with (7.3.17), we have the theorem. ■

7.5 Empirical study

Suppose that a bank selected a simple random sample of twenty states without replacement from the USA during 1997 and collected information (in thousands) on real (y) and nonreal estate farm loans (x). The selected states are CA, CT, FL, IL, ME, MS, MO, NE, NJ, NM, ND, OK, SC, TN, TX, UT, VA, WA, WV, and WI. For details of the data set, please see population-1 on page 1111 in singh (2003). However, assume the information on the real estate farm loans was not available for four states ME, ND, TX and VA. Parameters of the population is given below

$Y = 27771.73,\ X = 43908.12\ ,\bar{Y} = 555.43,\ \bar{X} = 878.16,\ S_x^2 = 1176526,\ S_y^2 = 342021.5,\ C_x^2 = 1.5256,\ C_y^2 = 1.1086,\ \lambda_{03} = 1.5936,\ S_{xy} = 509910.41,\ \rho_{xy} = 0.8038,\ N = 50,\ \beta = 0.4334, \lambda_{12} = 1.0982,\ \lambda_{21} = 0.9387,\ \lambda_{04} = 4.5247,\ \lambda_{22} = 2.8411.$

The percentage relative efficiency (PRE) of the proposed estimator are calculated with respect to $\bar{y}^*$ given by

$$PRE = \frac{Var(\bar{y}^*)}{min.MSE(T_{d_i}, T_{r_i})} \times 100$$

Table 7.1: MSE and PRE of the estimators

Estimators	MSE(PRE)	Proposed estimators	MSE(PRE)
$\bar{y}^*$	14811.97(100)	T	14133.39(104.80)
Strategy I			
t_{rn_2}	5242.04(282.56)	T_{rn_1}	5154.45(287.36)
t_{rn_3}	9743.77(152.01)	T_{rn_2}	9445.44(156.81)
t_{rn_1}	4739.97(312.49)	T_{rn_3}	4668.24(**317.29**)
Strategy II			
t_{rn_5}	8182.62(181.02)	T_{rn_4}	7971.19(185.81)
t_{rn_6}	11301.09(131.07)	T_{rn_5}	10901.74(135.86)
t_{rn_4}	7834.83(189.05)	T_{rn_6}	7640.78(**193.85**)
Strategy III			
t_{rn_8}	11871.38(124.77)	T_{rn_7}	11431.49(129.57)
t_{rn_9}	13254.64(111.75)	T_{rn_8}	12708.62(116.55)
t_{rn_6}	11717.11(126.41)	T_{rn_9}	11288.37(**131.21**)

Table 7.2: MSE and PRE of the estimators

Estimators	MSE(PRE)	Proposed estimators	MSE(PRE)
$\bar{y}^*$	14811.97(100)	T	14133.39(104.80)
Strategy I			
t_{rn_2}	5242.04(282.56)	T_{m_1}	5067.80(292.27)
t_{rn_3}	9743.77(152.01)	T_{m_2}	9064.70(163.40)
t_{rn_1}	4739.97(312.49)	T_{m_3}	4670.49(**317.14**)
Strategy II			
t_{rn_5}	8182.62(181.02)	T_{m_4}	7887.66(187.78)
t_{rn_6}	11301.09(131.07)	T_{m_5}	10621.99(139.44)
t_{rn_4}	7834.83(189.05)	T_{m_6}	7643.21(**193.79**)
Strategy III			
t_{rn_8}	11871.38(124.77)	T_{m_7}	11381.57(130.14)
t_{rn_9}	13254.64(111.75)	T_{m_8}	12575.49(117.78)
t_{rn_6}	11717.11(126.41)	T_{m_9}	11289.67(**131.19**)

7.6 Conclusion

In this chapter, we have concluded that the proposed estimators always perform better than the estimators proposed by Singh et al. (2012) under random non-response. The most important conclusion of this paper is that the traditional thought, that the ratio type estimator can at best match upto its regression counterpart, is vitiated. We have not only proved that the proposed regression provides an improvement over the traditional regression counterpart but also proved that the proposed ratio estimator can provide an improvement over both traditional regression estimator and proposed regression estimator provided their respective efficiency conditions (7.4.25) and (7.4.23) are satisfied. The computational result are given in table 7.1 also support this fact.

Chapter 8

Improved estimators for estimation of population variance presence of random non-response

8.1 Introduction

In estimation of population characteristics ancillary information closely related to the main characteristics plays a very important role. The parameters can be estimated more accurately by making use of information on an auxiliary variable that is correlated with study variable. The ratio method, regression method and their generalizations are good examples in this context. In this paper we have considered the problem of estimation of population variances. Srivastava and Jhajj (1980) suggested a class of estimators using auxiliary information for estimating finite population variance. Isaki (1983) proposed variance estimation using auxiliary information in sample survey then after Prasad and Singh (1990) proposed some improved ratio-type estimators of finite ppulation variance in sample surveys and Prasad and Singh (1992) also suggested an unbiased estimators of finite population variance using auxiliary information in sample surveys. Biradar and Singh (1994) proposed an alternative to ratio estimator of population variance, Cebrian and Garcia (1997) suggested the variance estimation using auxiliary information and Garcia and Cebrian (1996) proposed a repeated substitution method.

In survey sampling non-response is a major problem encountered by the practicioners. We know that the problem of non-response was first studied by Hansen and Hurvitz (1946) in relation to a mail survey. The non-response adversely affected the estimate of the population characteristics in survey sampling. Many authors dealt with the problem of non-response andsuggested different type of methods to estimate the population characteristics under non-response.

Tracy and Osahan (1994) proposed random non-response on study variable versus on study as well as auxiliary varibles. The problem of random non-response was considered by Singh and Joarder (1998) and they suggested that by using the auxiliary information we can improve the efficiency of the estimators of finite population variance. Singh et al. (2000) proposed a regression type estimators for random non-response in survey sampling, Ahmed et al. (2005) suggested the estimation of finite population variance in presence of random non-response using auxiliary variables and Singh et al. (2012) suggested a families of estimators for finite population variance under random non-response.

8.2 Distribution of random non-response, notations and expectations

Let $\Omega = (1, 2, ..., N)$ denote a population of N units from which a simple random sample of size n is drawn without replacement. If r $(r = 0, 1, 2, ..., (n-2))$ denotes the number of sampling units on which information could not be obtained due to a random non-response, then the remaining $(n-r)$ units can be treated as a simple random sample from U. It is assumed that r is less than $(n-1)$ i.e. $0 \leq r \leq (n-2)$. Singh and Joarder (1998) assumed that r has the following discrete distribution as

$$P(r) = \frac{n-r}{(nq+2p)} \binom{n-2}{r} p^r q^{n-2-r} \tag{8.2.1}$$

where p is the probability of non-response, $q = 1-p$ and $\binom{n-2}{r}$ represents that total number ways to obtain r non-response out of a possible $(n-2)$. They studied the effect of random non-response on the study and auxiliary variables of several estimators of variance. We define

$s_y^{*2} = (n-r-1)^{-1} \sum_{i=1}^{n-r} (y_i - \bar{y}^*)^2$ and $s_x^{*2} = (n-r-1)^{-1} \sum_{i=1}^{n-r} (x_i - \bar{x}^*)^2$ are conditionally unbiased estimators of

$S_y^{*2} = (N-1)^{-1} \sum_{i=1}^{N} \left(y_i - \bar{Y}\right)^2$ and $S_x^{*2} = (N-1)^{-1} \sum_{i=1}^{N} \left(x_i - \bar{X}\right)^2$, respectively, where $\bar{y}^* = (n-r)^{-1} \sum_{i=1}^{n-r} y_i$ and $\bar{x}^* = (n-r)^{-1} \sum_{i=1}^{n-r} x_i$.

Also, define $\mu_{ls} = (N-1)^{-1} \sum_{i=1}^{N} \left(y_i - \bar{Y}\right)^l \left(x_i - \bar{X}\right)^s$, $\lambda_{ls} = \frac{\mu_{ls}}{(\mu_{20})^{l/2} \left(\hat{\mu}_{02}^*\right)^{s/2}}$, $\theta^* = \left(\frac{1}{nq+2p} - \frac{1}{N}\right)$, $\theta = \left(\frac{1}{n} - \frac{1}{N}\right)$ and $\delta^* = \left(\frac{1}{nq+2p} - \frac{1}{n}\right)$

$\bar{X} = N^{-1} \sum_{i=1}^{N} x_i$: Population mean of the ith variate x_i

$\bar{Y} = N^{-1} \sum_{i=1}^{N} y_i$: Population mean of the ith variate y_i

Singh and Joarder (1998) obtained the following maximum likelihood estimator of p, λ_{ls} and μ_{ls} as

$$\hat{p} = \frac{(n-1+r) - \sqrt{(n-1+r)^2 - \frac{4rn(n-3)}{(n-2)}}}{2(n-3)} \tag{8.2.2}$$

$$\hat{\mu}_{ls}^* = (n-1+r)^{-1} \sum_{i=1}^{N} (y_i - \bar{y}^*)^l (x_i - \bar{x}^*)^s \tag{8.2.3}$$

and

$$\hat{\lambda}_{ls}^* = \frac{\hat{\mu}_{ls}^*}{\left(\hat{\mu}_{20}^*\right)^{l/2} \left(\hat{\mu}_{02}^*\right)^{s/2}} \tag{8.2.4}$$

If $r = 0$ then $\hat{p} = 0$ and, if $r = n-2$ then, $\hat{p} = 1$. Thus $\hat{p}$ is an admissible estimators of response probability p.

We have studied the effect of random non-response on the study and auxiliary variables of several estimators of variance under the following three strategies considered by Singh and Joarder (1998):

Strategy I: We are considering the situation when random non-response exists on both the study variable y and the auxiliary variable x and population variance S_x^2 ofthe auxiliary character is known.

Strategy II: Then we are considering the situation when information on variable y could not be obtained for r units while information on variable x is available andpopulation variance S_x^2 of the auxiliary variable is known.

Strategy III: Here we consider the situation when information on variable y could not be obtained for r units while information on the variable x is obtainedfor all the sample units, but the population variance S_x^2 of the auxiliary variable is unknown.

For each strategy, we have proposed a number of estimators and derived approximate bias and mean squared error. It is interesting that under this distribution of random non-response the exact bias and mean squared error expressions, up to first order approximation, exist for the proposed strategies.

Let us define

$s_y^{*2} = S_y^2 (1 + \epsilon_0)$, $s_x^{*2} = S_x^2 (1 + \epsilon_1)$ and $s_x^2 = S_x^2 (1 + \epsilon_2)$

Then, under the model using Singh and Joarder (1998) response, we have

$E(\varepsilon_i) = 0, (i = 0, 1, 2)$ $E(\varepsilon_0^2) = \theta^* (\lambda_{40} - 1)$, $E(\varepsilon_1^2) = \theta^* (\lambda_{04} - 1)$, $E(\varepsilon_2^2) = \theta (\lambda_{04} - 1)$, $E(\varepsilon_0 \varepsilon_1) = \theta^* (\lambda_{22} - 1)$, $E(\varepsilon_0 \varepsilon_2) = \theta (\lambda_{22} - 1)$ and $E(\varepsilon_1 \varepsilon_2) = \theta (\lambda_{04} - 1)$.

8.3 Existing estimators

Strategy I

Under this strategy, Singh and Joarder (1998) suggested the following estimator of finite population variance.

$$t_1^{'} = s_y^{*^2} \left(\frac{S_x^2}{s_x^{*2}} \right) \tag{8.3.1}$$

For this strategy, Ahmed et al. (2005) proposed the following estimators for finite population variance

$$t_{v_1} = s_y^{*^2} \left(\frac{S_x^2}{s_x^{*2}} \right)^{\alpha_1} \tag{8.3.2}$$

$$t_{v_2} = s_y^{*^2} + k_1 \left(S_x^2 - s_x^{*2} \right) \tag{8.3.3}$$

$$t_{v_3} = \frac{s_y^{*^2} S_x^2}{\theta_1 s_x^{*2} + (1 - \theta_1) S_x^2} \tag{8.3.4}$$

where α_1, k_1 and θ_1 are suitably chosen constants.

Theorem 8.3.1. *The Mean squared error of $t_t^{'}$ is given by*

$$MSE\left(t_1^{'}\right) = \theta^* S_y^4 \left(\lambda_{40} + \lambda_{04} - 2\lambda_{22}\right) \tag{8.3.5}$$

Theorem 8.3.2. *The minimum MSE of the estimators t_{v_i} $(i = 1, 2, 3)$ are equal and given by*

$$\min .MSE\left(t_{v_1}\right) = \min .MSE\left(t_{v_2}\right) = \min .MSE\left(t_{v_3}\right) = \theta^* S_y^4 \left[\left(\lambda_{40} - 1\right) - \frac{\left(\lambda_{22-1}\right)^2}{\left(\lambda_{04} - 1\right)}\right] \tag{8.3.6}$$

Strategy II

Under this strategy, Singh and Joarder (1998) proposed the following estimator of finite population variance given by

$$t_2^{'} = s_y^{*^2} \left(\frac{S_x^2}{s_x^2}\right) \tag{8.3.7}$$

For this case, Ahmed et al. (2005) proposed the following estimators for finite population variance given by

$$t_{v_4} = s_y^{*^2} \left(\frac{S_x^2}{s_x^2}\right)^{\alpha_2} \tag{8.3.8}$$

$$t_{v_5} = s_y^{*^2} + k_2 \left(S_x^2 - s_x^2\right) \tag{8.3.9}$$

$$t_{v_6} = \frac{s_y^{*^2} S_x^2}{\theta_2 s_x^2 + \left(1 - \theta_2\right) S_x^2} \tag{8.3.10}$$

where α_2, k_2 and θ_2 are suitably chosen constants.

Theorem 8.3.3. *The Mean squared error of $t_2^{'}$ is given by*

$$MSE\left(t_2^{'}\right) = S_y^4 \left[\theta^* \left(\lambda_{40} - 1\right) + \theta \left(\lambda_{04} - \lambda_{22}\right)\right] \tag{8.3.11}$$

Theorem 8.3.4. *The minimum MSE of the estimators t_{v_j} $(j = 4, 5, 6)$ are equal and given* *by*

$$\min .MSE\left(t_{v_4}\right)=\min .MSE\left(t_{v_5}\right)=\min .MSE\left(t_{v_6}\right)=S_y^4\left[\theta^*\left(\lambda_{40}-1\right)-\theta\left\{\frac{\left(\lambda_{22-1}\right)^2}{\left(\lambda_{04}-1\right)}\right\}\right] \tag{8.3.12}$$

Strategy III

Under strategy III, Singh and Joarder (1998) proposed the following estimator of finite population variance

$$t_3^{'}=s_y^{*^2}\left(\frac{s_x^2}{s_x^{*2}}\right) \tag{8.3.13}$$

For this strategy, Ahmed et al. (2005) proposed the following estimators for finite population variance given by

$$t_{v_7}=s_y^{*^2}\left(\frac{s_x^2}{s_x^{*2}}\right)^{\alpha_1} \tag{8.3.14}$$

$$t_{v_8}=s_y^{*^2}+k_3\left(s_x^2-s_x^{*2}\right) \tag{8.3.15}$$

$$t_{v_9}=\frac{s_y^{*^2}s_x^2}{\theta_3 s_x^{*2}+\left(1-\theta_3\right)s_x^2} \tag{8.3.16}$$

Theorem 8.3.5. *The mean squared error of $t_3^{'}$ is given by*

$$MSE\left(t_3^{'}\right)=S_y^4\left[\theta^*\left(\lambda_{40}-1\right)+\left(\theta^*-\theta\right)\left\{\left(\lambda_{04}-1\right)+2\left(\lambda_{22}-1\right)\right\}\right] \tag{8.3.17}$$

Theorem 8.3.6. *The minimum MSE of the estimators t_{v_i} $(i=7,8,9)$ are equal and given by*

$$\min .MSE\left(t_{v_7}\right)=\min .MSE\left(t_{v_8}\right)=\min .MSE\left(t_{v_9}\right)=S_y^4\left[\theta^*\left(\lambda_{40}-1\right)-\delta^*\left\{\frac{\left(\lambda_{22-1}\right)^2}{\left(\lambda_{04}-1\right)}\right\}\right] \tag{8.3.18}$$

8.4 Proposed estimators

The proposed difference and ratio type estimators under three different strategies, as mention above are given below.

Strategy I

$$T_{v_1} = \eta_1 s_y^{*^2} \left(\frac{S_x^2}{s_x^{*2}} \right)^{\beta_1} \tag{8.4.1}$$

$$T_{v_2} = \eta_2 s_y^{*^2} + k_1 \left(S_x^2 - s_x^{*2} \right) \tag{8.4.2}$$

$$T_{v_3} = \frac{\eta_3 s_y^{*^2} S_x^2}{\omega_1 s_x^{*2} + (1 - \omega_1) S_x^2} \tag{8.4.3}$$

Strategy II

$$T_{v_4} = \eta_4 s_y^{*^2} \left(\frac{S_x^2}{s_x^2} \right)^{\beta_2} \tag{8.4.4}$$

$$T_{v_5} = \eta_5 s_y^{*^2} + k_2 \left(S_x^2 - s_x^2 \right) \tag{8.4.5}$$

$$T_{v_6} = \frac{\eta_6 s_y^{*^2} S_x^2}{\omega_2 s_x^2 + (1 - \omega_2) S_x^2} \tag{8.4.6}$$

Strategy III

$$T_{v_7} = \eta_7 s_y^{*^2} \left(\frac{s_x^2}{s_x^{*2}} \right)^{\beta_3} \tag{8.4.7}$$

$$T_{v_8} = \eta_8 s_y^{*^2} + k_3 \left(s_x^2 - s_x^{*2} \right) \tag{8.4.8}$$

$$T_{v_9} = \frac{\eta_9 s_y^{*^2} s_x^2}{\omega_3 s_x^{*2} + (1 - \omega_3) s_x^2} \tag{8.4.9}$$

Theorem 8.4.1. *The bias and minimum MSE of the proposed ratio estimator T_{v_j} $(j = 1, 3, 4, 6, 7, 9)$ are given by*

$$Bias_{\eta_j} \left(T_{v_j} \right) = S_y^2 \left(\eta_j - 1 \right) + \eta_j Bias_1 \left(t_j \right) \tag{8.4.10}$$

and

$$\min MSE \left(T_{v_j} \right) = S_y^4 \left(1 - \frac{B_j^2}{A_j} \right) \tag{8.4.11}$$

where $Bias_{\eta_j} \left(T_{v_j} \right)$ is the first order bias with parameter η_j and $Bias_1 \left(t_{v_j} \right)$ is the first order *bias of the non-Searls counterpart with $\eta_j = 1$.*

Proof. The MSE of T_{v_j} $(j = 1, 3, 4, 6, 7, 9)$ is given by $MSE(T_{v_j}) = \bar{Y}^2\left[1 + \eta_j^2 A_j - 2\eta_j B_j\right]$
The optimum values of scalars involved are tabulated below for ready reference:
$\eta_{jopt} = \dfrac{B_j}{A_j}$; $(j = 1, 3, 4, 6, 7, 9)$
substituting the optimum value of η_j in $MSE\left(T_{v_j}\right)$ we get minimum MSE
$MSE\left(T_{v_j}\right) = S_y^4\left(1 - \dfrac{B_j^2}{A_j}\right)$
where

$$A_1 = \left[1 + \theta^* \left\{(\lambda_{40} - 1) + 2\beta_1^2(\lambda_{04} - 1) + \beta_1(\lambda_{04} - 1) - 4\beta_1(\lambda_{22} - 1)\right\}\right]$$

$$B_1 = \left[1 + \theta^* \left\{\tfrac{\beta_1^2}{2}(\lambda_{04} - 1) + \tfrac{\beta_1}{2}(\lambda_{04} - 1) - \beta_1(\lambda_{22} - 1)\right\}\right]$$

$$A_3 = \left[1 + \theta^* \left\{(\lambda_{40} - 1) + 2\omega_1^2(\lambda_{04} - 1) + \omega_1(\lambda_{04} - 1) - 4\omega_1(\lambda_{22} - 1)\right\}\right]$$

$$B_3 = \left[1 + \theta^* \left\{\tfrac{\omega_1^2}{2}(\lambda_{04} - 1) + \tfrac{\omega_1}{2}(\lambda_{04} - 1) - \omega_1(\lambda_{22} - 1)\right\}\right]$$

$$A_4 = \left[1 + \theta^*(\lambda_{40} - 1) + \theta\left\{2\beta_2^2(\lambda_{04} - 1) + \beta_2(\lambda_{04} - 1) - 4\beta_2(\lambda_{22} - 1)\right\}\right]$$

$$B_4 = \left[1 + \theta\left\{\tfrac{\beta_2^2}{2}(\lambda_{04} - 1) + \tfrac{\beta_2}{2}(\lambda_{04} - 1) - \beta_2(\lambda_{22} - 1)\right\}\right]$$

$$A_6 = \left[1 + \theta^*(\lambda_{40} - 1) + \theta\left\{2\omega_2^2(\lambda_{04} - 1) + \omega_2(\lambda_{04} - 1) - 4\omega_2(\lambda_{22} - 1)\right\}\right]$$

$$B_6 = \left[1 + \theta\left\{\tfrac{\omega_2^2}{2}(\lambda_{04} - 1) + \tfrac{\omega_2}{2}(\lambda_{04} - 1) - \omega_2(\lambda_{22} - 1)\right\}\right]$$

$$A_7 = \left[1 + \theta^*(\lambda_{40} - 1) + \delta^*\left\{2\beta_3^2(\lambda_{04} - 1) + \beta_3(\lambda_{04} - 1) - 4\beta_3(\lambda_{22} - 1)\right\}\right]$$

$$B_7 = \left[1 + \delta^*\left\{\tfrac{\beta_3^2}{2}(\lambda_{04} - 1) + \tfrac{\beta_3}{2}(\lambda_{04} - 1) - \beta_3(\lambda_{22} - 1)\right\}\right]$$
$$A_9 = \left[1 + \theta^*(\lambda_{40} - 1) + \delta^*\left\{2\omega_3^2(\lambda_{04} - 1) + \omega_3(\lambda_{04} - 1) - 4\omega_3(\lambda_{22} - 1)\right\}\right]$$

■

$$B_9 = \left[1 + \delta^*\left\{\tfrac{\omega_3^2}{2}(\lambda_{04} - 1) + \tfrac{\omega_3}{2}(\lambda_{04} - 1) - \omega_3(\lambda_{22} - 1)\right\}\right]$$

It is noteworthy that the simultaneous optimization w.r.t, η_j,β_j and θ_l of the expression of MSE is not possible and we use optimum value of $\beta_j = \beta_{jopt}$ when $\eta_j = 1$ and use this within $\eta_j = \eta_{jopt}$ to obtain .

Theorem 8.4.2. *The bias and minimum MSE of the proposed difference estimator T_{v_i} is given by*

$$Bias\left(T_{v_i}\right) = S_y^2\left(\eta_i - 1\right) \tag{8.4.12}$$

and

$$\min MSE_{\eta_i}\left(T_{v_i}\right) = \frac{S_y^4 MSE_1\left(t_{v_i}\right)}{S_y^4 + MSE_1\left(t_{v_i}\right)} \tag{8.4.13}$$

where $MSE_{\eta_i}\left(T_{v_i}\right)$ is the first order MSE with parameter η_i and $MSE_1\left(t_{v_i}\right)$ is the first order MSE of the conventional estimators counterpart with $\eta_i = 1$.

Proof. The MSE of T_{v_2} is given by

$$MSE\left(T_{v_2}\right) = \frac{\theta^* S_y^4\left[\left(\lambda_{40}-1\right) - \frac{\left(\lambda_{22-1}\right)^2}{\left(\lambda_{04}-1\right)}\right]}{\left[1+\theta^*\left[\left(\lambda_{40}-1\right) - \frac{\left(\lambda_{22-1}\right)^2}{\left(\lambda_{04}-1\right)}\right]\right]}$$

$$MSE\left(T_{v_2}\right) = \frac{S_y^4\left[MSE\left(t_2\right)\right]}{\left[S_y^4 + MSE\left(t_2\right)\right]}$$

the optimum values of scalars involved are given below:

$$\eta_2 = \frac{1}{\left[1+\theta^*\left\{\left(\lambda_{40}-1\right) - \frac{\left(\lambda_{22-1}\right)^2}{\left(\lambda_{04}-1\right)}\right\}\right]},\ k_1 = \eta_2\frac{S_y^2\left(\lambda_{22}-1\right)}{S_x^2\left(\lambda_{04}-1\right)}.$$

$$\eta_5 = \frac{1}{\left[1+\theta^*\left(\lambda_{40}-1\right) - \theta\left\{\frac{\left(\lambda_{22-1}\right)^2}{\left(\lambda_{04}-1\right)}\right\}\right]},\ k_2 = \eta_5\frac{S_y^2\left(\lambda_{22}-1\right)}{S_x^2\left(\lambda_{04}-1\right)}.$$

∎

$$\eta_8 = \frac{1}{\left[1+\theta^*\left(\lambda_{40}-1\right) - \delta^*\left\{\frac{\left(\lambda_{22-1}\right)^2}{\left(\lambda_{04}-1\right)}\right\}\right]},\ k_3 = \eta_8\frac{S_y^2\left(\lambda_{22}-1\right)}{S_x^2\left(\lambda_{04}-1\right)}$$

Corollary 8.4.3. *The proposed difference estimators for population mean T_{v_i} have always lesser MSE than the conventional estimators for population mean t_{v_i} $(i = 2, 5, 8)$.*

Proof. It may be easily observed from comparision of (8.4.11) with (8.4.13). It is interesting to note that simultaneous optimization w.r.t. the characterizing scalars η_i ,β_i and θ_j of the expression of MSE is possible for proposed difference estimator but not proposed ratioestimators. ■

Theorem 8.4.4. *The proposed difference estimation method is better than proposed ratio estimation methods* $(i = 2, 5, 8)$ *iff*

$$\eta_{iopt} > \frac{B_j^2}{A_j} \tag{8.4.14}$$

and vice versa. Otherwise both are equally efficient in case of equality in (8.4.14).

Proof. It may be easily observed from (8.4.13) that the MSE of proposed difference estimators $T_{_i}$ is given by

$$\min MSE\left(T_{_i}\right) = S_y^4\left(1 - \eta_{iopt}\right) \tag{8.4.15}$$

Comparing (8.4.15)with (8.4.14), we have the theorem. ■

Theorem 8.4.5. *The proposed ratio estimation method is better than the conventional estimators* t_{v_i} , $(i = j = 1, 3, 4, 6, 7, 9)$ *iff*

$$\frac{B_j^2}{A_j} > \left\{1 - \frac{\min .MSE\left(t_i\right)}{S_y^4}\right\} \tag{8.4.16}$$

and vice versa. Otherwise both are equally efficient in case of equality in (8.4.16)

Proof. It may be easily observed from (8.4.11) that the MSE of proposed ratio estimators T_{v_j} is given by

$$\min MSE\left(T_{v_j}\right) = S_y^4\left(1 - \frac{B_j^2}{A_j}\right) \tag{8.4.17}$$

Comparing (8.4.17)with (8.3.18), we have the theorem. ■

8.5 Empirical study

For empirical study data are related to the total cultivated area during 1978-79 and the area under wheat in the two consecutive years 1979-80 or a sample of 16 villages selected in the Baghpat Tehsil of Meerut District (UP) for the survey on fertlizer practices. The following parametrs are given below

$N = 50,\ n = 20,\ r = 4,\ p = 0.23355,\ q = 0.76645,\ \theta^* = 0.043307,\ \theta = 0.03, \delta^* = 0.013307, \bar{x}^* = 1123.62,\ \bar{y}^* = 735.77,\ s_y^{*^2} = 432778.2,\ s_x^{*^2} = 1622949,\ \hat{\lambda}^*_{40} = 2.0781,\ \hat{\lambda}_{04} = 2.6128$ and $\hat{\lambda}^*_{22} = 1.1775, \beta_1 = \beta_2 = \beta_3 = \omega_1 = \omega_2 = \omega_3 = 0.110057044.$

The percentage relative efficiency (PRE) of the proposed estimator is calculated by

$$PRE = \frac{Var(t_i^{'})}{min.MSE(T_{v_j})} \times 100$$

The results obtained are reported in the table given below.

Table 8.1: MSE and PRE of the estimators

Strategy I		Strategy II		Strategy II	
Estimators	MSE(PRE)	Estimators	MSE(PRE)	Estimators	MSE(PRE)
$t_1^{'}$	18947012798 (100)	$t_2^{'}$	16809532992 (100)	$t_3^{'}$	13649093903 (100)
$t_{v_1} = t_{v_2} = t_{v_3}$	8586258928 (220.666)	$t_{v_4} = t_{v_5} = t_{v_6}$	8634946650 (194.668)	$t_{v_7} = t_{v_8} = t_{v_9}$	8696025021 (156.957)
T_{v_1}	8154458390 **(232.351)**	T_{v_4}	8216169296 **(204.590)**	T_{v_7}	8293333721 **(164.579)**
T_{v_2}	8209892648 (230.782)	T_{v_5}	8254394544 (203.643)	T_{v_8}	8310190563 (164.245)
T_{v_3}	8154458390 **(232.351)**	T_{v_6}	8216169296 **(204.590)**	T_{v_9}	8310190563 (164.245)

It can be easily observed that the proposed estimators $T_{v_j}, j = 1, 2, ..., 9$ always perform better than the conventional estimators $t_{v_i}, i = 1, 2, ..., 9$ under the respective strategy.

8.6 Conclusion

The most important conclusion of this paper is that the proposed estimators provide a considerable improvement over the conventional regression estimator, which is also the BLUE. Further, it can be concluded that the proposed regression (difference) type estimators T_{v_j}, $j = 2, 5, 8$ always perform better than the conventional estimators t_{v_i}, $i = 2, 5, 8$ as establish by corollary 8.4.3. The proposed ratio type estimators T_{v_j}, $j = 1, 3, 4, 6, 7, 9$ also perform better than the conventional estimators t_{v_i}, $i = 1, 3, 4, 6, 7, 9$ as established by theorem 8.4.5. The computational results are given in table 8.1, confirm this fact. Also, the proposed ratio type estimators T_{v_j}, $j = 1, 3, 4, 6, 7, 9$ perform better than proposed difference type estimators T_{v_j}, $j = 2, 5, 8$ under the optimality condition given by (8.4.14).

Bibliography

[1] Ahmed, M. S., Titi, O. A., Rawi, Z. A. and Dayyeh, W. A. (2005): Estimation of finite population variance in presence of random non-response using auxiliary variables. *Infr. Mang. Sci.*, 16(2), 73-82.

[2] Ahmed, M. S., Al-Titi, O., Al-Rawi, Z. and Abu-Dayyeh, W. (2006): Estimation of a population mean using different imputation methods. *Statistics in Transition*, 7(6): 1247-1264.

[3] Ahmed, M. S., Titi, O. A., Rawi, Z. A. and Dayyeh, W. A. (2009): Estimation of finite population mean using random non-response survey sampling. *J. Appl. Statist. Sci.*, 17(1), 1-12.

[4] Bhushan, S. and Naqvi, N.(2015): Generalized efficient classes of estimators in presence of non-response using two auxiliary variables. *Jour. Stat. Mang. Syst.*, 18(6), 573-602.

[5] Bhushan, S. and Pandey, A. P. (2016): Optimality of ratio type estimation methods for population mean in presence missing data. *Comm. Statist. Theory Methods*, DOI:10.1080/03610926.2016.1167906.

[6] Bhushan, S. and Pandey, A. P. (2016): Optimal imputation of missing data for estimation of population mean. *Jour. Stat. Mang. Syst.*, 19(6), 755-769.

[7] Biradar, R. S. and Singh, H. P. (1994): An alternative to ratio estimator of population variance. *Ass. Stats. Rev.*, 8(2), 18-33.

[8] Cebrian, A. A. and Garcia, M. R. (1997): Variance estimation using auxiliary information: an almost unbiased multivariate ratio estimator. *Metrika*, 45, 171-178.

[9] Cochran, W. G. (1977): Sampling techniques, 3rd edn. *John Wiley and Sons*, New York.

[10] Diana, G. and Tommasi, C. (2003): Optimal estimation for finite population mean in two phase sampling. *Stat. Meth. App.*, 12, 41-48.

[11] Diana, G. and Perri, P. F. (2007): Estimation of finite population mean using multi-auxiliary information. *Metron*, LXV, 99-112.

[12] Diana, G. and Perri, P. F. (2010): Improved estimators of the population mean for missing data. *Comm. Statist. Theory Methods*, 39, 3245-3251.

[13] Diana, G. and Perri, P. F. (2013): A class of estimators in two- phase sampling with subsampling the non-respondents. *Appl. Math. Comp.*, 219, 10033-10043.

[14] El-Badry, M. A. (1956): A sampling procedure for mailed questionnaires. *J. Am. Stat. Asso.*, 51, 209-227.

[15] Garcia, M. R. and Cebrian, A. A. (1996): Repeated substitution method: the ratio estimator for the population variance. *Metrika*, 43, 101-105.

[16] Gupta, S. and Shabbir, J. (2008): On improvement in estimating the population mean in simple random sampling. *J. Appl. Stat.*, 35(5), 559-566.

[17] Hansen, M. H. and Hurwitz W. N. (1946): The problem of non-response in sample surveys. *J. Am. Stat. Asso.*, 41, 517- 529.

[18] Heitjan, D. F. and Basu, S. (1996): Distinguishing 'Missing at Random' and 'Missing Completely at Random'. *The American Statistician*, 50, 207-213.

[19] Isaki, C. T. (1983): Variance estimation using auxiliary information, *J. Am. Stat. Asso.*, 7, 117-123.

[20] Kalton, G., Kasprzyk, D. and Santos, R. (1981): Issues of nonresponse and imputation in the Survey of Income and Program Participation. *Current Topics in Survey Sampling*, D. Krewski, R. Platek and J.N.K. Rao, eds., 455-480*Academic Press*, New York.

[21] Kalton, G. and Kasprzyk, D. (1982): Imputing for missing survey responses, *Proc. of the Am. Stat. Asso., Survey Research Methods Section*, 22-31.

[22] Kadilar, C. and Cingi, H. (2004): Rato estimators in simple random sampling. *Appl. Math. Computat.*, 151:893-902.

[23] Kadilar, C. and Cingi, H. (2008): Estimators for the population mean in the case missing data. *Comm. Statist. Theory Methods*, 37, 2226-2236.

[24] Khare, B. B. and Srivastava, S. (1993): Estimation of population mean uing auxiliary character in presence of non-response. *Nat. Acad. Sci. Lett.*, India 16, 111-114.

[25] Khare, B. B. and Srivastava, S. (1995): Study of conventional and alternative two-phase sampling ratio, product and regression estimators in presence of non-response. Proc. *Indian Nat. Sci. Acad.*, 65, 195-203.

[26] Khare, B. B. and Srivastava, S. (1997): Transformed ratio type estimators for the population mean in the presence of non-response. *Comm. Statist. Theory Methods*, 26, 1779-1791.

[27] Koyuncu, N. and Kadilar, C. (2010): On improvement in estimating population mean in stratified random sampling. *J. Appl. Stat.*, 37(6), 999-1013.

[28] Lee, H., Rancourt, E. and Sarndal, C. E. (1994): Experiments with Variance Estimation from Survey Data with Imputed Values. *Journal of Official Statistics*, 10, 231-243.

[29] Lee, H., Rancourt, E. and Sarndal, C. E. (1995): Variance estimation in the presence of imputed data for the generalized estimation system, *Proc. of the Am. Stat. Asso., Survey Research Methods Section*, 384-389.

[30] Murthy, M. N. (1977): Sampling: Theory and methods. *Statistical Publishing Society*, Calcutta.

[31] Mohamed, C., Sedory, S. A. and Singh, S. (2017): Imputation using higher order moments of an auxiliary variable. *Commun. Stat. Simul. Comput.*, DOI:10.1080/03610918.2016.1208235.

[32] Thakur, N. S., Yadav, K. and Pathak, S. (2012): Imputation using regression estimators for estimating population mean in two phase sampling. *J. Reliab. Statist. Studies*, 5(2), 21-31.

[33] Okafor, F. C. and Lee, H. (2000): Double sampling for ratio and regression estimation with sub sampling the non respondents. *Surv. Methodol*, 26(2), 183-188.

[34] Olkin, I. (1958): Multi-variate ratio estimation for finite population. *Biometrika*, 43, 154-163.

[35] Prasad, B. and Singh, H. P. (1990): Some improved ratio-type estimators of finite ppulation variance in sample surveys. *Comm. Statist. Theory methods*, 19(3), 1127-1139.

[36] Prasad, B. and Singh, H. P. (1992): Unbiased estimators of finite population variance using auxiliary information in sample surveys. *Comm. Statist. Theory methods*, 21(5), 1367-1376.

[37] Rao, P. S. R. S. (1983): Randomization Approach in Incomplete Data in Sample Surveys, Vol. 2, eds W.G. Madow, I. Olkin, and D.B. Rubin, pp. 33-44. *Academic Press*, New York.

[38] Rao, P. S. R. S. (1986): Ratio estimation with sub sampling the non-respondents. *Survey Methodology*, 12, 217-230.

[39] Rao, P. S. R. S. (1987): Ratio and regression estimates with sub sampling the non-respondents. *Int. Statist. Assoc. Meet, Tokyo,* Japan, Sep., 2-16.

[40] Reddy, V. N. (1978): A study on the use of prior knowledge on certain population parameters in estimation. *Sankhya*, C 40, 29–37.

[41] Rubin, R. B. (1976): Inference and missing data. *Biometrika*, 63(3), 581-592.

[42] Rueda, M. and Gonzalez, S. (2004): Missing data and auxiliary information in surveys. *Computat. Statist.*, 19, 551-567.

[43] Rueda, M., Gonzalez, S. and Arcos, A. (2005): Indirect methods of imputation of missing data based on available units. *Appl. Math. Computat.*, 64, 249-261.

[44] Sande, I. G. (1979): A personal view of hot deck approach to automatic edit and imputation. *Journal Imputation Procedure*, Survey methodology, 5, 238-246.

[45] Searls, D. T. (1964): The Utilization of a known Coefficient of Variation in the Estimation Procedure. *J. Am. Stat. Asso.*, 59, 1225-1226.

[46] Singh, S. and Singh, R. (1979): On random non- response in unequal sampling. *Sankhya*, 41, 27-37.

[47] Singh, S. and Singh, R. (1985): On random non- response in double sampling with difference estimator. *Comm. Statist. Simulation Comput.*, 14, 747-757.

[48] Singh, S. and Singh, R. (1991): Imputation methods in two-dimensional surveys. *In: Recent Advances in Agriculture Statistics Research*, New Delhi: Wiley Eastern Ltd.

[49] Singh, S. and Joarder, A. H. (1998): Estimation of finite population variance using random non-response in survey sampling, *Metrika*, 98, 241-249.

[50] Singh, S., Joarder, A. H. and Tracy, D. S. (2000): Regression type estimators for random non-response in survey sampling, *Statistica*, LX(1), 39-44.

[51] Singh, S. and Horn, S. (2000): Compromised imputation in survey sampling. *Metrika*, 51, 267-276.

[52] Singh, S., Joarder, A. H. and Tracy, D. S. (2000): Regression type estimators for random non response in survey sampling. *Statistica*, 60(1), 39-44.

[53] Singh, H. P. and Tracy, D. S. (2001): Estimation of population mean in presence of random non-response in sample surveys. *Staistika*, 61, 231-248.

[54] Singh, S. (2003): Advanced sampling theory with applications. *Kluwer Academic Publishers*, 41, 27-37.

[55] Singh, S. and Deo, B. (2003): Imputation by power transformation. *Statistical Papers*, 44, 555-579.

[56] Singh, H. P., Chandra, P., Joarder, A. H. and Singh, S. (2007): Families of estimators of mean, ratio and product of a finite population using random non-response. *Test*, 16, 565-597.

[57] Singh, H.P and Kumar, S.(2008): A regression approach to the estimation of the finite population mean in the presence of non-response. *Aust. N. Z. J. Stat.*, 50(4), 395-408.

[58] Singh, S. (2009): A new method of imputation in survey sampling. *Statistics* , 43(5), 499-511.

[59] Singh, H. P. and Kumar, S. (2009a): A general class of estimators of the population mean in survey sampling using auxiliary information with sub sampling the non-respondents. *The Korean J. Appl. Statist.*, 22, 387-402.

[60] Singh, H.P and Kumar, S. (2010): Estimation of mean in presence of non-response using two phase sampling scheme. *Stat. Papers*, 51, 559-582.

[61] Singh, H. P., Kumar, S. and Kozak, M. (2010): Improved estimation of finite population mean using sub sampling to deal with non-response in two phase sampling scheme. *Comm. Statist. Theory Methods*, 39, 791-802.

[62] Singh, H. P. and Kumar, S. (2010b): Improved estimation of population mean under the two phase sampling with sub sampling the non-respondents. *J. Stat. Plan. Infer.*, 140, 2536-2550.

[63] Singh, H. P. and Solanki, R. S. (2011): Estimation of finite population mean using random non-response survey sampling. *Pak. J. Statist. Oper. Res.*, 7, 21-41.

[64] Singh, H. P., Tailor, R., Kim, J. M. and Singh, S. (2012): Families of estimators of finite population variance using random non-response in survey sampling. *The Kor. J. App. Stat.*, 25(4), 681-695.

[65] Shabbir, J. and Khan, N. S. (2013): On estimating the finite population mean using two auxiliary variables in two phase sampling in the presence of non-response. *Comm. Statist. Theory Methods*, 42, 4127-4145.

[66] Singh, H. P. and Solanki, R. S. (2013): An efficient class of estimators for the population mean using auxiliary information. *Comm. Statist. Theory Methods*, 42, 145-163.

[67] Singh, G. N., Majhi, D., Prasad, S. and Homa, I. (2013): Assessment of non-response under ratio method of imputation in two occasion successive sampling. *J. Statist. Theory App.*, 12(4), 403-418.

[68] Srivastava, S. K. and Jhajj, H. S. (1980): A class of estimators using auxiliary information for estimating finite population variance, *Sankhya*, C, 42(1-2), 87-96.

[69] Srivastava, S. K. and Jhajj, H. S. (1981): A class of estimators of the population mean in survey sampling using auxiliary information. *Biometrika* 68, 341-343.

[70] Tabasum, R. and Khan, I. A.(2004): Double sampling for ratio estimation with non-response. *J. Indian Soc. Agriculture Statist.*, 58(3), 300-306.

[71] Toutenburg, H. and Srivastava, V. K. (1998): Estimation of ratio of population means in survey sampling when some observations are missing. *Metrika*, 48, 177-187.

[72] Toutenburg, H., Srivastava, V. K. and Shalabh (2008): Amputation versus imputation of missing values through ratio method in sample surveys. *Statistical Papers*, 49, 237-247.

[73] Tracy, D. S. and Osahan, S. S. (1994): Ranom non-response on study variable versus on study as well as auxiliary varibles, *Statistica*, 54, 163-168.

[74] Walsh, J.E. (1970): Generalization of ratio estimator for population total. *Sankhya*, A, 32, 99-106.