23236351, 23232701, 23237721, 23234116
23235733, 23232317, 23236735, 23239437

विश्वविद्यालय अनुदान आयोग
बहादुरशाह जफर मार्ग
नई दिल्ली-110 002
UNIVERSITY GRANTS COMMISSION
BAHADURSHAH ZAFAR MARG
NEW DELHI-110 002

F. No. 42-42/2013 (SR)

[12 MAR 2013]

The Under Secretary (FD-III)
University Grants Commission
New Delhi-110002

Sub:- UGC support for the Major Research Project in Physical Sciences, Bio-Sciences, Maths , Medical, Agricultural Sciences and Engineering & Chemistry to University/College Teachers – Project entitled,

"Search of good rotation patterns on successive occasions and its applications"

Sir,

I am to refer to your letter forwarding the application of **Dr. Kumari Priyanka** of your institution for financial assistance under the above scheme and to convey the Commission's approval & sanction an on account grant of Rs. 7,36,800/- **(Rupees: seven lakh thirty six thousand eight hundred only)** to the Principal, **Shivaji College, Raja Garden, New Delhi-110027** in r/o Major Research Project of **Dr. Kumari Priyanka**, Department of Statistics for the period of 3 years w.e.f. 1.4.2013 as detailed below:-

S.No	ITEMS	AMOUNT APPROVED	GRANT RELEASED AS Ist INSTALMENT	Category
A.	**Non - Recurring**		3,25,000/-	GEN
1.	Books & Journals	75,000/-		
2.	Equipment (as per proposal)	2,50,000/-		
B.	**Recurring**		4,11,800/-	
1.	Honorarium to Retd. Teacher @ Rs. 12, 000/- p.m.	nil		
2.	Project Fellow @14,000/- p.m. for initial 2 years and Rs. 16,000/- p.m. from the third year onwards.	5,28,000/-		
3.	Chemical/ Glassware / Consumable	nil		
4.	Hiring Services	nil		
5.	Contingency	75,000/-		
6.	Travel/Field Work	1,00,000/-		
7.	Special Need	nil		
8.	Overhead Charges @ Rs. 10% approved recurring Grant (Except Travel & Field Work)	60,300/-		
	Total (A + B)	10,88,300/-	7,36,800/-	

The acceptance Certificate in prescribed format (Annexure-1 available on the UGC web-site) may be sent to the undersigned within one month from the issue of the award letter failing which the project may be treated as cancelled.

If the terms & conditions are acceptable, as per guideline which are available on UGC web-site www.ugc.ac.in the Demand Draft/ Cheque being sent may be retained. Otherwise the same may be returned in original to the UGC by Registered Post in variably with in 15 days from the receipt of the Demand Draft/Cheque in favour of Secretary, UGC, New Delhi.

Principal Investigators should ensure that the statement of expenditure & utilization Certificate to the effect that the grant has been utilized for the purpose for which it has been sanctioned shall be furnished to the University Grants Commission in time.

The first instalment of the grant shall comprise of 100% of the Non –Recurring including Over Head Charges, and 50% of the total Recurring grant.

S.O. Acc
4/4/13

1. The sanctioned amount is debitable to the Major Head 4 (i) a (31) **Rs. 4,11,800/-** & 4 (i) a (35) **Rs. 3,25,000/-** and is valid for payment during financial year 2012-13.
2. The amount of the Grant shall be drawn by the Under Secretary (drawing and Disbursing Office), University Grants Commission on the Grants-in-aid Bill and shall be disbursed to and credited to the University/College, **Shivaji College, Raja Garden, New Delhi-110027** through Cheque/Demand Draft/ Mail Transfer.
3. The Grants is subject to the adjustment of the basis of Utilization Certificate in the prescribed performa submitted by the University/Colleges/institution.
4. The University/College shall maintain proper accounts of the expenditure out of the grants which shall be utilized only on approved items of expenditure.
5. The Utilization Certificate of the effect that the grant has been utilized for the purpose for which it has been sanctioned shall be furnished to the University Grants Commission as early as possible after the close of the current financial year.
6. The assets acquired wholly or substantially out of University Grant Commission's grant shall not be disposed or encumbered of utilized for the purposes other that those for which the grant was given, without proper sanctioned of the University Grants Commission and should, at any time the College/University ceased in function such assets shall revert to the University Grant Commission.
7. A Register of assets acquired wholly or substantially out of the grant shall be maintained by the University/College in the prescribed form.
8. The grantee institution shall ensure the utilization of grant-in-aid for which it is being sanction/paid. In case non-utilization/part utilization, the simple interest @ 10% per annum as amended from time to time on unutilized amount from the date of drawl to the date of refund as per provisions contained in General Financial Rules of Govt. of India will be charged.
9. The interest earned by the University/College/Institute on this grants in aid shall be treated as additional grant and may be shown in the Utilization Certificate/Statement of expenditure to be furnished by grantee institution.
10. The University/College/Institute shall follow strictly all the instructions issued by the Government of India from time to time with regard to reservation of posts for Scheduled Castes/Scheduled Tribes/OBC/PH etc.
11. The University/College shall fully implement to Official Language Policy of Union Govt. and comply with the Official Language Act, 1963 and Official Languages (Use for Official purposes of the Union) Rules, 1978 etc.
12. The sanction issues in exercise of the delegation of powers vide Commission Office Order No. 25/92 dated May 01, 1992.
13. An amount of **Rs.** ---------- out the grant of **Rs.** ----------- sanctioned vide letter No. F. 42-42/2013 (SR) dated has been utilized by University/College/Institution for he purpose for which it was sanctioned. Utilization Certificate for **Rs.** ------------ has already been entered at S. No. ------- now we may enter Utilization Certificate for **Rs.** --------------- S. No ----- and in the U. C. Registrar at page No. -----.
14. It is also certified from the B.C.R. that the funds are available under the scheme. Entered in BCR at S.No 2717/05
The above grant is sanctioned against the budget provision of **Rs.** .... during the current financial year leaving a balance of **Rs.** -------------------------------------- under the head of Account 4 (i) a (31) **Rs. 4,11,800/-** & 4 (i) a (35) **Rs. 3,25,000/-**
15. The funds to the extent are available under the Scheme.
16. The University/Institution/College is strictly following the UGC regulations on curbing the menace of ragging in Higher Educational Institutions, 2009.

(Dr. (Mrs.) Urmila Devi)
Joint Secretary

Copy forwarded for information and necessary action for:-

1. The Principal, **Shivaji College, Raja Garden, New Delhi-110027,**
Acknowledgement for the receipt of DD / Cheque / Mail Transfer for **Rs. 7,36,800/-** may be sent to the Under Secretary, Finance Division III, UGC.
2. **Dr. Kumari Priyanka,** Principal Investigator, Department of **Mathematics**
Shivaji College, Raja Garden, New Delhi, 110027
3. office of the Director General of Audit, Central Revenues, A. G. C. R. Building, I. P. Estate, New Delhi.
4. The Registrar, **University of Delhi, Delhi**

(Pramod Sharma)
Section Officer

FINAL REPORT ON

"Search of Good Rotation Patterns on Successive Occasions and its Applications"

Major Research Project No. [42-42(2013)/SR] Submitted to

University Grant Commission, New Delhi, India

BY

Dr. KUMARI PRIYANKA
Principle Investigator
Assistant Professor
Department of Mathematics
Shivaji College
(University of Delhi)
New Delhi -110027
India

CERTIFICATE

*I, Dr. Kumari Priyanka, declare that the work presented in this report is original and carried out by me with the help of project fellow Ms. Richa Mittal, during the complete tenure of Major Research Project No. [42 - 42(2013)/SR] entitled "**Search of Good Rotation Patterns on Successive Occasions and its Applications**" financially sponsored by U.G.C., New Delhi, India.*

Kumari Priyanka 30/06/2016

Dr. KUMARI PRIYANKA
Principal Investigator
UGC Sponsored MRP No. (42-42(2013)/SR)
Title : "Search of . . . and its applications"
Department of Mathematics
Shivaji College (University of Delhi)
New Delhi-110027, India

Acknowledgement

First of all, I would like to convey my heartfelt thanks to my beloved daughter Miss Priyasha Kunjal for her constant cooperation and understanding throughout the period of project work.

I cannot forget the sacrifice of my husband Dr. Deo Nandan Kumar, Associate Professor, Department of Chemistry, Deshbandhu College, University of Delhi who gladly took over most of the familial responsibilities in spite of his heavy workload. Standing firmly by me through the ups and downs of this period, he led me to success with his gentle encouragement. Hence, I extend my heartfelt gratitude to him without his constant inspiration and support throughout my career I would not have achieved this goal.

I share my inherent respect and deep sense of gratitude to Principle, Shivaji College, University of Delhi, New Delhi India for her invaluable blessing, unstinted support and constant encouragement throughout the course of my project work.

I also express my deep sense of gratitude to Prof. Ajay Kumar (Former HOD, Department of Mathematics, University of Delhi) and Prof. V. Ravichandran (HOD, Department of Mathematics, University of Delhi) for their encouragement and the facility extended in connection with this project work.

I deeply acknowledge the financial assistance provided by University Grant Commission, New Delhi, India for carrying out the entire research work.

I also extend my sincere thanks to the project fellow Miss Rich Mittal for her sincere hard work for the project.

I would also like to express my regards and thanks to the honorable referees and editorial board of Statistics in Transition-New Series, Hacettepe Journal of Mathematics and Statistics, Communications in Statistics-Theory and Methods, International Journal of Mathematics and Statistics, Journal of Statistics Applications & Probability Letters, Communications for Statistical Applications and Methods for publishing the work embodied in this project report.

My heartfelt gratitude and indebtedness are due also to my parents and parent in-laws without whose blessings it would have been impossible for me to have this research realized.

Lines of sincere gratitude are extended to the staff and administration of Shivaji College for their valuable help in dealing with financial and administrative matters at all levels.

Sincere thanks and regards to my friend colleagues in Shivaji College, University of Delhi, India for their help, cooperation and encouragement in carrying out the present work.

Date: 30/6/2016

Kumari Priyanka

Dr. KUMARI PRIYANKA
Principal Investigator
UGC Sponsored MRP No. [42-42(2013)/SR]
Title : "Search of . . . and its applications"
Department of Mathematics
Shivaji College (University of Delhi)
New Delhi-110027, India

SYNOPSIS

Nature permits certain changes in different real life phenomenon such as social, economic, agricultural, medicinal, financial and even vital statistics related to life span of human being etc. with the change of time. To study these changes with the change of time, the different real life phenomenon need to be observed more than once since single time observation contains the subsisting situation of the study variable but not the change, over the time period. So observation need to be made at several occasions. Such a design of observations is known as Successive sampling or rotation sampling in statistical surveys which is considered a very strong statistical tool for analyzing change occurred in the phenomenon over a span of time.

The work done focuses on searching effective rotation patterns for the estimation of different population parameters like population mean and population median on successive occasions in two occasion successive sampling. The entire work has been divided in six units, in each unit population parameter has been estimated under certain set of assumption, underlying the situations for survey has been conducted.

The First unit has been devoted to the estimation of population median at current occasion in two occasion successive sampling. Various estimators have been proposed under different chapters and they have been compared to some of well-known estimators existing in the literature of successive sampling.

In chapter-1, the work deals with the problem of estimation of population median at current occasion in two-occasion successive sampling. Best linear unbiased estimators have been proposed by utilizing additional auxiliary information which is stable in nature and readily available on both the occasions.

Chapter-2 deals with the problem of estimation of finite population median at current occasion, in two occasion successive (rotation) sampling. A class of estimators has been proposed for the estimation of population median at current occasion, which includes many existing estimators as a particular case.

Chapter-3 is an attempt to explore the rotation patterns using exponential ratio type estimators for the estimation of finite population median at current occasion in two occasion rotation sampling.

Chapter-4 makes an attempt to explore the analysis on longitudinal surveys in which same units are investigated on several occasions. Multivariate exponential ratio type estimator has been proposed for the estimation of finite population median at current occasion in two occasion longitudinal surveys. Information on several additional auxiliary variables which are stable over time and readily available on both the occasions has been utilized.

In chapter-5, the problem of estimation of finite population median at current occasion in two occasion successive sampling has been considered using the additional auxiliary variate which is dynamic over time and is readily available at both the occasions.

Looking at the effective gain in precision of the estimates and decreased cost of the survey by using the exponential ratio type estimators in two occasion successive sampling, Unit-II has been devoted to the estimation of population mean by utilizing the exponential ratio type estimators since these are least utilized estimators in two occasion successive sampling.

Chapter-6 considers the problem of longitudinal analysis of population mean in two occasion successive sampling. The usefulness of exponential type estimators in enhancing the working efficiency of different ratio type estimators for population mean, when embedded with auxiliary information which is stable over time in two occasion successive sampling have been explored.

Chapter-7 deals with the problem of estimation of the population mean in presence of multi auxiliary information in two occasion rotation sampling. A multivariate exponential ratio type estimator has been proposed to estimate population mean at current (second) occasion using information on p-additional auxiliary variates which are positively correlated to study variates and are stable in nature over successive occasion.

The key and fundamental purpose of sampling over successive waves lies in the varying nature of study character, it so may happen with ancillary information if the time lag between two successive waves is sufficiently large. Chapter-8 consumes the varying nature of auxiliary information and modern approaches have been proposed to estimate population mean over two successive waves. Four exponential ratio type estimators have been designed. Cost models have also been worked out to minimize the total cost of the survey design over two successive waves.

Unit-III carry forward the idea of estimating population mean at current occasion in two occasion successive sampling but here one more aspect of surveys has been taken in to consideration that some-times in surveys, some units or the whole sample tends to be non-informative or non-responding due to any of the reason. The reason of non-response may include the absence of sample unit at said place, refusal to response or lost information etc. In such a situation, analysis of real state of facts is troubled. Unit-III explores the exponential ratio type estimators in the presence of non-response in two occasion successive sampling with the application of technique of imputation to deal with non-response.

Chapter-9 takes in consideration that while sample surveys are conducted, prompt chances of non-response of sample units leads to incompleteness of data and analyzing such data may result in false inference of facts. So utilizing the method of imputation with the aid of a completely known auxiliary character correlated to the study character and is stable in nature over the occasions, an affective estimation procedure has been suggested to deal with non-response for estimating population mean in two occasion successive sampling. A vast study has been done to elaborate the properties of the proposed estimator through theoretical and empirical entails considering that (i) non-response may arise on both occasions, (ii) it may occur only at first occasion or (iii) it may occur only at second occasion while comparing the proposed estimator with the same estimator having complete response for all sample units at each occasion.

In chapter-10, it has been discussed that the occurrence of non-response is very much plebeian in surveys, which troubles the analysis and hence an inappropriate inference is left out. To counterbalance the sour effects of the incompleteness, fresh imputation techniques have been proposed with the aid of multi-auxiliary variates for the estimation of population mean on successive waves.

Chapter-11 considers that encountering non-response is quite prone in sample surveys however smart be the design, which sours the analysis and hence the results. An effort has been made to exploit the non-response by using a completely fresh approach of imputation technique to estimate the population mean in two occasion successive sampling, utilizing completely known auxiliary information which is dynamic in nature and pronto over the occasions.

Unit-III provides a tool to negotiate with the non-response of sample units due to sensitivity of issue, although non-response may creep due to many reasons. What if non-response is due to stigmatizing character of study variable? In such surveys there is a possibility that in place of non-response, respondent simply under or over response the real facts due to social desirability and inclination. If a certain privacy level is ensured to the respondents then they may respond truthfully. Such a technique known as scrambled response technique has been explored to estimate population mean of a sensitive character.

The work done in chapter-12 is an attempt to use non-sensitive auxiliary character and scrambled response techniques to estimate population mean of a sensitive character. Various estimators using Scrambled Response Techniques (SRT) to estimate the population mean of a sensitive character have been proposed in sampling over two successive waves. Two models; Additive (ASRM) and Multiplicative (MSRM) scrambled response model have been used and the estimators have been discussed under both the models. Further pros and cons for two models in successive sampling have been illustrated. The model for optimum total cost of the survey has also been designed and discussed.

Unit-V illustrates the findings of the work done in the previous four units and makes recommendations of the work done in previous chapters on basis of requirement of survey design.

It also illustrates the further scopes for present work to be explored in future through different other survey sampling techniques.

Unit-VI show cases all the literature available in survey sampling which has been refereed to carry out the work done in this study.

INDEX

Complete Statistics of Work incorporated in the Report

Published	**:**	06
Accepted	**:**	04
Under Consideration	**:**	02
Total	**:**	**12**

UNIT - I

SEARCH OF GOOD ROTATION PATTERNS FOR ESTIMATION OF POPULATION MEDIAN AT CURRENT OCCASION

CHAPTER – 1*

Effective Rotation Patterns for Median Estimation in Successive Sampling

---

* Following is the publication based on the work of this chapter:--

1. Priyanka, K. and Mittal, R. (2014): Effective Rotation Patterns for Median Estimation in Successive Sampling. ***Statistics in Transition-new series***, Vol. 15, No. 2, 197-220.

Effective Rotation Patterns for Median Estimation in Successive Sampling

1. Introduction

When the value of study character of a finite population is subject to change (dynamic) over time, a survey carried out on a single occasion will provide information about the characteristics of the surveyed population for the given occasion only and will not give any information on the nature of change of the characteristic over different occasions and the average value of the characteristic over all occasions or the most recent occasion. To meet these requirements, sampling is done on successive occasions that provide a strong tool for generating the reliable estimates at different occasions. The problem of sampling on two successive occasions was first considered by Jessen (1942), and latter this idea was extended by Patterson (1950), Narain (1953), Eckler (1955), Gordon (1983), Arnab and Okafor (1992), Feng and Zou (1997), Singh and Singh (2001), Singh and Priyanka (2008a), Singh et al.(2012), Bandyopadhyay and Singh(2014) and many others.

All the above studies were concerned with the estimation of population mean or variance on two or more occasion.

There are many problems of practical interest which involves variables with extreme values that strongly influence the value of mean. In such situations the study variable is having highly skewed distributions. For example, the study of environmental issues, the study of social evil such as abortions, the study of income, expenditure etc. In these situations, the mean may offer results which are not representative enough because it moves with the direction of the asymmetry. The median, on the other hand is unaffected by extreme values.

Most of the studies related to medians have been developed by assuming simple random sampling or its ramification in stratified random sampling (Gross (1980), Sedransk and Meyer (1978), Smith and Sedransk (1983) and considering only the variable of interest without making explicit use of auxiliary variables. Some of the researchers namely Chambers and Dunstan(1986), Kuk and Mak (1989), Rao et al. (1990), Rueda et al.(1998), Allen et al. (2002), Singh and Solanki (2013) etc. make use of auxiliary variables to estimate the population median.

It is to be mentioned that a large number of estimators for estimating the population mean at current occasion have been proposed by various authors, however, only a few efforts (namely Martinez-Miranda et al. (2005), Singh et al. (2007) and Rueda et al. (2008)), Gupta et al. (2008) have been made to estimate the population median on the current occasion in two occasions successive sampling. It is well known that the use of auxiliary information at the estimation stage can typically increase the precision of estimates of a parameter. To the best of our knowledge, no effort has been made to use additional auxiliary information readily available on both the occasions to estimate population median at current occasion in two- occasion successive sampling.

Motivated with the above arguments and utilizing the information on an additional auxiliary variable, readily available on both the occasions, the best linear unbiased estimators for estimating the population median on current occasion in two-occasion successive sampling have been proposed. It has been assumed that the additional auxiliary variable is stable over the two-occasions.

The work is spread over ten sections. Sample structure and notations have been discussed in section 2. In section 3 the proposed estimator has been formulated. Properties of proposed estimators including variances are derived under section 4. Minimum variance of the prosed estimator is derived in section 5. Practicability of the proposed estimator is also discussed. In section 6 optimum replacement policies are discussed. Section 7 contains comparison of proposed estimator with the natural sample median estimator, when there is no matching from the previous occasion and the estimator when no additional auxiliary information has been used. Practicability of the estimator Δ is also discussed. In section 8 simulation studies have been carried out to investigate the performance of the proposed estimators. The results obtained as a result of empirical and simulation studies have been elaborated in section 9. Finally the conclusion of the entire work has been presented in section 10.

2. Sample Structures and notations

Let U = (U_1, U_2, - - -, U_N) be the finite population of N units, which has been sampled over two occasions. It is assumed that the size of the population remains unchanged but values of unit change over two occasions. The character under study be denoted by x (y) on the first (second) occasions respectively. It is further assumed that information on an auxiliary variable z (with known population median) is available on both the occasions. A simple random sample (without replacement) of n units is taken on the first occasion. A random sub sample of m = n λ units is retained (matched) for use on the second occasion. Now, at the current occasion a simple random sample (without replacement) of u= (n - m) = nμ units is drawn afresh from the remaining (N - n) units of the population so that the sample size on the second occasion is also n. λ and μ, (λ+ μ =1) are the fractions of matched and fresh samples respectively at the second (current) occasion. The following notations are considered for the further use:

M_x, M_y, M_z : Population median of x, y and z respectively.

$\hat{M}_{x(n)}, \hat{M}_{x(m)}, \hat{M}_{y(m)}, \hat{M}_{y(u)}, \hat{M}_{z(n)}, \hat{M}_{z(m)}, \hat{M}_{z(u)}$: Sample median of the respective variables of the sample sizes shown in suffices.

$\rho_{yx}, \rho_{xz}, \rho_{yz}$: The Correlation coefficient between the variables shown in suffices.

3. Formulation of Estimator

To estimate the population median M_y on the current (second) occasion, the minimum variance linear unbiased estimator of M_y under SRSWOR sampling scheme have been proposed and is given as

$$T=\left\{\alpha_1\hat{M}_{y(u)}+\alpha_2\hat{M}_{y(m)}\right\}+\left\{\alpha_3\hat{M}_{x(m)}+\alpha_4\hat{M}_{x(n)}\right\}+\left\{\alpha_5\hat{M}_{z(u)}+\alpha_6\hat{M}_{z(m)}+\alpha_7\hat{M}_{z(n)}+\alpha_8 M_z\right\} \qquad (1)$$

where $\alpha_i\,(i=1, 2, ---, 8)$ are constants to be determined so that

(i) The estimator T becomes unbiased for M_y and

(ii) The variance of T attains a minimum

For unbiasedness, the following conditions must hold

$(\alpha_1+\alpha_2)=1$, $(\alpha_3+\alpha_4)=0$ and $(\alpha_5+\alpha_6+\alpha_7+\alpha_8)=0$.

Substituting $\alpha_1=\varphi_1$, $\alpha_3=\beta_1$ and $\alpha_8=-(\alpha_5+\alpha_6+\alpha_7)$ in equation (1), the estimator T takes the following form

$$T = \left\{\varphi_1\hat{M}_{y(u)} + (1-\varphi)\hat{M}_{y(m)}\right\} + \beta_1\left\{\hat{M}_{x(m)} - \hat{M}_{x(n)}\right\} + \left\{\alpha_5\left(\hat{M}_{z(u)} - M_z\right) + \alpha_6\left(\hat{M}_{z(m)} - M_Z\right) + \alpha_7\left(\hat{M}_{z(n)} - M_Z\right)\right\}$$

$$= \varphi_1\left\{\hat{M}_{y(u)} + k_1\left(\hat{M}_{z(u)} - M_Z\right)\right\} + (1-\varphi_1)\left\{\hat{M}_{y(m)} + k_2\left(\hat{M}_{x(m)} - \hat{M}_{x(n)}\right) + k_3\left(\hat{M}_{z(m)} - M_Z\right) + k_4\left(\hat{M}_{z(n)} - M_Z\right)\right\}$$

$$T = \varphi_1 T_1 + (1-\varphi_1)\, T_2 \qquad (2)$$

where $T_1=\hat{M}_{y(u)} + k_1\left(\hat{M}_{z(u)} - M_Z\right)$ is based on the sample of size u drawn afresh at current occasion and the estimator

$T_2=\left\{\hat{M}_{y(m)} + k_2\left(\hat{M}_{x(m)} - \hat{M}_{x(n)}\right) + k_3\left(\hat{M}_{z(m)} - M_Z\right) + k_4\left(\hat{M}_{z(n)} - M_Z\right)\right\}$ is based on the sample of size m matched form previous occasion .

$k_1=\frac{\alpha_5}{\varphi_1}$, $k_2=\frac{\beta_1}{1-\varphi_1}$, $k_3=\frac{\alpha_6}{1-\varphi_1}$, $k_4=\frac{\alpha_7}{1-\varphi_1}$ and φ_1 are the unknown constants to be determined so as to minimize the variance of estimator T.

Remark 3.1: For estimating the median on each occasion, the estimator T_1 is suitable, which implies that more belief on T_1 could be shown by choosing φ_1 as 1 (or close to 1), while for estimating the change from one occasion to the next, the estimator T_2 could be more useful so φ_1 be chosen as 0(or close to 0). For asserting both the problems simultaneously, the suitable (optimum) choice of φ_1 is required.

4. Properties of the estimator T

The properties of the proposed estimator T are derived under the following assumptions:

(i) Population size is sufficiently large (i.e. $N\rightarrow\infty$), therefore finite population corrections are ignored.

(ii) As $N\rightarrow\infty$, the distribution of bivariate variable (a, b) where a and $b\in\{x, y, z\}$ and $a \neq b$ approaches a continuous distribution with marginal densities $f_a(\cdot)$ and $f_b(\cdot)$ for a and b respectively, see Kuk and Mak (1989).

(iii) The marginal densities $f_x(\cdot)$, $f_y(\cdot)$ and $f_z(\cdot)$ are positive.

(iv) The sample medians $\hat{M}_{x(n)}, \hat{M}_{x(m)}, \hat{M}_{y(m)}, \hat{M}_{y(u)}, \hat{M}_{z(n)}, \hat{M}_{z(m)}$ and $\hat{M}_{z(u)}$ are consistent and asymptotically normal (see Gross (1980)).

(v) Following Kuk and Mak (1989), let P_{ab} be the proportion of elements in the population such that $a \leq M_a$ and $b \leq M_b$ where a and $b\in\{x, y, z\}$ and $a \neq b$.

(vi) Following large sample approximations are assumed:

$\hat{M}_{y(u)}=M_y(1+e_0)$, $\hat{M}_{y(m)}=M_y(1+e_1)$, $\hat{M}_{x(m)}=M_x(1+e_2)$, $\hat{M}_{x(n)}=M_x(1+e_3)$, $\hat{M}_{z(u)}=M_z(1+e_4)$, $\hat{M}_{z(m)}=M_z(1+e_5)$ and $\hat{M}_{z(n)}=M_z(1+e_6)$ such that $|e_i| < 1 \;\forall\; i = 0, 1, 2, 3, 4, 5, 6$.

The values of various related expectations can be seen in Allen et al. (2002) and Singh (2003). Under the above transformations, the estimator T_1 and T_2 takes the following forms:

$$T_1 = M_y(1+e_0) + k_1 M_z e_4 \tag{3}$$

$$T_2 = M_y(1+e_1) + k_2 M_x (e_2 - e_3) + M_z (k_3 e_5 + k_4 e_6) \tag{4}$$

Thus we have the following theorems:

Theorem 4.1: T is unbiased estimator of M_y.

Proof: Since T_1 and T_2 are difference and difference-type estimators, respectively so they are unbiased for M_y. The combined estimator T is a convex linear combination of T_1 and T_2, hence it is also an unbiased estimator of M_y.

Theorem 4.2: Ignoring the finite population corrections, the variance of T is

$$V(T) = \varphi_1^2 V(T_1) + (1-\varphi_1)^2 V(T_2) \tag{5}$$

where $V(T_1) = \frac{1}{u}\xi_1$ (6)

and $V(T_2) = \frac{1}{m}\xi_2 + \left(\frac{1}{m} - \frac{1}{n}\right)\xi_3 + \frac{1}{n}\xi_4$ (7)

$\xi_1 = A_1 + k_1^2 A_2 + 2k_1 A_3$, $\xi_2 = A_1 + k_3^2 A_2 + 2k_3 A_3$, $\xi_3 = k_2^2 A_4 + 2k_2 A_5 + 2k_2 k_3 A_6$,

$\xi_4 = k_4^2 A_2 + 2k_4 A_3 + 2k_3 k_4 A_2$, $A_1 = \frac{1}{4}\{f_y(M_y)\}^{-2}$, $A_2 = \frac{1}{4}\{f_z(M_z)\}^{-2}$,

$A_3 = (P_{yz} - 0\times 25)\{f_y(M_y)\}^{-1}\{f_z(M_z)\}^{-1}$, $A_4 = \frac{1}{4}\{f_x(M_x)\}^{-2}$,

$A_5 = (P_{yx} - 0\times 25)\{f_y(M_y)\}^{-1}\{f_x(M_x)\}^{-1}$ and $A_6 = (P_{xz} - 0\times 25)\{f_x(M_x)\}^{-1}\{f_z(M_z)\}^{-1}$.

Proof: The variance of T is given by

$$V(T) = E(T - M_y)^2 = E\left[\varphi_1(T_1 - M_y) + (1-\varphi_1)(T_2 - M_y)\right]^2$$

$$=\varphi_1^2\ V(T_1)+(1-\varphi_1)^2 V(T_2)+\varphi_1(1-\varphi_1)\ \text{cov}(T_1,T_2) \qquad (8)$$

where $V(T_1)=E(T_1-M_y)^2$ and $V(T_2)=E(T_2-M_y)^2$.

As T_1 and T_2 are based on two independent samples of sizes u and m respectively, hence $\text{cov}\,(T_1, T_2)=0$.

Now, substituting the expressions of T_1 and T_2 from equations (3) and (4) in equation (8), taking expectations and ignoring finite population corrections, we have the expression for variance of T as in equation (5).

5. Minimum Variance of the Estimator T

Since, the variance of the estimator T in equation (5) is the function of unknown constants k_1, k_2, k_3, k_4 and φ_1, therefore it is minimized with respect to k_1, k_2, k_3, k_4 and φ_1 and subsequently the optimum values of k_1, k_2, k_3, k_4 and φ_1 are obtained as

$$k_1^*=\frac{-A_3}{A_2} \qquad (9)$$

$$k_2^*=\frac{A_3A_4A_6-A_2A_4A_5}{A_4\left(A_2A_4-A_6^2\right)} \qquad (10)$$

$$k_3^*=\frac{-A_3A_4+A_5A_6}{\left(A_2A_4-A_6^2\right)} \qquad (11)$$

$$k_4^*=\frac{A_3A_6^2-A_2A_5A_6}{A_2\left(A_2A_4-A_6^2\right)} \qquad (12)$$

$$\varphi_{1opt\times}=\frac{V(T_2)}{V(T_1)+V(T_2)} \qquad (13)$$

Using the optimum values of k_i's $(i=1, 2, 3, 4)$ in equation (6) and (7), we get the optimum variances of T_1 and T_2 as

$$V(T_1)_{opt.} = \frac{1}{u} A_7 \tag{14}$$

$$V(T_2)_{opt.} = \frac{1}{m} A_8 + \left(\frac{1}{m} - \frac{1}{n}\right) A_9 + \frac{1}{n} A_{10} \tag{15}$$

where $A_7 = A_1 + k_1^{*2} A_2 + 2k_1^* A_3$, $A_8 = A_1 + k_3^{*2} A_2 + 2k_3^* A_3$, $A_9 = k_2^{*2} A_4 + 2k_2^* A_5 + 2k_2^* k_3^* A_6$ and $A_{10} = k_4^{*2} A_2 + 2k_4^* A_3 + 2k_3^* k_4^* A_2$.

Further substituting the values of $V(T_1)_{opt.}$ and $V(T_2)_{opt.}$ from equations (14) and (15) in equation (13), we get the optimum values of $\varphi_{1opt.}$ with respect to k_i^*'s $(i=1, 2, 3, 4)$ as

$$\varphi^*_{1opt.} = \frac{V(T_2)_{opt.}}{V(T_1)_{opt.} + V(T_2)_{opt.}} \tag{16}$$

Again substituting the value of $\varphi^*_{1opt.}$ from equation (16) in equation (5), we get the optimum variance of T as

$$V(T)_{opt.} = \frac{V(T_1)_{opt.} V(T_2)_{opt.}}{V(T_1)_{opt.} + V(T_2)_{opt.}} \tag{17}$$

Further, substituting the value from (14) and (15) in equation (16) and (17), we get the simplified values of $\varphi^*_{1opt.}$ and $V(T)_{opt.}$ as

$$\varphi^*_{1opt.} = \frac{\mu(A_{11} + \mu A_{12})}{\mu^2 A_{12} + \mu^2 A_{13} + A_7} \tag{18}$$

$$V(T)_{opt.} = \frac{1}{n} \frac{A_7 (A_{11} + \mu A_{12})}{(\mu^2 A_{12} + \mu A_{13} + A_7)} \tag{19}$$

where $A_{11} = A_8 + A_{10}$, $A_{12} = A_9 - A_{10}$, $A_{13} = A_{11} - A_7$ and μ is the fraction of fresh sample at current occasion for the estimator T.

5.1 Estimator T in practice

The main difficulty in using the proposed estimator T defined in equation (2), is the availability of k_i's $(i = 1, 2, 3, 4)$ as the optimum values of k_i's $(i = 1, 2, 3,)$ depends on the population parameters $P_{yx}, P_{yz}, P_{xz}, f_y(M_y), f_x(M_x)$ and $f_z(M_z)$. If these parameters are known, the proposed estimator can be easily implemented. Otherwise, which is the most often situation in practice, the unknown population parameters are replaced by their respective sample estimates. The population proportions P_{yx}, P_{yz} and P_{xz} are replaced by the sample estimates $\hat{P}_{yx}, \hat{P}_{yz}$ and $\hat{P}_{xz}$ respectively and the marginal densities $f_y(M_y), f_x(M_x)$ and $f_z(M_z)$ can be substituted by their kernel estimator or nearest neighbour density estimator or generalized nearest neighbour density estimator related to the kernel estimator (Silverman (1986)). Here, the marginal densities $f_y(M_y)$, $f_x(M_x)$ and $f_z(M_z)$ are replaced by $\hat{f}_y(\hat{M}_{y(m)}), \hat{f}_x(\hat{M}_{x(n)})$ and $\hat{f}_z(\hat{M}_{z(n)})$ respectively, which are obtained by the method of generalized nearest neighbour density estimation related to the kernel estimator.

Remark 5.1.1: To estimate $f_x(M_x)$, by the generalized nearest neighbour density estimator related to the kernel estimator, following procedure has been adopted:

Choose an integer $h \approx n^{1/2}$ and define the distance $d(x_1, x_2)$ between two points on the line to be $|x_1 - x_2|$.

For $\hat{M}_{x(n)}$ define $d_1(\hat{M}_{x(n)}) \le d_2(\hat{M}_{x(n)}) \le - - - \le d_n(\hat{M}_{x(n)})$ to be the distances, arranged in ascending order, from $\hat{M}_{x(n)}$ to the points of the sample.

The generalized nearest neighbour density estimate is defined by

$$\hat{f}(\hat{M}_{x(n)}) = \frac{1}{nd_h(\hat{M}_{x(n)})} \sum_{i=1}^{n} K\left(\frac{\hat{M}_{x(n)} - x_i}{d_h(\hat{M}_{x(n)})}\right) \tag{20}$$

where the kernel function K, satisfies the condition $\int_{-\infty}^{\infty} K(x)\, dx = 1$.

Here, the kernel function is chosen as Gaussian Kernel given by $K(x) = \frac{1}{2\pi} e^{-\left(\frac{1}{2}x^2\right)}$.

Similarly, the estimate of $f_y(M_y)$ and $f_z(M_z)$ can be obtained.

Remark 5.1.2: For estimating $f_y(M_y)$, P_{yz} and P_{yx} we are having two independent samples of sizes u and m respectively at current occasion. So, either of the two can be used, but in general for good sampling design in successive sampling $u \le m$. So, in the present work $f_y(M_y)$, P_{yz} and P_{yx} are estimated from sample of size m, matched from first occasion.

Therefore, under the above substitutions of the unknown population parameters by their respective sample estimates, the estimator T takes the following form:

$$T^* = \psi_1 T_1^* + (1 - \psi_1)\, T_2^* \tag{21}$$

where $$T_1^* = \hat{M}_{y(u)} + k_1^{**}\left(\hat{M}_{z(u)} - M_Z\right) \tag{22}$$

and $$T_2^* = \left\{\hat{M}_{y(m)} + k_2^{**}\left(\hat{M}_{x(m)} - \hat{M}_{x(n)}\right) + k_3^{**}\left(\hat{M}_{z(m)} - M_Z\right) + k_4^{**}\left(\hat{M}_{z(n)} - M_Z\right)\right\} \tag{23}$$

$$k_1^{**} = \frac{-A_3^*}{A_2^*},\; k_2^{**} = \frac{A_3^*A_4^*A_6^* - A_2^*A_4^*A_5^*}{A_4^*\left(A_2^*A_4^* - A_6^{*2}\right)},\; k_3^{**} = \frac{-A_3^*A_4^* + A_5^*A_6^*}{\left(A_2^*A_4^* - A_6^{*2}\right)},\; k_4^{**} = \frac{A_3^*A_6^{*2} - A_2^*A_5^*A_6^*}{A_2^*\left(A_2^*A_4^* - A_6^{*2}\right)},$$

$$A_1^* = \frac{1}{4}\left\{\hat{f}_y\left(\hat{M}_{y(m)}\right)\right\}^{-2},\quad A_2^* = \frac{1}{4}\left\{\hat{f}_z\left(\hat{M}_{z(n)}\right)\right\}^{-2},$$

$$A_3^* = \left(\hat{P}_{yz} - 0.25\right)\left\{\hat{f}_y\left(\hat{M}_{y(m)}\right)\right\}^{-1}\left\{\hat{f}_z\left(\hat{M}_{z(n)}\right)\right\}^{-1},\qquad A_4^* = \frac{1}{4}\left\{\hat{f}_x\left(\hat{M}_{x(n)}\right)\right\}^{-2},$$

$$A_5^* = \left(\hat{P}_{yx} - 0.25\right)\left\{\hat{f}_y\left(\hat{M}_{y(m)}\right)\right\}^{-1}\left\{\hat{f}_x\left(\hat{M}_{x(n)}\right)\right\}^{-1} \text{ and}$$

$$A_6^* = \left(\hat{P}_{xz} - 0.25\right)\left\{\hat{f}_x\left(\hat{M}_{x(n)}\right)\right\}^{-1}\left\{\hat{f}_z\left(\hat{M}_{z(n)}\right)\right\}^{-1}.$$

ψ_1 is an unknown constant to be determined so as to minimize the mean square error of the estimator T^*.

Remark 5.1.3: The proposed estimator T is difference-type estimator so, after replacing the unknown population parameters by their respective sample estimates it becomes regression-type estimator. Hence, up to the first order of approximations the estimator T^* will be equally precise to that of the estimator T (see Singh and Priyanka (2008a)). Therefore, similar conclusions are applicable for T^* as that of T.

6. Optimum Replacement Policy

To determine the optimum value of μ (fraction of sample to be taken afresh at second occasion) so that M_y may be estimated with maximum precision, we minimize $V(T)_{opt.}$ in equation (19) with respect to μ and hence we get the optimum value of μ as

$$\mu_{opt.^*} = \frac{-S_2 \pm \sqrt{S_2^2 - S_1 S_3}}{S_1} = \mu_0 \text{ (say)} \tag{24}$$

where, $S_1 = A_{12}^2$, $S_2 = A_{11}A_{12}$ and $S_3 = A_{11}A_{13} - A_7 A_{12}$.

From equation (24), it is obvious that the real value of $\mu_{opt.}$ exists if $S_2^2 - S_1 S_3 \geq 0$. For certain situation, there might be two values of $\mu_{opt.}$ satisfying the above condition, hence to choose a value of $\mu_{opt.}$, it should be remembered that $0 \leq \mu_{opt.} \leq 1$. All other values of $\mu_{opt.}$ are inadmissible. In case if both the values of $\mu_{opt.}$ are admissible, we choose the minimum of these two as μ_0. Substituting the value of $\mu_{opt.}$ from equation (24) in (19) we have

$$V(T)_{opt.^*} = \frac{1}{n} \frac{A_7 \left(A_{11} + \mu_0 A_{12}\right)}{\left(\mu_0^2 A_{12} + \mu_0 A_{13} + A_7\right)} \tag{25}$$

where $V(T)_{opt.^*}$ is the optimum value of T with respect μ.

7. Efficiency Comparison

To study the performance of the estimator T, the percent relative efficiencies of T with respect to (i) $\hat{M}_{y(n)}$, the natural estimator of M_y, when there is no matching and (ii) the estimator Δ, when no additional auxiliary information is used at any occasion, have been computed for two natural population data. The estimator Δ is defined under the same circumstances as the estimator T, but in the absence of information on additional auxiliary variable z on both the occasions and is proposed as

$$\Delta=\left\{\delta_1\hat{M}_{y(u)}+\delta_2\hat{M}_{y(m)}\right\}+\left\{\delta_3\hat{M}_{x(m)}+\delta_4\hat{M}_{x(n)}\right\} \quad (26)$$

where $\delta_i\ (i=1, 2, 3, 4)$ are constants to be determined so that

(i) The estimator Δ becomes unbiased for M_y and

(ii) The variance of Δ attains a minimum.

For unbiasedness, the following conditions must hold

$(\delta_1+\delta_2)=1$ and $(\delta_3+\delta_4)=0$.

Substituting $\delta_1=\varphi_2$ and $\delta_3=\beta_2$ in equation (26), the estimator Δ takes the following form

$$\Delta=\left\{\varphi_2\hat{M}_{y(u)}+(1-\varphi_2)\hat{M}_{y(m)}\right\}+\beta_2\left(\hat{M}_{x(m)}-\hat{M}_{x(n)}\right)$$

$$=\varphi_2\hat{M}_{y(u)}+(1-\varphi_2)\left\{\hat{M}_{y(m)}+k_5\left(\hat{M}_{x(m)}-\hat{M}_{x(n)}\right)\right\}$$

$$\Delta=\varphi_2\Delta_1+(1-\varphi_2)\Delta_2 \quad (27)$$

where, the estimator $\Delta_1=\hat{M}_{y(u)}$ is based on the fresh sample of size u and the estimator $\Delta_2=\left\{\hat{M}_{y(m)}+k_5\left(\hat{M}_{x(m)}-\hat{M}_{x(n)}\right)\right\}$ is based on the matched sample of size m, $k_5=\frac{\beta_2}{(1-\varphi_2)}$ and φ_2 are the unknown constants to be determined so as to minimize the variance of estimator Δ. Following the methods discussed in Sections 4, 5 and 6, the optimum value of k_5, $\mu_{1opt.}$ (optimum value of fraction of fresh sample for the estimator Δ), variance of $\hat{M}_{y(n)}$ and optimum variance of Δ ignoring the finite population corrections are given by

$$k_5^* = \frac{-A_5}{A_4} \tag{28}$$

$$\mu_{1opt.^*} = \frac{-A_1 \pm \sqrt{A_1\left(A_1 + A_{14}\right)}}{A_{14}} = \mu^*\left(\text{say}\right) \tag{29}$$

$$V\left(\hat{M}_{y(n)}\right) = \frac{1}{n} A_1 \tag{30}$$

$$V\left(\Delta\right)_{opt.^*} = \frac{1}{n}\frac{A_1\left(A_1 + \mu^* A_{14}\right)}{\left(\mu^{*2} A_{14} + A_1\right)} \tag{31}$$

where $A_{14} = \frac{-A_5^2}{A_4}$.

The optimum values of μ, μ_1 and percent relative efficiencies E_1 and E_2 of the estimator T with respect to the estimator $\hat{M}_{y(n)}$ and Δ are computed for two natural populations and results are shown in Tabe-2, where

$$E_1 = \frac{V\left(\hat{M}_{y(n)}\right)}{V\left(T\right)_{opt^*.}} \times 100 \text{ and } E_2 = \frac{V\left(\Delta\right)_{opt^*.}}{V\left(T\right)_{opt^*.}} \times 100$$

7.1 Estimator Δ in practice

The main difficulty in using the proposed estimator Δ defined in equation (27), is the availability of k_5, as the optimum values of k_5 depends on the population parameters P_{yx}, $f_y\left(M_y\right)$ and $f_x\left(M_x\right)$. If these parameters are known, the estimator Δ can easily be implemented otherwise the unknown population parameters are replaced by their respective sample estimates as discussed in subsection 5.1. Hence, in this scenario the estimator Δ takes the following form:

$$\Delta^{*}=\psi_2\Delta_1+\left(1-\psi_2\right)\Delta_2^{*} \tag{32}$$

where $\Delta_2^{*}=\left\{\hat{M}_{y(m)}+k_5^{**}\left(\hat{M}_{x(m)}-\hat{M}_{x(n)}\right)\right\}$, $k_5^{**}=\dfrac{-A_5^{*}}{A_4^{*}}$ and ψ_2 is the unknown constants to be determined so as to minimize the mean squared error of the estimator Δ^{*}.

Remark 7.1.1: Since, Δ^{*} is regression-type estimator corresponding to the difference-type estimator Δ, hence up to the first order of approximations similar conclusions are applicable to Δ^{*} as that of Δ (See Singh and Priyanka (2008a)).

Remark 7.1.2: For simulation study the proposed estimator T^{*} and Δ^{*} are considered instead of the proposed estimators T and Δ respectively.

8. Monte Carlo Simulation

Empirical validation can be carried out by Monte Carlo Simulation. Real life situations of completely known two finite populations have been considered.

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

The first population comprise of N = 51 states of United States. Let y_i represent the number of abortions during 2007 in the i^{th} state of U. S., x_i be the number of abortions during 2005 in the i^{th} state of U. S. and z_i denote the number of abortions during 2004 in the i^{th} state of U. S. The data are presented in Figure 1.

Similarly, the second population consists of N=41 corn producing states of United States. We assume y_i the production of corn (in million bushels) during 2009 in the i^{th} state of U.S., x_i be the production of corn (in million bushels) during 2008 in the i^{th} state of U. S. and z_i denote the production of corn (in million bushels) during 2007 in the i^{th} state of U. S. The data are represented by means of graph in Figure 2.

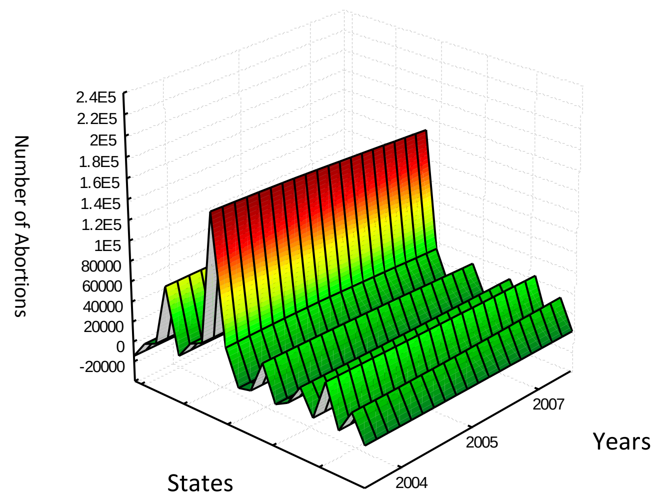


Figure 1: Number of abortions during 2004, 2005 and 2007 versus different states of US

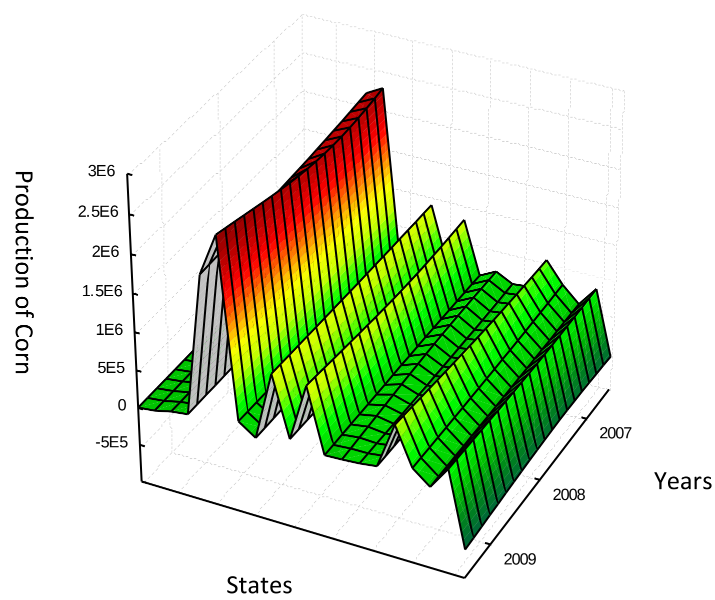


Figure 2: Production of corn during 2007, 2008 and 2009 versus different states of US

The graphs in Figure1 and Figure 2 show that the number of abortions and the production of corn in different states are skewed towards right. One reason of skewness for the population-I may be the distribution of population in different states, that is the states having larger population are expected to have larger number of abortion cases. Similarly for population-II, the states having larger area for farming are expected to have larger production of corn. Thus skewness of data indicates that the use of median may be a good measure of central location than mean in these situations.

For performing the Monte Carlo Simulation in the considered population-I, 5000 samples of n=20 states were selected using simple random sampling without replacement in the year 2005. The sample medians $\hat{M}_{x(n)|k}$ and $\hat{M}_{z(n)|k}$, k =1, 2, - - -,5000 were computed and the parameters $f_x(M_x)$, $f_z(M_z)$ and P_{xz} were estimated by the method given in Remark 5.1.1. From each one of the selected samples, m=17 states were retained and new u=3 states were selected out of N – n =51 – 20 = 31 states using simple random sampling without replacement in the year 2007. From the m units retained in the sample at the current occasion, the sample medians $\hat{M}_{x(m)|k}$, $\hat{M}_{y(m)|k}$ and $\hat{M}_{z(m)|k}$, k = 1, 2,- - -,5000 were computed and the parameters $f_y(M_y)$, P_{yz} and P_{xz} were estimated. From the new unmatched units selected on the current occasion the sample medians $\hat{M}_{y(u)|k}$ and $\hat{M}_{z(u)|k}$, k = 1, 2,- - -,5000 were computed. The parameters ψ_1 and ψ_2 are selected between 0.1 and 0.9 with a step of 0.1.

The percent relative efficiencies of the proposed estimator T^* with respect to $\hat{M}_{y(n)}$ and Δ^* are respectively given by:

$$E_{1sim} = \frac{\sum_{k=1}^{5000}\left[\hat{M}_{y(n)|k} - M_y\right]^2}{\sum_{k=1}^{5000}\left[T_k^* - M_y\right]^2}\times 100 \quad \text{and} \quad E_{2sim} = \frac{\sum_{k=1}^{5000}\left[\Delta_k^* - M_y\right]^2}{\sum_{k=1}^{5000}\left[T_k^* - M_y\right]^2}\times 100$$

For better analysis, this simulation experiments were repeated for different choices of μ.

Similar steps are also followed for Population-II. The simulation results in Table 3, Table 4 and Table 5 show the comparison of the proposed estimator T^* with respect to the estimators $\hat{M}_{y(n)}$ and Δ^* respectively. For convenience the different choices of μ are considered as different sets for the considered Population-I and Population-II, which are shown below:

Sets	Population-I	Population-II
I	n =20; μ = 0.15 (m =17, u =3)	n=15; μ = 0.13 (m =13, u =2)
II	n =20; μ = 0.25 (m = 15, u =5)	n=15; μ = 0.20 (m =12, u =3)
III	n =20; μ = 0.35 (m = 13, u =7)	n=15; μ = 0.30 (m = 10, u =5)
IV	n =20; μ = 0.50 (m = 10, u =10)	n=15; μ = 0.40 (m = 9, u =6)

Table 1: Descriptive statistics for Population-I and Population-II

	Population-I			Population-II		
	Abortions 2004 (z)	Abortions 2005 (x)	Abortions 2007 (y)	Production of Corn in 2007 (z)	Production of Corn in 2008 (x)	Production of Corn in 2009 (y)
Mean	23963.14	23651.76	23697.65	317997	294918.2	319313.7
Median	11010.00	10410.00	9600.00	83740	66650	79730
Standard Deviation	38894.81	38487.71	39354.65	565641.6	530483.7	563103.3
Kurtosis	12.02669	12.39229	14.42803	6.838888	6.492807	6.036604
Skewness	3.275197	3.310767	3.527683	2.638611	2.595704	2.499771
Minimum	80	70	90	2997	2475	2635
Maximum	208180	208430	223180	2376900	2188800	2420600
Count	51	51	51	41	41	41

Table 2: Comparison of the proposed estimator T (at optimal conditions) with respect to the estimators $\hat{M}_{y(n)}$ and Δ (at optimal conditions)

	Population - I	**Population-II**
μ_0	0.5411	0.6669
μ^*	0.6800	0.7642
E_1	1407.5	1401.3
E_2	1034.9	916.80

Table 3: Monte Carlo Simulation results when the proposed estimator T^* is compared to $\hat{M}_{y(n)}$ for population-I and population-II

	Population-I				Population-II			
Set	I	II	III	IV	I	II	III	IV
$\psi_1 \downarrow$	E_{1sim}	E_{1sim}	E_{1sim}	E_{1sim}	E_{1sim}	E_{1sim}	E_{1sim}	E_{1sim}
0.1	338.42	285.75	294.74	191.46	762.21	747.03	127.19	321.48
0.2	330.71	291.82	320.22	238.4	860.29	644.25	140.93	364.51
0.3	315.85	288.81	333.44	254.30	971.34	536.15	154.84	397.27
0.4	282.71	288.70	326.08	276.75	1097.6	427.33	166.51	420.99
0.5	248.64	268.90	322.70	295.47	1219.7	340.46	172.53	413.40
0.6	210.41	249.90	299.55	301.46	1377.0	262.76	175.98	413.49
0.7	178.81	220.94	269.87	304.12	1529.3	206.40	172.93	398.24
0.8	152.05	194.11	245.61	297.46	1707.7	166.72	166.51	369.96
0.9	127.19	168.82	216.58	289.94	1855.9	136.86	161.50	336.32

Table 4: Monte Carlo Simulation results for Population-I when the proposed estimator T^* is compared to Δ^*

$\psi_1 \downarrow$	$\psi_2 \rightarrow$		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	E_{2sim}	I	329.1	470.4	707.2	1017.2	1590.3	2211.0	2869.2	4255.0	5490.3
		II	269.4	272.6	291.4	424.8	681.0	752.7	1023.3	1511.8	1790.9
		III	285.6	233.2	273.0	320.1	430.9	624.4	770.1	1126.7	1353.6
		IV	205.2	188.5	168.7	168.4	198.1	230.3	318.0	419.5	559.2
0.2	E_{2sim}	I	340.3	456.3	714.2	1078.2	1685.3	2268.1	3064.6	4227.3	5437.1
		II	285.8	282.7	312.6	461.3	678.1	824.9	1150.8	1600.8	2034.9
		III	295.9	251.1	279.7	344.3	457.5	636.8	831.4	1126.8	1428.8
		IV	242.3	199.2	177.2	182.9	222.9	269.7	351.5	483.4	631.6
0.3	E_{2sim}	I	325.9	440.9	688.6	1071.6	1547.1	2158.4	2979.3	4060.1	5145.1
		II	288.6	285.4	336.3	475.3	677.2	839.5	1187.6	1643.4	1983.4
		III	298.7	264.8	287.5	358.9	456.2	642.1	852.9	1159.3	1466.2
		IV	261.4	216.4	192.2	198.1	247.3	294.9	391.5	529.6	681.6
0.4	E_{2sim}	I	298.2	411.3	624.7	967.3	1430.2	1975.9	2648.7	3594.8	4721.6
		II	284.9	282.3	329.8	454.1	659.4	842.4	1152.1	1600.3	1946.5
		III	289.6	265.6	284.4	341.2	460.3	635.6	857.8	1142.6	1440.9
		IV	279.6	231.6	204.9	212.9	263.5	314.2	419.5	559.7	739.3
0.5	E_{2sim}	I	262.6	358.2	548.2	883.8	1247.1	1709.9	2238.4	3128.2	4213.1
		II	266.7	263.7	312.7	430.3	620.7	789.8	1072.8	1468.6	1775.0
		III	274.8	251.4	270.1	327.9	442.0	616.1	820.8	1111.1	1404.6
		IV	296.9	246.8	219.2	222.8	273.9	331.8	440.8	586.7	765.7
0.6	E_{2sim}	I	230.1	310.8	463.6	754.2	1078.0	1509.3	2016.2	2669.3	3583.8
		II	248.8	244.8	283.3	403.9	565.8	730.9	1004.8	1336.5	1673.8
		III	249.3	238.5	253.4	314.6	412.2	574.3	775.3	1016.9	1336.2
		IV	303.9	256.0	226.1	231.7	283.7	343.1	456.8	600.3	783.1
0.7	E_{2sim}	I	194.5	257.1	396.7	625.2	920.4	1275.6	1753.0	2249.7	2955.3
		II	226.0	216.7	252.9	352.7	512.4	656.3	907.6	1182.0	1473.9
		III	226.1	214.6	226.1	285.9	382.3	532.1	706.8	898.9	1208.2
		IV	305.8	258.3	227.1	235.5	284.2	346.9	459.8	599.8	788.4
0.8	E_{2sim}	I	159.8	221.7	341.1	523.4	757.4	1095.9	1515.0	1960.0	2478.9
		II	193.4	190.9	228.7	320.2	438.1	580.6	825.6	1037.5	1328.2
		III	201.6	194.7	205.2	265.1	347.7	481.8	628.9	800.2	1082.0
		IV	299.9	256.9	223.5	233.7	283.7	341.6	453.7	589.5	772.5
0.9	E_{2sim}	I	136.5	186.4	289.7	440.6	635.9	939.3	1269.8	1663.2	2125.0
		II	172.9	165.9	202.6	288.7	373.1	514.3	709.8	894.3	1160.4
		III	182.2	167.1	185.0	234.8	309.8	418.6	552.9	722.3	930.8
		IV	293.8	245.8	216.8	225.3	272.8	329.7	438.3	574.2	742.7

Table 5: Monte Carlo Simulation results for population-II when the proposed estimator T^* is compared to Δ^*

$\psi_1 \downarrow$	$\psi_2 \rightarrow$		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	E_{2sim}	I	1126.40	2860.5	5849.0	9978.9	14402.0	22607.0	30230.0	40853.0	46469.0
		II	961.19	1757.9	3077.6	5323.8	7930.8	11637.0	14805.0	20847.0	26905.0
		III	274.83	264.72	298.76	362.76	515.77	742.68	1006.7	1174.6	1320.8
		IV	448.87	445.82	537.81	641.19	1000.5	1320.8	1757.2	2256.2	3038.8
0.2	E_{2sim}	I	873.59	2198.3	4489.6	7729.9	11800.0	17466.0	22954.0	31590.0	3644.3
		II	831.99	1472.2	2545.2	4305.6	6678.7	9960.1	13156.0	17250.0	23024.0
		III	302.79	284.98	314.11	406.01	562.11	821.52	995.42	1259.0	1522.1
		IV	495.59	481.24	567.79	708.65	1010.5	1426.0	1852.1	2354.0	3098.0
0.3	E_{2sim}	I	621.89	1594.20	3184.1	5627.4	8573.0	12582.0	16513.0	22385.0	27277.0
		II	682.77	1169.0	2044.1	3405.3	5386.4	7770.3	10373.0	13378.0	17978.0
		III	328.74	312.90	338.97	448.28	617.43	89.51	1079.6	1333.3	1719.8
		IV	528.81	521.64	667.01	761.28	1069.9	1502.1	1953.7	2645.4	3251.4
0.4	E_{2sim}	I	441.33	1136.90	2342.9	4039.8	6230.6	8970.8	11971.0	16010.0	20221.0
		II	540.36	905.32	1585.1	2637.0	4066.8	5938.0	8098.8	10354.0	13708.0
		III	349.27	334.32	366.96	469.80	658.16	909.27	1131.5	1455.1	1817.1
		IV	557.80	535.90	625.09	792.63	1111.7	1534.2	2022.3	2703.7	3360.2
0.5	E_{2sim}	I	325.32	829.35	1693.8	2954.8	4550.0	6503.2	8647.7	11725.0	14875.0
		II	423.09	685.55	1205.1	2062.0	3128.3	4491.7	6008.1	7843.8	10477.0
		III	358.42	347.77	382.11	498.04	683.40	938.99	1172.6	1524.7	1908.0
		IV	552.30	537.56	627.89	796.60	1104.7	1536.0	2036.20	2690.1	3371.6
0.6	E_{2sim}	I	247.94	628.85	1282.4	2233.8	3406.2	4921.7	6612.4	8869.5	11284.0
		II	326.45	531.46	954.37	1614.8	2416.2	3449.1	4720.8	6152.4	8021.9
		III	369.80	356.29	390.36	507.65	697.08	953.09	1193.9	1553.5	1966.7
		IV	545.08	519.34	607.57	778.51	1081.1	1486.7	1976.3	2607.6	3256.7
0.7	E_{2sim}	I	191.82	481.70	989.78	1738.2	2659.8	3832.4	5161.5	6844.7	8705.7
		II	256.24	421.16	747.44	1246.6	1864.4	2796.1	3789.1	4836.2	6404.1
		III	368.09	357.34	391.04	507.07	692.18	943.99	1198.0	1548.7	1972.1
		IV	523.74	448.94	569.41	738.38	1020.9	1405.1	1886.9	2452.8	3067.3
0.8	E_{2sim}	I	154.29	383.89	790.48	1385.5	2112.4	3041.20	4114.9	5376.9	6949.5
		II	206.36	335.56	604.62	1004.1	1507.5	2283.7	3062.3	3868.2	5119.9
		III	361.45	347.49	391.04	490.64	667.61	915.93	1161.0	1510.2	1915.8
		IV	488.89	463.14	526.20	689.27	941.81	1304.0	1735.1	2254.4	2837.2
0.9	E_{2sim}	I	124.89	310.43	635.21	1100.2	1714.1	2458.4	3302.5	4362.3	5601.2
		II	169.07	271.88	498.12	826.69	1245.4	1855.6	2493.5	3169.4	4211.6
		III	346.69	330.68	379.63	469.72	629.28	869.77	1114.2	1438.0	1843.1
		IV	445.87	413.45	477.73	615.16	848.82	1179.9	1569.1	2032.7	2622.9

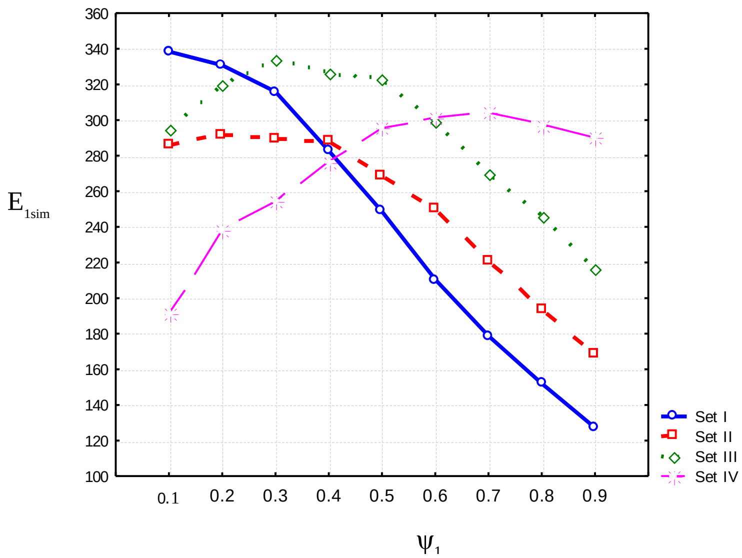


Figure 3: PRE of the estimator T^* with respect to $\hat{M}_{y(n)}$ for Population-I

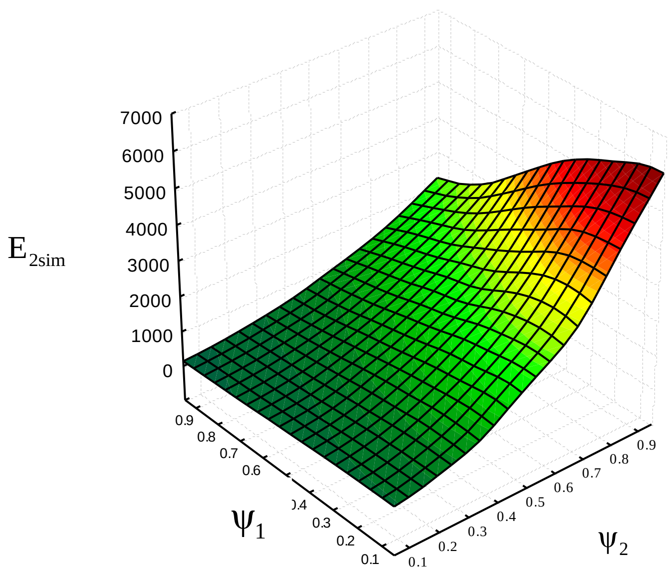


Figure 4: PRE of estimator T^* with respect to Δ^* for set-I for Population-I

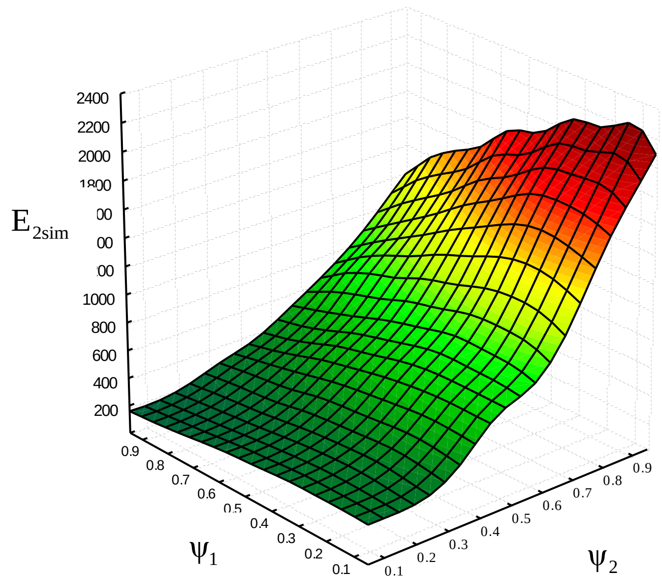


Figure 5: PRE of estimator T^* with respect to Δ^* for set-II for Population-I

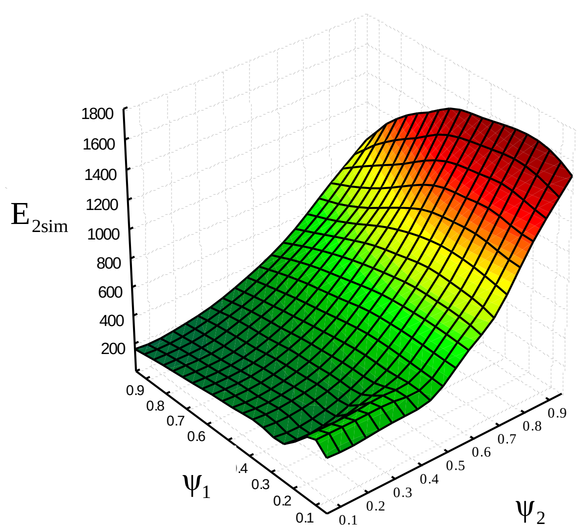


Figure 6: PRE of estimator T^* with respect to Δ^* for set-III for Population-I

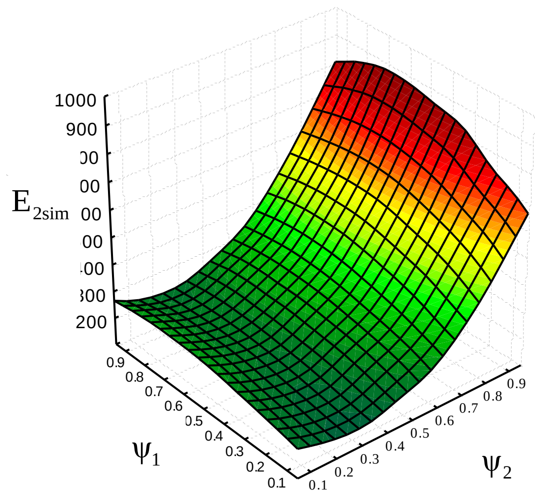


Figure 7: PRE of estimator T^* with respect to Δ^* for set-IV for Population-I

9. Analysis of Empirical and Simulation Results

1. From table 2, it is visible that the optimum values of μ (fraction of fresh sample to be drawn at current occasion) exist and this value for the estimator T is less than that of the estimator Δ for both the considered populations. This indicates that the use of additional auxiliary information at both the occasion reduces the cost of the survey.

2. Appreciable gain is observed in terms of precision indicating the proposed estimator T (at optimal condition) is preferable over the estimators $\hat{M}_{y(n)}$ and Δ (at optimal condition). This result justifies the use of additional auxiliary information at both the occasions in two-occasion successive sampling.

3. The following conclusion may be observed from Table 3 and Figure 3:

(i) For Set-I of Population-I, the value of E_{1sim} decreases as the value of ψ_1 increases. This result is expected as for Set-I, the value of μ is very less, however for Set-I of Population-II, E_{1sim} increases with the increasing value of ψ_1.

(ii) For Set-II, III and IV, of the Population-I, the value of E_{1sim} first increases and then start decreasing with the increasing value of ψ_1, however no specific pattern is observed for set II, III and IV of Population-II.

(iii) For all the considered combinations appreciable gain in precision is observed when the proposed estimator is compared with sample median estimator. Hence, the use of additional auxiliary information at both the occasions is highly justified.

4. The following points may be noted from Table 4, Table 5 and Figures 4, 5, 6 and 7:

(i) For fixed value of ψ_1 and ψ_2, the value of E_{2sim} decreases with the increasing value of μ, except for few combinations of ψ_1 and ψ_2 for Population-I, however, no specific pattern is observed for Population-II.

(ii) For fixed value of ψ_1 and μ and increasing value of ψ_2, the value of E_{2sim} also increases, except for few combinations.

(iii) For fixed value of ψ_2, and lower value of μ, the value of E_{2sim} decreases with increasing value of ψ_1 however for higher value of μ, the value of E_{2sim} increases with the increasing value of ψ_1 except for few combinations.

(iv) Tremendous gain in precision is obtained for all the considered cases.

10. Conclusion

From the analysis of empirical and simulation results it can be concluded that the proposed estimator T is favourable in terms of efficiency with respect to the standard sample median estimator, where there is no matching from previous occasion. The estimator T also proves to be much better than the estimator Δ, when no additional auxiliary information is used at any occasion. Therefore, the use of additional auxiliary information at both the occasions in two occasion successive sampling for estimating population median at current occasion is highly rewarding in terms of precision and reducing the total cost of survey. Hence, the proposed estimators may be recommended for further use by survey practitioners.

CHAPTER – 2*

A Class of Estimators for Population Median in Two Occasion Rotation Sampling

---

* Following is the publication based on the work of this chapter:--

Priyanka, K. and Mittal, R. (2015): A Class of Estimators for Population Median in Two Occasion Rotation Sampling. ***HJMS***, Vol. 44, No. 1, 189 – 202.

A Class of Estimators for Population Median in Two Occasion Rotation Sampling

1. Introduction

When both, the characteristic and the composition of the population change over time, then the cross-sectional surveys at a particular point of time become important. The survey estimates are therefore time specific, a feature that is particularly important in some context. For example, the unemployment rate is a key economic indicator that varies over time, the rate may change from one month to the next because of a change in the economy (with business laying off or recruiting new employees). To deal with such kind of circumstances, sampling is done on successive occasions with partial replacement of the units.

The problem of sampling on two successive occasions was first considered by Jessen (1942), and latter this idea was extended by Patterson (1950), Narain (1953), Eckler (1955), Gordon (1983), Arnab and Okafor (1992), Feng and Zou (1977), Singh and Singh (2001), Singh and Priyanka (2008a), Singh et al. (2012) and many others. All the above efforts were devoted to the estimation of population mean or variance on two or more occasion successive sampling.

Often, there are many practical situations where variables involved, consists of extreme values and resulting strong influence on the value of mean. In such cases the study variable is having a highly skewed distribution. For example, the study of environmental issues, the study of income as well as expenditure, the study of social evils such as abortions etc…In these situations, the mean as a measure of central tendency may not be representative of the population because it moves with the direction of asymmetry leaving the median as a better measure since it is not affected by extreme values.

Most of the studies related to median have been developed by assuming simple random sampling or its ramification in stratified random sampling (Gross (1980), Sedransk and Meyer (1978), Smith and Sedransk (1983)).

As noted earlier, a large number of estimators for estimating the population mean at current occasion have been proposed by various authors, but only a few efforts (namely Martinez-Miranda et al. (2005), Singh et al. (2007) and Rueda and Munoz (2008)) have been made to estimate the population median on current occasion in two occasion successive sampling.

The present work develops a one-parameter class of estimators that estimate the population median on the current occasion in two-occasion successive sampling. The proposed class of estimators includes some of the estimators proposed by Singh et al. (2007) for second quantile as particular cases.

Asymptotic expressions for bias and mean square error including the asymptotic convergence of the proposed class of estimators are derived. The optimum replacement strategies are discussed. The proposed class of estimators at optimum conditions is compared with sample median estimator when there is no matching from the previous occasion as well as with some of the estimators due to Singh et al. (2007) and few other members of its class. Theoretical results are justified by empirical interpretation with the help of some natural populations.

2. Sample Structure and Notations

Let $U = \left(U_1, U_2, \ldots, U_N\right)$ be the finite population of N units, which has been sampled over two occasions. It is assumed that size of the population remains unchanged but values of units change over two occasions. The character under study be denoted by x (y) on the first (second) occasions respectively. Simple random sample (without replacement) of n units is taken on the first occasion. A random subsample of $m = n\lambda$ units is retained for use on the second occasion. Now at the current occasion a simple random sample (without replacement) of u= (n-m) = nμ units is drawn afresh from the remaining (N-n) units of the population so that the sample size on the second occasion is also n. μ and λ, $(\mu + \lambda = 1)$ are the fractions of fresh and matched samples respectively at the second (current) occasion. The following notations are considered for the further use:

M_x, M_y : Population median of the variables x and y respectively.

$\hat{M}_x(n), \hat{M}_x(m), \hat{M}_y(m), \hat{M}_y(u)$: Sample medians of the respective variables shown in suffices and based on the sample sizes given in braces.

$f_x(M_x), f_y(M_y)$: The marginal densities of variables x and y respectively.

3. Proposed Class of Estimators

To estimate the population median M_y on the current (second) occasion, two independent estimators are suggested. One is based on sample of the size u= nμ drawn afresh on the current (second) occasion and which is given by

$$T_u = \hat{M}_y(u) \tag{1}$$

Second estimator is a one-parameter class of estimators based on the sample of size $m = n\lambda$ common to the both occasions and is defined as

$$T_m(d) = \hat{M}_y(m)\left[\frac{(A + C)\hat{M}_x(n) + f\,B\,\hat{M}_x(m)}{(A + f\,B)\hat{M}_x(n) + C\,\hat{M}_x(m)}\right] \tag{2}$$

$$A = (d\text{-}1)(d\text{-}2),\ B = (d\text{-}1)(d\text{-}4),\ C = (d\text{-}2)(d\text{-}3)(d\text{-}4) \text{ and } f = \frac{n}{N}$$

where d is a non-negative constant, identified to minimize the mean square error of the estimator $T_m(d)$.

Now considering the convex linear combination of the estimators T_u and $T_m(d)$, a class of estimators for M_y is proposed as

$$\hat{T}_d = \varphi\, T_u + (1\text{-}\varphi)\, T_m(d) \tag{3}$$

where φ is an unknown constant to be determined so as to minimise the mean squared error of the class of the estimators $\hat{T}_d$.

Remark 3.1: For estimating the median on each occasion, the estimator T_u is suitable, which implies that more belief on T_u could be shown by choosing φ as 1 (or close to 1), while for estimating the change from occasion to occasion, the estimator $T_m(d)$ could be more useful so φ might be chosen 0 (or close to 0). For asserting both problems simultaneously, the suitable (optimum) choice of φ is desired.

Remark 3.2: The following estimators can be identified as a particular case of the suggested class of estimators $\hat{T}_d$ to estimate population median on the current occasion in two occasion successive (rotation) sampling for different values of the unknown parameter ' d ':

$(i)\ \hat{T}_1 = \varphi_1\ T_u + (1 - \varphi_1)\ T_m(1);\ (\text{Ratio type estimator})$

$(ii)\ \hat{T}_2 = \varphi_2\ T_u + (1 - \varphi_2)\ T_m(2);\ (\text{Product type estimator})$

$(ii)\ \hat{T}_3 = \varphi_3\ T_u + (1 - \varphi_3)\ T_m(3);\ (\text{Dual to Ratio type estimator})$

where $T_m(1) = \hat{M}_y(m)\left[\dfrac{\hat{M}_x(n)}{\hat{M}_x(m)}\right],$

$$T_m(2) = \hat{M}_y(m)\left[\frac{\hat{M}_x(m)}{\hat{M}_x(n)}\right],$$

$$T_m(3) = \hat{M}_y(m)\left[\frac{n\,\hat{M}_x(n) - m\,\hat{M}_x(m)}{(n - m)\hat{M}_x(n)}\right]$$

and φ_i (i=1, 2, 3) are unknown constants to be determined so as to minimize the mean squared errors of the estimators $\hat{T}_i$ (i=1, 2, 3).

Remark 3.3: The Ratio and Product type estimators, proposed by Singh et al. (2007) for second quantile become particular cases of the proposed family of the estimators $\hat{T}_d$ for $d = 1$ and 2 respectively.

4. Properties of the Proposed Class of Estimators

The properties of the proposed class of estimators $\hat{T}_d$ are derived under the following assumptions:

(i) Population size is sufficiently large (i.e. $N \to \infty$), therefore finite population corrections are ignored.

(ii) As $N \to \infty$, the distribution of the bivariate variable (x, y) approaches a continuous distribution, which depend on population under consideration with marginal densities $f_x(.)$ and $f_y(.)$ respectively, (see Kuk and Mak(1989)).

(iii) The marginal densities $f_x(.)$ and $f_y(.)$ are positive.

(iv)The sample medians $\hat{M}_y(u), \hat{M}_y(m), \hat{M}_x(m)$ and $\hat{M}_x(n)$ are consistent and asymptotically normal (see Gross (1980)).

(v) Following Kuk and Mak (1989), P_{yx} is assumed to be the proportion of elements in the population such that $x \le \hat{M}_x$ and $y \le \hat{M}_y$.

(vi)The following large sample approximations are assumed:

$$\hat{M}_y(u) = M_y(1 + e_0),\ \hat{M}_y(m) = M_y(1 + e_1),\ \hat{M}_x(m) = M_x(1 + e_2), \hat{M}_x(n) = M_x(1 + e_3)$$

such that $|e_i| < 1 \ \forall \ i = 0, 1, 2,$ and 3.

The values of various related expectations can be seen in Allen et al. (2002) and Singh (2003). Under the above transformations, the estimators T_u and $T_m(d)$ takes the following forms:

$$T_u = M_y(1 + e_0) \tag{4}$$

$$\begin{aligned} T_m(d) = M_y[&1 + e_1 + d_1e_3 + d_2e_2 - d_3e_3 - d_4e_2 - d_1d_3e_3^2 - d_1d_4e_2e_3 - d_2d_3e_2e_3 \\ &- d_2d_4e_2^2 + d_3^2e_3^2 + d_4^2e_2^2 + 2d_3d_4e_2e_3 + (d_1 - d_3)e_1e_3 + (d_2 - d_4)e_1e_2] \end{aligned} \tag{5}$$

where $d_1 = \dfrac{A + C}{A + f B + C}$, $d_2 = \dfrac{f B}{A + f B + C}$, $d_3 = \dfrac{A + f B}{A + f B + C}$ and $d_4 = \dfrac{C}{A + f B + C}$.

Thus we have the following theorems:

Theorem 4.1: The bias of the estimator $\hat{T}_d$ to the first order of approximation is obtained as

$$B\left(\hat{T}_d\right) = (1 - \varphi)\ B\ \{T_m(d)\} \tag{6}$$

where $$B\{T_m(d)\} = \frac{1}{n}\ Q_1 + \frac{1}{m}\ Q_2 \tag{7}$$

$$Q_1 = \left(- d_1d_3 - d_1d_4 - d_2d_3 + d_3^2 + 2d_3d_4\right)\frac{\{f_x(M_x)\}^{-2}}{4\ M_x^{\ 2}}$$
$$+\ (d_1 - d_3)(P_{yx} - 0.25)\frac{\{f_y(M_y)\}^{-1}\{f_x(M_x)\}^{-1}}{M_y\ M_x}$$

and $$Q_2 = \left(- d_2d_4 + d_4^2\right)\frac{\{f_x(M_x)\}^{-2}}{4\ M_x^2}$$
$$+\ (d_2 - d_4)(P_{yx} - 0.25)\frac{\{f_y(M_y)\}^{-1}\{f_x(M_x)\}^{-1}}{M_y\ M_x}\ .$$

Proof: The bias of the estimator $\hat{T}_d$ is given by

$$B\{\hat{T}_d\} = E\ \{\hat{T}_d - M_y\}$$
$$= \varphi\ B\ \{T_u\} + (1 - \varphi)B\ \{T_m(d)\} \tag{8}$$

Since, the estimator T_u is unbiased for M_y and $T_m(d)$ is biased for M_y, so the bias of the estimator $T_m(d)$ is given by

$$B\ \{T_m(d)\} = E\ \{T_m(d) - M_y\}$$

Now, substituting the value of $T_m(d)$ from equation (5) in the above equation we get the expression for bias of $T_m(d)$ as in equation (7).

Finally substituting the value of $B\{T_m(d)\}$ in equation (8), we get the expression for the $B\{\hat{T}_d\}$ as in equation (6).

Theorem 4.2: The mean square error of the estimator $\hat{T}_d$ is given by

$$M\left(\hat{T}_d\right) = \varphi^2 \ V\left(T_u\right) + \left(1 - \varphi\right)^2 \ M\left(T_m\left(d\right)\right)_{opt.} \tag{9}$$

where $$V\left(T_u\right) = \frac{1}{u}\frac{\left\{f_y\left(M_y\right)\right\}^{-2}}{4} \tag{10}$$

and $$M\left(T_m\left(d\right)\right)_{opt.} = \frac{1}{m}A_1 + \left(\frac{1}{m} - \frac{1}{n}\right)\left\{\alpha^{*2}A_2 + 2\alpha^* A_3\right\} \tag{11}$$

where $A_1 = \dfrac{\left\{f_y\left(M_y\right)\right\}^{-2}}{4}$, $A_2 = \dfrac{\left\{f_x\left(M_x\right)\right\}^{-2}}{4}\left[\dfrac{M_y^2}{M_x^2}\right]$,

$$A_3 = \left(P_{yx} - 0.25\right)\left\{f_y\left(M_y\right)\right\}^{-1}\left\{f_x\left(M_x\right)\right\}^{-1}\left[\frac{M_y}{M_x}\right],$$

$\alpha^* = \left[\alpha\right]_{d = d_0}$, $\alpha = \left(d_2 - d_4\right) = \left(d_3 - d_1\right) = \dfrac{f\,B - C}{A + f\,B + C}$ and d_0 is the optimum value of d.

Proof: The mean square error of the estimator $\hat{T}_d$ is given by

$$M\left(\hat{T}_d\right) = E\left[\hat{T}_d - M_y\right]^2 = E\left[\varphi\left(T_u - M_y\right) + \left(1 - \varphi\right)\left\{T_m\left(d\right) - M_y\right\}\right]^2$$

$$= \varphi^2 \ V\left(T_u\right) + \left(1 - \varphi\right)^2 M\left[T_m\left(d\right)\right] + 2\,\varphi\left(1 - \varphi\right) Cov\left(T_u, T_m\left(d\right)\right) \tag{12}$$

where $$V\left(T_u\right) = E\left[T_u - M_y\right]^2 \tag{13}$$

and $$M\left[T_m\left(d\right)\right] = E\left[T_m\left(d\right) - M_y\right]^2 \tag{14}$$

As T_u and $T_m\left(d\right)$ are based on two independent samples of sizes u and m respectively, hence $Cov\left(T_u, T_m\left(d\right)\right) = 0$. Now, substituting the values of T_u and $T_m\left(d\right)$ from equations (4) and (5) in equation (13) and (14) respectively, taking expectations and ignoring finite population corrections we get the expression for $V\left(T_u\right)$ as in equation (10) and mean square error of $T_m\left(d\right)$ is obtained as

$$M\left[T_m\left(d\right)\right] = \left[\frac{1}{m}\ A_1 + \left(\frac{1}{m} - \frac{1}{n}\right)\left\{\alpha^2 A_2 + 2\,\alpha\,A_3\right\}\right]$$

where $A_1 = \frac{\left\{f_y\left(M_y\right)\right\}^{-2}}{4}$, $A_2 = \frac{\left\{f_x\left(M_x\right)\right\}^{-2}}{4}\left[\frac{M_y^2}{M_x^2}\right]$

$$A_3 = \left(P_{yx} - 0.25\right)\left\{f_y\left(M_y\right)\right\}^{-1}\left\{f_x\left(M_x\right)\right\}^{-1}\left[\frac{M_y}{M_x}\right]$$

and $\alpha = \left(d_2 - d_4\right) = \left(d_3 - d_1\right) = \frac{f B - C}{A + f B + C}$

The mean square error of the $T_m(d)$ is a function of α, which in turns is a function of d, hence it can be minimized for d, and therefore we have

$$\frac{\partial\left\{M\left[T_m(d)\right]\right\}}{\partial d} = 0$$

This gives $\alpha = \frac{-A_3}{A_2}$, assuming $\frac{\partial\alpha}{\partial d} \neq 0$ which in turns yields a cubic equation in 'd' given by

$$z_1 d^3 + z_2 d^2 + z_3 d + z_4 = 0 \qquad (15)$$

where $z_1 = \left(\frac{A_3}{A_2} - 1\right)$, $z_2 = \left(f + 9\right) + \frac{A_3}{A_2}\left(f - 8\right)$, $z_3 = \left(-5 f - 26\right) + \frac{A_3}{A_2}\left(23 - 5 f\right)$

and $z_4 = \left(4 f + 24\right) + \frac{A_3}{A_2}\left(4 f - 22\right)$.

Now for given values of $M_x, M_y, f_x\left(M_x\right)$ and $f_y\left(M_y\right)$ one will get the three optimum values of d for which $M\left[T_m(d)\right]$ attains the minimum value. The possibility of getting negative or imaginary roots cannot be ruled out. However, Singh and Shukla (1987) has pointed out that for any choice of $f, M_x, M_y, f_x\left(M_x\right)$ and $f_y\left(M_y\right)$, there exists at least one positive real root of the equation (15) ensuring that $M\left[T_m(d)\right]$ attaints its minimum within the parameter space $(0, \infty)$. Since, there may exist at most three optimum values of d, a criterion for suitable value of optimum d may be set as follows: "Out of all possible values of optimum d, choose $d = d_0$ as an adequate choice, which makes $\left|B\left[T_m(d)\right]\right|$ smallest".

Hence, the minimum mean square error of $T_m(d)$ is given by

$$M\left[T_m(d)\right]_{opt.} = \frac{1}{m} A_1 + \left(\frac{1}{m} - \frac{1}{n}\right) A_4 \qquad (16)$$

where $A_1 = \frac{\left\{f_y\left(M_y\right)\right\}^{-2}}{4}$, $\quad A_4 = \alpha^{*2} A_2 + 2\alpha^* A_3$, and $\quad \alpha^* = [\alpha]_{d = d_0}$.

Further, substituting the expression for $V(T_u)$ and $M\left[T_m(d)\right]_{opt.}$ in equation (12) we get the expression for $M(\hat{T}_d)$ as in equation (9).

Remark 4.1: The cubic equation (15) depends on the population parameters P_{yx}, $f_y(M_y)$ and $f_x(M_x)$. If these parameters are known, the proposed estimator can be easily applied. Otherwise, which is the most often situation in practice, the unknown population parameters are replaced by their sample estimates. The population proportion P_{yx} can be replaced by the sample estimate P_{yx} and the marginal densities $f_y(M_y)$ and $f_x(M_x)$ can be substituted by their kernel estimator or nearest neighbour density estimator or generalized nearest neighbour density estimator related to the kernel estimator (Silverman (1986)). Here, the marginal densities $f_y(M_y)$ and $f_x(M_x)$ are replaced by $\hat{f}_y\left(\hat{M}_y(m)\right)$ and $\hat{f}_x\left(\hat{M}_x(n)\right)$ respectively, which are obtained by method of generalized nearest neighbour density estimation related to kernel estimator.

To estimate $f_y(M_y)$ and $f_x(M_x)$, by generalized nearest neighbour density estimator related to the kernel estimator, following procedure has been adopted:

Choose an integer $h \approx n^{1/2}$ and define the distance $\delta(x_1, x_2)$ between two points on the line to be $|x_1 - x_2|$.

For $\hat{M}_x(n)$, define $\delta_1\left(\hat{M}_x(n)\right) \leq \delta_2\left(\hat{M}_x(n)\right) \leq \cdots \leq \delta_n\left(\hat{M}_x(n)\right)$ to be the distances, arranged in ascending order, from $\hat{M}_x(n)$ to the points of the sample.

The generalized nearest neighbour density estimate is defined by

$$\hat{f}\left(\hat{M}_x(n)\right) = \frac{1}{n\,\delta_h\left(\hat{M}_x(n)\right)}\sum_{i=1}^{n} K\left[\frac{\hat{M}_x(n) - x_i}{\delta_h\left(\hat{M}_x(n)\right)}\right]$$

where the kernel function K, satisfies the condition $\int_{-\infty}^{\infty} K(x)\,dx = 1$.

Here, the kernel function is chosen as Gaussian Kernel given by $K(x) = \frac{1}{2\pi}\, e^{-\left(\frac{1}{2}x^2\right)}$.

The estimate of $f_y(M_y)$ can be obtained by the above explained procedure in similar manner.

Theorem 4.3: The estimator $\hat{T}_d$, its bias and mean square error are asymptotically convergent to the estimator $\hat{T}_1$, its bias and mean square error respectively for large d.

Proof: Taking limit as $d \to \infty$ in equation (3) we get

$$\lim_{d\to\infty}\hat{T}_d = \varphi\, T_u + (1 - \varphi)\lim_{d\to\infty} T_m(d)$$

Since, $d \neq 0$, dividing numerator and denominator of the second term in R.H.S. of above equation by d^3 and taking limit as $d \to \infty$, we have

$$\lim_{d\to\infty}\hat{T}_d = \varphi\, T_u + (1 - \varphi)\, T_m(1) = \hat{T}_1$$

This is the ratio type estimator to estimate population median in two occasion rotation sampling as given in Remark 3.2. Similarly, using the expressions of bias and mean square error of the estimator $\hat{T}_d$, it is easy to see that

$$\lim_{d\to\infty} B\left\{\hat{T}_d\right\} = B\left\{\hat{T}_1\right\}$$

and

$$\lim_{d\to\infty} M\left\{\hat{T}_d\right\} = M\left\{\hat{T}_1\right\}.$$

Thus the proposed class of estimators converges to a well-defined estimator even if one chooses arbitrary, a larger value of the unknown parameter d. The bias and mean squared error also tends asymptotically to that of ratio type estimator to estimate finite population median. There is no need to bother about the existence of the estimator while choosing a larger value of d.

5. Minimum Mean Squared Error of the Proposed Class of Estimators $\hat{T}_d$

Since, mean squared error of $\hat{T}_d$ in equation (9) is function of unknown constant φ, therefore, it is minimized with respect to φ and subsequently the optimum value of φ is obtained as

$$\varphi_{opt.} = \frac{M\left\{T_m(d)\right\}_{opt.}}{V(T_u) + M\left\{T_m(d)\right\}_{opt.}} \tag{17}$$

and substituting the value of $\varphi_{opt.}$ from equation (17) in equation (9), we get the optimum mean square error of the estimator $\hat{T}_d$ as

$$M\left(\hat{T}_d\right)_{opt.} = \frac{V(T_u).\, M\left\{T_m(d)\right\}_{opt.}}{V(T_u) + M\left\{T_m(d)\right\}_{opt.}} \tag{18}$$

Further, by substituting the values from equation (10) and equation (11) in equation (18), we get the simplified value of $M\left(\hat{T}_d\right)_{opt.}$ as

$$M\left(\hat{T}_d\right)_{opt.} = \frac{A_1\left[A_1 + \mu A_4\right]}{n\left[A_1 + \mu^2 A_4\right]} \tag{19}$$

where $\mu\,(= u/n)$ is the fraction of fresh sample drawn on the current (second) occasion. Again $M\left(\hat{T}_d\right)_{opt.}$ derived in equation (19) is the function of μ. To estimate the population median on each occasion the better choice of μ is 1(case of no matching); however, to estimate the change in median from one occasion to the other, μ should be 0(case of complete matching). But intuition suggests that an optimum choice of μ is desired to devise the amicable strategy for both the problems simultaneously.

6. Optimum Replacement Policy

The key design parameter affecting the estimates of change is the overlap between successive samples. Maintaining high overlap between repeats of a survey is operationally convenient, since many sampled units have been located and have some experience in the survey. Hence to decide about the optimum value of μ (fraction of sample to be drawn afresh on current occasion) so that M_y may be estimated with maximum precision, we minimize $M(\hat{T}_d)_{opt.}$ in equation (19) with respect to μ .

The optimum value of μ so obtained is one of the two roots given by

$$\mu = \frac{-A_1 \pm \sqrt{A_1(A_1+A_4)}}{A_4} \tag{20}$$

The real value of μ exists, iff $A_1(A_1+A_4) \geq 0$. For any situation, which satisfies this condition, two real values of μ may be possible , hence in choosing a value of μ, care should be taken to ensure that $0 \leq \hat{\mu} \leq 1$, all other values of μ are inadmissible. If both the real values of μ are admissible, the lowest one will be the best choice as it reduces the total cost of the survey. Substituting the admissible value of μ say μ_0 from equation (20) in equation (19), we get the optimum value of the mean square error of the estimator $\hat{T}_d$ with respect to φ and μ both as

$$M(\hat{T}_d)_{opt.^*} = \frac{A_1[A_1 + \mu_0 A_4]}{n[A_1 + \mu_0^2 A_4]}$$

7. Efficiency Comparison

To evaluate the performance of the estimator $\hat{T}_d$, the estimator $\hat{T}_d$ at optimum conditions is compared with respect to the estimator $\hat{M}_y(n)$(the sample median), when there is no matching from previous occasion. Since, $\hat{M}_y(n)$ is unbiased for population median, its variance for large N is given by

$$V\left[\hat{M}_y(n)\right] = \frac{1}{n}\frac{\left\{f_y(M_y)\right\}^{-2}}{4} \tag{21}$$

The percent relative efficiency of the estimator $\hat{T}_d$ (under optimal condition) with respect to $\hat{M}_y(n)$ is given by

$$\text{P.R.E.}\left(\hat{T}_d, \hat{M}_y(n)\right) = \frac{V\left[\hat{M}_y(n)\right]}{M\left(\hat{T}_d\right)_{opt.^*}} \times 100 \tag{22}$$

The estimator $\hat{T}_d$ (at optimal conditions) is also compared with respect to the estimators $\hat{T}_1$, $\hat{T}_2$ and $\hat{T}_3$ respectively. Hence for large N, the expressions for optimum mean squared errors of $\hat{T}_1$, $\hat{T}_2$ and $\hat{T}_3$ are given by

$$M\left(\hat{T}_1\right)_{opt.^*} = \frac{A_1\left[A_1 + \mu_1 A_5\right]}{n\left[A_1 + \mu_1^2 A_5\right]}, \qquad M\left(\hat{T}_2\right)_{opt.^*} = \frac{A_1\left[A_1 + \mu_2 A_6\right]}{n\left[A_1 + \mu_2^2 A_6\right]}$$

and $$M\left(\hat{T}_3\right)_{opt.^*} = \frac{A_1\left[A_1 + \mu_3 A_7\right]}{n\left[A_1 + \mu_3^2 A_7\right]}$$

where $$\mu_1 = \frac{-A_1 \pm \sqrt{A_1^2 + A_1 A_5}}{A_5}, \qquad \mu_2 = \frac{-A_1 \pm \sqrt{A_1^2 + A_1 A_6}}{A_6}$$

$$\mu_3 = \frac{-A_1 \pm \sqrt{A_1^2 + A_1 A_7}}{A_7}, \qquad A_1 = \frac{\left\{f_y\left(M_y\right)\right\}^{-2}}{4},$$

$$A_5 = A_2 - 2A_3, \quad A_6 = A_2 + 2A_3 \quad \text{and} \quad A_7 = \left(\frac{f}{1+f}\right)^2 A_2 + 2\left(\frac{f}{1+f}\right) A_3 .$$

where $A_2 = \frac{\left\{f_x\left(M_x\right)\right\}^{-2}}{4}\left[\frac{M_y^2}{M_x^2}\right]$ and $A_3 = \left(P_{yx} - 0.25\right)\left\{f_y\left(M_y\right)\right\}^{-1}\left\{f_x\left(M_x\right)\right\}^{-1}\left[\frac{M_y}{M_x}\right].$

The percent relative efficiencies of $\hat{T}_d$ at optimum conditions with respect to the estimators $\hat{T}_i$ for i=1, 2 and 3 at optimum conditions are given by

$$\text{P.R.E.}\left(\hat{T}_d, \hat{T}_i\right) = \frac{M\left[\hat{T}_i\right]_{opt.^*}}{M\left(\hat{T}_d\right)_{opt.^*}} \times 100 \text{ for } i=1,2 \text{ and } 3$$

8. Numerical Illustrations

The various results obtained in previous sections are now illustrated using two natural populations.

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

In the first case, a real life situation consisting N=51 states of United States has been considered. Let y_i represent the number of abortions during 2007 in the i^{th} state of U.S. and x_i be the number of abortions during 2005 in the i^{th} state of U.S. The data are presented pictorially in Figure 8.1 as under:

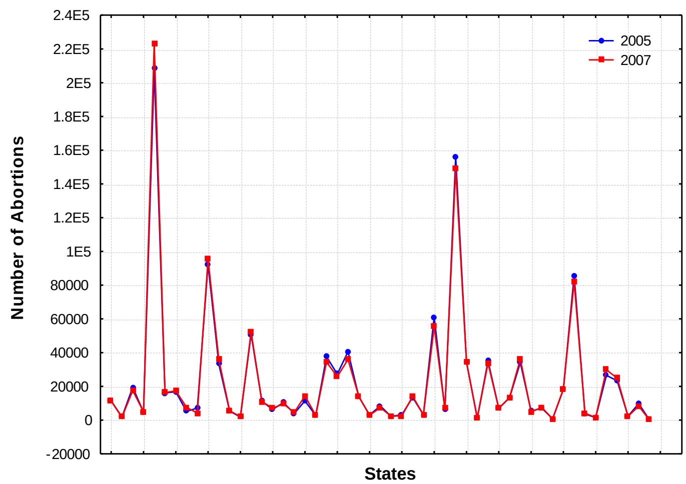


Figure 8.1 Number of Abortions during 2005 and 2007 versus different states of U.S.

Similarly in the second case, the study population consist of N=51 states of United States for year 2004. Let y_i (study variable) be the percent of bachelor degree holders or more in the year 2004 in the i^{th} state of U.S. and x_i be the percent of bachelor degree holders or more in the year 2000 in the i^{th} state of U.S. The data are represented pictorially in Figure 8.2 as under:

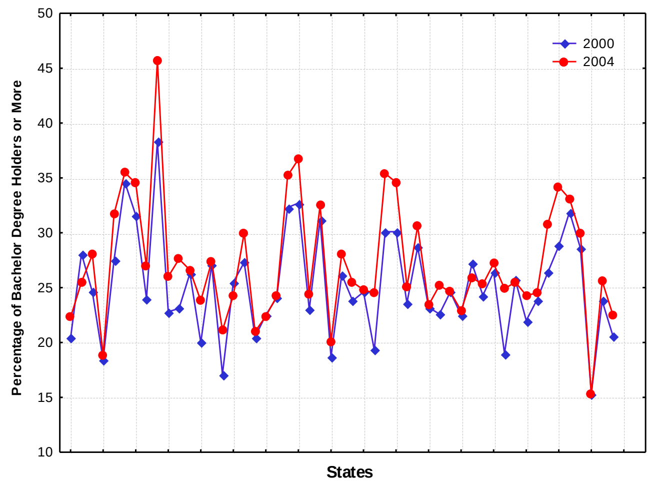


Figure 8.2 Percent of Bachelor Degree Holders or More during 2000 and 2004 versus Different States of U.S.

The graph in Figure 8.1 shows that the distribution of number of abortions in different states is skewed towards right. Similar graph is obtained for population-II as indicated in Figure 8.2. One reason of skewness may be the distribution of population in different states, that is, the states having larger populations are expected to have larger number of abortion cases and the larger percent of bachelor degree holders or more for the second case as well. Thus skewness of the data indicates that the use of median may be a good measure of central location than mean in such a situation.

Based on the above description, the descriptive statistics for both populations have been computed and are presented in Table 1.

Table 1 Descriptive Statistics for Population-I and Population-II

	Population-I		**Population-II**	
	Number of Abortions in 2005	**Number of Abortions in 2007**	**% of Bachelor Degree Holders or more in 2000**	**% of Bachelor Degree Holder or more in 2004**
Mean	23651.76	23697.65	27.19	27.17
Standard Error	5389.35	5510.75	0.65	0.75
Median	10410.00	9600.00	24.60	25.50
Standard Deviation	38487.71	39354.65	4.66	5.40
Kurtosis	12.39	14.42	0.29	1.67
Skewness	3.31	3.52	0.40	0.89
Minimum	70.00	90.00	15.30	15.30
Maximum	208430.00	223180.00	30.30	45.70

For the two populations under consideration, the cubic equation (15) is solved for d for some choices of "f". The optimum mean square errors of the proposed class of estimators are found to be same for all the three values of "d" obtained. So, using the criteria set in the proof of theorem 4.1, Table 2 shows the best choice of the optimum value of "d" for different choices of "f" for both, Population-I and Population-II.

Table 2: Best choice of d for Population-I and Population-II, for different choices of f

f	**Population-I**			**Population-II**		
	d	$\lvert Bias \rvert$	d_0	d	$\lvert Bias \rvert$	d_0
0.9800	10.0002 2.4170 1.4705	3.6526 0.3097 4.1206	**2.4170**	22.8356 2.3533 1.2030	0.1419 0.1089 0.1467	**2.3553**
0.1960	10.7520 2.6449 1.3740	1.8948 1.2919 2.1515	**2.6449**	2.5878 25.5834 1.1537	1.3940 0.0702 0.0748	**25.5834**
0.2941	11.5280 2.8115 1.3146	1.3005 1.5131 1.4675	**11.5280**	28.3715 2.7621 1.1244	0.0486 0.1526 0.0504	**28.3715**
0.3922	12.3230 2.9414 1.2729	0.9984 1.5271 1.1168	**12.3230**	31.1885 2.8979 1.1047	0.0367 0.1562 0.0381	**31.1885**
0.4902	13.1327 3.0462 1.2417	0.8141 1.4584 0.9026	**13.1327**	34.0268 3.0070 1.0905	0.0296 0.1532 0.0306	**34.0268**

Table 3: Optimum value of μ and percent relative efficiencies of $\hat{T}_d$ at optimum conditions With respect to $\hat{M}_y(n)$ and $\hat{T}_i$ for i=1, 2 and 3 at optimum conditions

	Population-I	Population-II
f	0.9800	0.9800
d_0	2.4170	2.3553
μ_0	0.6800	0.6271
P.R.E. $\left(\hat{T}_d, \hat{M}_y(n)\right)$	136.00	125.41
P.R.E. $(\hat{T}_d, \hat{T}_1)$	103.33	100.16
P.R.E. $(\hat{T}_d, \hat{T}_2)$	206.73	173.48
P.R.E. $(\hat{T}_d, \hat{T}_3)$	128.93	120.81

9. Interpretation of Results and Conclusion

(i) From Table 2, it can clearly be seen that the real optimum value of d always exists for both the considered populations. This justifies the feasibility of the proposed class of estimators $\hat{T}_d$.

(ii) From Table 3, it can be seen that the optimum value of μ also exist for both the considered populations. Hence, it indicates that the proposed class of estimators $\hat{T}_d$ is quite feasible under optimal conditions.

(iii) Table 3 indicates that the proposed class of estimators $\hat{T}_d$ at optimum conditions is highly preferable over sample median estimator $\hat{M}_y(n)$. It also performs better than the estimators $\hat{T}_1$ and $\hat{T}_2$ which are the estimators proposed by Singh et al. (2007) for second quantile. It also proves to be highly efficient than the estimator $\hat{T}_3$ which is a Dual to Ratio type estimator, a member of its own class.

Hence, it can be concluded that the estimation of median at current occasion is certainly feasible in two occasion successive sampling. The enchanting convergence property of proposed class of estimators $\hat{T}_d$ justifies the incorporation of unknown parameter in the structure of proposed class of estimators, since the optimum value of the parameter always exists. Hence the proposed class of estimators $\hat{T}_d$ can be recommended for its further use by survey practitioners.

CHAPTER – 3*

Searching Effective Rotation Patterns for Population Median using Exponential Type Estimators in Two-Occasion Rotation Sampling

* Following is the publication based on the work of this chapter:--

1. Priyanka, K. and Mittal, R. (2016): Searching Effective Rotation Patterns for Population
Median using Exponential Type Estimators in Two-Occasion Rotation Sampling. ***Communication in Statistics (Theory and Methods),*** DOI: 10.1080/03610926.2014.944661.

Searching Effective Rotation Patterns for Population Median using Exponential Type Estimators in Two-Occasion Rotation Sampling

1. Introduction

When both, the characteristic and the composition of the population change over time, then the cross-sectional surveys at a particular point of time become important. The survey estimates are therefore time specific, a feature that is particularly important in some context. For example, the unemployment rate is a key economic indicator that varies over time, the rate may change from one month to the next because of a change in the economy (with business laying off or recruiting new employees). To deal with such kind of circumstances, sampling is done on successive occasions with partial replacement of the units.

The problem of sampling on two successive occasions was first considered by Jessen (1942), latter this idea was extended by Patterson (1950), Narain (1953), Singh and Priyanka (2008a), Singh et al. (2013a) and many others.

All the above efforts were devoted to the estimation of population mean or variance on two or more occasion successive sampling while there are many practical situations where variables involved, consists of extreme values and resulting strong influence on the value of mean. In such cases the study variable is having highly skewed distribution and mean may offer the result not enough to be representative because it moves with the direction of asymmetry. The median, on the other hand does not suffer from extreme values.

Most of the studies related to median have been developed by assuming simple random sampling or its ramification in stratified random sampling (Gross (1980), Sedransk and Meyer (1978), Smith and Sedransk (1983)) considering only the variable of interest without making explicit use of auxiliary variables. It is well known that the use of auxiliary information at the estimation stage can typically increase the precision of estimates of a parameter.

Exponential type estimators also play a vital role in increasing the precision of the estimates. Bahl and Tuteja (1991) was the first to propose the exponential ratio and product type estimators for the estimation of finite population mean.

Motivated with their work, the present work develops more effective and relevant estimators using exponential ratio type estimators for population median at current occasion in two occasion successive sampling. Properties of the proposed estimators are discussed. Optimum replacement strategies are elaborated for the proposed estimators.

Proposed estimators at optimum conditions are compared with the sample median estimator when there is no matching from the previous occasion as well with the ratio type estimator proposed by Singh et al. (2007) for second quantile, when no additional auxiliary information was used at any occasion. The behaviours of the proposed estimator are justified by empirical interpretations and validated by the means of simulation study with the help of some natural populations.

2. Sample Structure and Notations

Let $U = (U_1, U_2, \ldots, U_N)$ be the finite population of N units, which has been sampled over two occasions. It is assumed that size of the population remains unchanged but values of units change over two occasions. The character under study be denoted by x (y) on the first (second) occasions respectively. It is assumed that information on an auxiliary variable z, whose population median is known and stable over occasions is readily available on both the occasions and highly correlated to x and y respectively. Simple random sample (without replacement) of n units is taken on the first occasion. A random subsample of $m = n\lambda$ units is retained for use on the second occasion. Now at the current occasion a simple random sample (without replacement) of u= (n-m) = nμ units is drawn afresh from the remaining (N-n) units of the population so that the sample size on the second occasion is also $n.\ \mu$ and $\lambda(\mu + \lambda = 1)$ are the fractions of fresh and matched samples respectively at the second (current) occasion. The following notations are considered for the further use:

M_x, M_y, M_z : Population median of the variables x, y and z respectively.

$\hat{M}_y(u), \hat{M}_z(u)$: Sample median of variables y and z based on the sample size u.

$\hat{M}_x(m), \hat{M}_y(m), \hat{M}_z(m)$: Sample median of variables x, y and z based on the sample size m.

$\hat{M}_x(n), \hat{M}_z(n)$: Sample medians of variables x and z based on the sample size n.

$f_x(M_x), f_y(M_y), f_z(M_z)$: The marginal densities of variables x, y and z respectively.

3. Proposed Estimators T_{ij} (i, j =1, 2)

To estimate the population median M_y on the current (second) occasion, two sets of estimators have been proposed utilizing the concept of exponential ratio type estimators. First set of estimators $\{T_{1u}, T_{2u}\}$ is based on sample of the size u= nμ drawn afresh on the current

(second) occasion and the second set of estimators $\{T_{1m}, T_{2m}\}$ is based on sample size $m = n\lambda$ common to the both occasions. The two sets of the proposed estimators are given as

$$T_{1u} = M_z \left(\frac{\hat{M}_y(u)}{\hat{M}_z(u)} \right) \tag{1}$$

$$T_{2u} = \hat{M}_y(u) \exp\left(\frac{M_z - \hat{M}_z(u)}{M_z + \hat{M}_z(u)} \right) \tag{2}$$

$$T_{1m} = \hat{M}_x(n) \left(\frac{\hat{M}_y(m)}{\hat{M}_x(m)} \right) \exp\left(\frac{M_z - \hat{M}_z(m)}{M_z + \hat{M}_z(m)} \right) \tag{3}$$

$$T_{2m} = \hat{M}_x^*(n) \left(\frac{\hat{M}_y^*(m)}{\hat{M}_x^*(m)} \right) \tag{4}$$

where $\hat{M}_y^*(m) = \hat{M}_y(m) \exp\left(\frac{M_z - \hat{M}_z(m)}{M_z + \hat{M}_z(m)} \right), \hat{M}_x^*(m) = \hat{M}_x(m) \exp\left(\frac{M_z - \hat{M}_z(m)}{M_z + \hat{M}_z(m)} \right)$

and $\hat{M}_x^*(n) = \hat{M}_x(n)\ \exp\left(\frac{M_z - \hat{M}_z(n)}{M_z + \hat{M}_z(n)}\right).$

Considering the convex linear combination of the two sets of estimators $T_{iu}\ (i = 1, 2)$ and $T_{jm}\ (j = 1, 2)$, we have the final estimators of population median M_y on the current occasion as

$$T_{ij} = \varphi_{ij} T_{iu} + \left(1 - \varphi_{ij}\right) T_{jm}\ ;\ (i, j = 1, 2) \tag{5}$$

where $\varphi_{ij}\ (i, j = 1, 2)$ are the unknown constants to be determined so as to minimise the mean squared error of the estimators T_{ij} (i, j=1, 2).

Remark 3.1: For estimating the median on each occasion, the estimators $T_{iu}\ (i = 1, 2)$ are suitable, which implies that more belief on T_{iu} could be shown by choosing $\varphi_{ij}\ (i, j = 1, 2)$ as 1 (or close to 1), while for estimating the change from occasion to occasion, the estimators $T_{jm}\ (j=1, 2)$ could be more useful so φ_{ij} might be chosen as 0 (or close to 0). For asserting both problems simultaneously, the suitable (optimum) choices of φ_{ij} are desired.

4. Properties of the Proposed Estimators T_{ij} (i, j =1, 2)

4.1. Assumptions

The properties of the proposed estimators $T_{ij}\ (i, j = 1, 2)$ are derived under the following assumptions:

(i) Population size is sufficiently large (i.e. $N \to \infty$), therefore finite population corrections are ignored.

(ii) As $N \to \infty$, the distribution of the bivariate variable (a, b) where a and $b \in \{x, y, z\}$ and $a \neq b$ approaches a continuous distribution with marginal densities $f_a(.)$ and $f_b(.)$ respectively, (see Kuk and Mak (1989)).

(iii) The marginal densities $f_x(.)$, $f_y(.)$ and $f_z(.)$ are positive.

(iv) The sample medians $\hat{M}_y(u), \hat{M}_y(m), \hat{M}_x(m), \hat{M}_x(n), \hat{M}_z(u), \hat{M}_z(m)$ and $\hat{M}_z(n)$ are consistent and asymptotically normal (see Gross (1980)).

(v) Following Kuk and Mak (1989), let P_{ab} be the proportion of elements in the population such that $a \le \hat{M}_a$ and $b \le \hat{M}_b$ where $a \text{ and } b \in \{x, y, z\}$ and $a \neq b$.

(vi) Following large sample approximations are assumed:

$\hat{M}_y(u) = M_y(1 + e_0)$, $\hat{M}_y(m) = M_y(1 + e_1)$, $\hat{M}_x(m) = M_x(1 + e_2)$, $\hat{M}_x(n) = M_x(1 + e_3)$ $\hat{M}_z(u) = M_z(1 + e_4)$, $\hat{M}_z(m) = M_z(1 + e_5)$ and $\hat{M}_z(n) = M_z(1 + e_6)$ such that $|e_i| < 1$ $\forall$ i = 0, 1, 2, 3, 4, 5 and 6 .

The values of various related expectations can be seen in Allen et al. (2002) and Singh (2003).

4.2. Bias and Mean Square Error of the Estimators T_{ij} (i, j =1, 2)

The estimators T_{iu} and T_{jm} $(i, j=1, 2)$ are ratio, exponential ratio, ratio to exponential ratio and chain type ratio to exponential ratio type in nature respectively. Hence they are biased for population median M_y. Therefore, the final estimators $T_{ij}(i, j=1, 2)$ defined in equation (5) are also biased estimators of M_y. Bias $B(.)$ and mean square errors $M(.)$ of the proposed estimators $T_{ij}(i, j=1, 2)$ are obtained up to first order of approximations and thus we have following theorems:

Theorem 4.2.1. Bias of the estimators $T_{ij}(i, j=1, 2)$ to the first order of approximations are obtained as

$$B(T_{ij}) = \varphi_{ij}\, B(T_{iu}) + (1 - \varphi_{ij})\, B(T_{jm});\ (i, j=1,2) \tag{6}$$

where
$$B(T_{1u}) = \frac{1}{u}\left(\frac{\left[f_z(M_z)\right]^{-2} M_y}{4\, M_z^2} - \frac{(4\, P_{yz} - 1)\left[f_y(M_y)\right]^{-1}\left[f_z(M_z)\right]^{-1}}{4\, M_z}\right) \tag{7}$$

$$B\left(T_{2u}\right) = \frac{1}{u}\left(\frac{3\left[f_z\left(M_z\right)\right]^{-2} M_y}{32\, M_z^2} - \frac{\left(4\,P_{yz} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_z\left(M_z\right)\right]^{-1}}{8\, M_z}\right) \tag{8}$$

$$\begin{aligned} B\left(T_{1m}\right) = &\frac{1}{m}\left(\frac{\left[f_x\left(M_x\right)\right]^{-2} M_y}{4\, M_x^2} + \frac{3\left[f_z\left(M_z\right)\right]^{-2} M_y}{32\, M_z^2} - \frac{\left(4\,P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1}}{4\, M_x}\right. \\ &\left. - \frac{\left(4\,P_{yz} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_z\left(M_z\right)\right]^{-1}}{8\, M_z} + \frac{\left(4\,P_{xz} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_z\left(M_z\right)\right]^{-1} M_y}{8\, M_x\, M_z}\right) \\ &+ \frac{1}{n}\left(\frac{\left(4\,P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1}}{4\, M_x} - \frac{\left(4\,P_{xz} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_z\left(M_x\right)\right]^{-1} M_y}{8\, M_x\, M_z}\right. \\ &\left. - \frac{\left[f_x\left(M_x\right)\right]^{-2} M_y}{4\, M_x^2}\right) \end{aligned} \tag{9}$$

$$\begin{aligned} B\left(T_{2m}\right) = &\frac{1}{m}\left(\frac{\left[f_x\left(M_x\right)\right]^{-2} M_y}{4\, M_x^2} - \frac{\left(4\,P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1}}{4\, M_x}\right) + \frac{1}{n}\left(\frac{3\left[f_z\left(M_z\right)\right]^{-2} M_y}{32\, M_z^2}\right. \\ &\left. + \frac{\left(4\,P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1}}{4\, M_x} - \frac{\left[f_x\left(M_x\right)\right]^{-2} M_y}{4\, M_x^2} - \frac{\left(4\,P_{yz} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_z\left(M_z\right)\right]^{-1}}{8\, M_z}\right) \end{aligned} \tag{10}$$

Proof: The bias of the estimators $T_{ij}\left(i, j = 1, 2\right)$ are given by

$$B\left(T_{ij}\right) = E\left[T_{ij} - M_y\right] = \varphi_{ij}\, B\left(T_{iu}\right) + \left(1 - \varphi_{ij}\right) B\left(T_{jm}\right)$$

where $B\left(T_{iu}\right) = E\left[T_{iu} - M_y\right]$ and $B\left(T_{jm}\right) = E\left[T_{jm} - M_y\right]$

Using large sample approximations assumed in Section 4.1 and retaining terms upto the first order of approximations, the expression for $B\left(T_{iu}\right)$ and $B\left(T_{jm}\right)$ are obtained as in equations (7) - (10) and hence the expression for bias of the estimators $T_{ij}\left(i, j = 1, 2\right)$ are obtained as in equation (6).

Theorem 4.2.2. Mean square errors of the estimators $T_{ij}(i, j = 1, 2)$ to the first order of approximations are obtained as

$$M(T_{ij}) = \varphi_{ij}^2 M(T_{iu}) + (1 - \varphi_{ij})^2 M(T_{jm}) + 2\varphi_{ij}(1 - \varphi_{ij})Cov(T_{iu}, T_{jm}); \text{ (i, j=1,2)} \quad (11)$$

where $M(T_{1u}) = \frac{1}{u} A_1$ (12)

$$M(T_{2u}) = \frac{1}{u} A_2 \quad (13)$$

$$M(T_{1m}) = \frac{1}{m} A_3 + \frac{1}{n} A_4 \quad (14)$$

$$M(T_{2m}) = \frac{1}{m} A_5 + \frac{1}{n} A_6 \quad (15)$$

$$A_1 = \left\{ \frac{[f_y(M_y)]^{-2}}{4} + \frac{[f_z(M_z)]^{-2} M_y^2}{4 M_z^2} - \frac{(4P_{yz} - 1)[f_y(M_y)]^{-1}[f_z(M_z)]^{-1} M_y}{2 M_z} \right\},$$

$$A_2 = \left\{ \frac{[f_y(M_y)]^{-2}}{4} + \frac{[f_z(M_z)]^{-2} M_y^2}{16 M_z^2} - \frac{(4P_{yz} - 1)[f_y(M_y)]^{-1}[f_z(M_z)]^{-1} M_y}{4 M_z} \right\},$$

$$A_3 = \left\{ \frac{[f_y(M_y)]^{-2}}{4} + \frac{[f_x(M_x)]^{-2} M_y^2}{4 M_x^2} + \frac{[f_z(M_z)]^{-2} M_y^2}{16 M_z^2} - \frac{(4P_{xy} - 1)[f_x(M_x)]^{-1}[f_y(M_y)]^{-1} M_y}{2 M_x} \right.$$
$$\left. - \frac{(4P_{yz} - 1)[f_y(M_y)]^{-1}[f_z(M_z)]^{-1} M_y}{4 M_z} + \frac{(4P_{xz} - 1)[f_x(M_x)]^{-1}[f_z(M_z)]^{-1} M_y^2}{4 M_x M_z} \right\},$$

$$A_4 = \left\{ \frac{(4P_{xy} - 1)[f_x(M_x)]^{-1}[f_y(M_y)]^{-1} M_y}{2 M_x} - \frac{(4P_{xz} - 1)[f_x(M_x)]^{-1}[f_z(M_z)]^{-1} M_y^2}{4 M_x M_z} \right.$$
$$\left. - \frac{[f_x(M_x)]^{-2} M_y^2}{4 M_x^2} \right\},$$

$$A_5 = \left\{ \frac{\left[f_y\left(M_y\right)\right]^{-2}}{4} + \frac{\left[f_x\left(M_x\right)\right]^{-2} M_y^2}{4\,M_x^2} - \frac{\left(4\,P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1} M_y}{2\,M_x} \right\}$$

and

$$A_6 = \left\{ \frac{\left[f_z\left(M_z\right)\right]^{-2} M_y^2}{16\,M_z^2} - \frac{\left[f_x\left(M_x\right)\right]^{-2} M_y^2}{4\,M_x^2} + \frac{\left(4\,P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1} M_y}{2\,M_x} \right.$$
$$\left. - \frac{\left(4\,P_{yz} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_z\left(M_z\right)\right]^{-1} M_y}{4\,M_z} \right\}.$$

Proof: The mean square errors of the estimators T_{ij} are given by

$$M\left(T_{ij}\right) = E\left[T_{ij} - M_y\right]^2 = E\left[\varphi_{ij}\left(T_{iu} - M_y\right) + \left(1 - \varphi_{ij}\right)\left\{T_{jm} - M_y\right\}\right]^2$$
$$= \varphi_{ij}^{\,2}\, M\left(T_{iu}\right) + \left(1 - \varphi_{ij}\right)^2 M\left[T_{jm}\right] + 2\,\varphi_{ij}\left(1 - \varphi_{ij}\right) \mathrm{Cov}\left(T_{iu}, T_{jm}\right)$$

where $M\left(T_{iu}\right) = E\left[T_{iu} - M_y\right]^2$ and $M\left[T_{jm}\right] = E\left[T_{jm} - M_y\right]^2$; $\quad$ (i, j=1, 2)

The estimators T_{iu} and T_{jm} are based on two independent samples of sizes u and m respectively, hence $\mathrm{Cov}\left(T_{iu}, T_{jm}\right) = 0$; (i, j = 1, 2). Using large sample approximations assumed in section 4.1 and retaining terms upto the first order of approximations, the expression for $M\left(T_{iu}\right)$ and $M\left(T_{jm}\right)$ are obtained as given in equations (12) - (15) and hence the expressions for mean square error of estimators $T_{ij}\left(i, j = 1, 2\right)$ are obtained.

Remark 4.2.1: The mean square errors of the estimators T_{ij}(i, j=1, 2) in equation (11) depend on the population parameters P_{xy}, P_{yz}, P_{xz}, $f_x\left(M_x\right)$, $f_y\left(M_y\right)$ and $f_z\left(M_z\right)$. If these parameters are known, the properties of proposed estimators can be easily studied. Otherwise, which is the most often situation in practice, the unknown population parameters are replaced by their sample estimates. The population proportions P_{xy}, P_{yz} and P_{xz} can be replaced by the sample estimate P_{xy}, P_{yz} and P_{xz} and the marginal densities

$f_y(M_y)$, $f_x(M_x)$ and $f_z(M_z)$ can be substituted by their kernel estimator or nearest neighbour density estimator or generalized nearest neighbour density estimator related to the kernel estimator (Silverman (1986)). Here, the marginal densities $f_y(M_y)$, $f_x(M_x)$ and $f_z(M_z)$ are replaced by $\hat{f}_y(\hat{M}_y(m))$, $\hat{f}_x(\hat{M}_x(n))$ and $\hat{f}_z(\hat{M}_z(n))$ respectively, which are obtained by method of generalized nearest neighbour density estimation related to kernel estimator.

To estimate $f_y(M_y)$, $f_x(M_x)$ and $f_z(M_z)$, by generalized nearest neighbour density estimator related to the kernel estimator, following procedure has been adopted:

Choose an integer $h \approx n^{1/2}$ and define the distance $d(x_1, x_2)$ between two points on the line to be $|x_1 - x_2|$.

For $\hat{M}_x(n)$, define $d_1(\hat{M}_x(n)) \le d_2(\hat{M}_x(n)) \le --- \le d_n(\hat{M}_x(n))$ to be the distances, arranged in ascending order, from $\hat{M}_x(n)$ to the points of the sample.

The generalized nearest neighbour density estimate is defined by

$$\hat{f}(\hat{M}_x(n)) = \frac{1}{n\, d_h(\hat{M}_x(n))} \sum_{i=1}^{n} K\left[\frac{\hat{M}_x(n) - x_i}{d_h(\hat{M}_x(n))}\right]$$

where the kernel function K, satisfies the condition $\int_{-\infty}^{\infty} K(x)\, dx = 1$.

Here, the kernel function is chosen as Gaussian Kernel given by $K(x) = \frac{1}{2\pi} e^{-\left(\frac{1}{2}x^2\right)}$.

The estimate of $f_y(M_y)$ and $f_z(M_z)$ can be obtained by the above explained procedure in similar manner.

5. Minimum Mean Square Errors of the Proposed Estimators T_{ij} (i, j =1, 2)

Since the mean square errors of the estimators $T_{ij}(i, j = 1, 2)$ given in equation (11) are the functions of unknown constants $\varphi_{ij}(i, j = 1, 2)$, therefore, they are minimized with respect to φ_{ij} and subsequently the optimum values of φ_{ij} are obtained as

$$\varphi_{i\,j_{opt.}} = \frac{M\left(T_{j\,m}\right)}{M\left(T_{i\,u}\right) + M\left(T_{j\,m}\right)};\ (i, j = 1, 2) \qquad (16)$$

Now substituting the values of $\varphi_{i\,j_{opt.}}$ in equation (11), we obtain the optimum mean square errors of the estimators $T_{i\,j}\left(i, j = 1, 2\right)$ as

$$M\left(T_{i\,j}\right)_{opt.} = \frac{M\left(T_{i\,u}\right) . M\left(T_{j\,m}\right)}{M\left(T_{i\,u}\right) + M\left(T_{j\,m}\right)};\ \left(i, j = 1, 2\right) \qquad (17)$$

Further, substituting the values of the mean square error of the estimators defined in equation (12) to equation (15) in equation (16) and (17), the simplified values $\varphi_{i\,j_{opt.}}$ and $M\left(T_{i\,j}\right)_{opt.}$ are obtained as

$$\varphi_{11_{opt.}} = \frac{\mu_{11}\left[\mu_{11}\ A_4 - \left(A_3 + A_4\right)\right]}{\left[\mu_{11}^2\ A_4 - \mu_{11}\left(A_3 + A_4 - A_1\right) - A_1\right]} \qquad (18)$$

$$\varphi_{12_{opt.}} = \frac{\mu_{12}\left[\mu_{12}\ A_6 - \left(A_5 + A_6\right)\right]}{\left[\mu_{12}^2\ A_6 - \mu_{12}\left(A_5 + A_6 - A_1\right) - A_1\right]} \qquad (19)$$

$$\varphi_{21_{opt.}} = \frac{\mu_{21}\left[\mu_{21}\ A_4 - \left(A_3 + A_4\right)\right]}{\left[\mu_{21}^2\ A_4 - \mu_{21}\left(A_3 + A_4 - A_2\right) - A_2\right]} \qquad (20)$$

$$\varphi_{22_{opt.}} = \frac{\mu_{22}\left[\mu_{22}\ A_6 - \left(A_5 + A_6\right)\right]}{\left[\mu_{22}^2\ A_6 - \mu_{22}\left(A_5 + A_6 - A_2\right) - A_2\right]} \qquad (21)$$

$$M\left(T_{11}\right)_{opt.} = \frac{1}{n}\frac{\left[\mu_{11}\ C_1 - C_2\right]}{\left[\mu_{11}^2\ A_4 - \mu_{11}\ C_3 - A_1\right]} \qquad (22)$$

$$M\left(T_{12}\right)_{opt.} = \frac{1}{n}\frac{\left[\mu_{12}\ C_4 - C_5\right]}{\left[\mu_{12}^2\ A_6 - \mu_{12}\ C_6 - A_1\right]} \qquad (23)$$

$$M\left(T_{21}\right)_{opt.} = \frac{1}{n}\frac{\left[\mu_{21}\ C_7 - C_8\right]}{\left[\mu_{21}^2\ A_4 - \mu_{21}\ C_9 - A_2\right]} \qquad (24)$$

$$M(T_{22})_{opt.} = \frac{1}{n}\frac{[\mu_{22}\ C_{10} - C_{11}]}{[\mu_{22}^2\ A_6 - \mu_{22}\ C_{12} - A_2]} \quad (25)$$

where

$C_1 = A_1A_4$, $C_2 = A_1A_3 + A_1A_4$, $C_3 = A_3 + A_4 - A_1$, $C_4 = A_1A_6$, $C_5 = A_1A_5 + A_1A_6$,

$C_6 = A_5 + A_6 - A_1$, $C_7 = A_2A_4$, $C_8 = A_2A_3 + A_2A_4$, $C_9 = A_3 + A_4 - A_2$, $C_{10} = A_2A_6$

$C_{11} = A_2A_5 + A_2A_6$, $C_{12} = A_5 + A_6 - A_2$ and $\mu_{ij}(i, j = 1, 2)$ are the fractions of the sample drawn afresh at the current(second) occasion.

Remark 5.1: $M(T_{ij})_{opt.}$ derived in equation (22) - (25) are the functions of $\mu_{ij}(i, j = 1, 2)$. To estimate the population median on each occasion the better choices of $\mu_{ij}(i, j = 1, 2)$ are 1(case of no matching); however, to estimate the change in median from one occasion to other, $\mu_{ij}(i, j = 1, 2)$ should be 0(case of complete matching). But intuition suggests that the optimum choices of $\mu_{ij}(i, j = 1, 2)$ are desired to devise the amicable strategy for both the problems simultaneously.

6. Optimum Replacement Strategies for the Estimators T_{ij} (i, j =1, 2)

The key design parameter affecting the estimates of change is the overlap between successive samples. Maintaining high overlap between repeats of a survey is operationally convenient, since many sampled units have been located and have some experience in the survey. Hence to decide about the optimum value of $\mu_{ij}(i, j = 1, 2)$ (fractions of samples to be drawn afresh on current occasion) so that M_y may be estimated with maximum precision and minimum cost, we minimize the mean square errors $M(T_{ij})_{opt.}(i, j = 1, 2)$ in equation (22) to (25) with respect to $\mu_{ij}(i, j = 1, 2)$ respectively.

The optimum value of $\mu_{ij}(i, j = 1, 2)$ so obtained is one of the two roots given by

$$\mu_{11} = \frac{D_2 \pm \sqrt{D_2^2 - D_1\ D_3}}{D_1} \quad (26)$$

$$\mu_{12} = \frac{D_5 \pm \sqrt{D_5^2 - D_4 D_6}}{D_4} \tag{27}$$

$$\mu_{21} = \frac{D_8 \pm \sqrt{D_8^2 - D_7 D_9}}{D_7} \tag{28}$$

$$\mu_{22} = \frac{D_{11} \pm \sqrt{D_{11}^2 - D_{10} D_{12}}}{D_{10}} \tag{29}$$

where

$D_1 = A_4C_1,\ D_2 = A_4C_2,\ D_3 = A_1C_1 + C_2C_3,\ D_4 = A_6C_4,\ D_5 = A_6C_5,\ D_6 = A_1C_4 + C_5C_6$

$D_7 = A_4C_7,\ D_8 = A_4C_8,\ D_9 = A_2C_7 + C_8C_9,\ D_{10} = A_6C_{10},\ D_{11} = A_6C_{11}$ and $D_{12} = A_2C_{10} + C_{11}C_{12}$.

The real values of $\mu_{ij}(i, j = 1, 2)$ exist, iff $D_2^2 - D_1 D_3 \geq 0,\ D_5^2 - D_4 D_6 \geq 0,\ D_8^2 - D_7 D_9 \geq 0,$ and $D_{11}^2 - D_{10} D_{12} \geq 0$. For any situation, which satisfies these conditions, two real values of $\mu_{ij}(i, j = 1, 2)$ may be possible , hence to choose a value of $\mu_{ij}(i, j = 1, 2)$, it should be taken care of that $0 \leq \mu_{ij} \leq 1$, all other values of $\mu_{ij}(i, j = 1, 2)$ are inadmissible. If both the real values of $\mu_{ij}(i, j = 1, 2)$ are admissible, the lowest one will be the best choice as it reduces the total cost of the survey. Substituting the admissible value of μ_{ij} say $\mu_{ij}^{(0)}(i, j = 1, 2)$ from equation (26) to (29) in equation (22) to (25) respectively , we get the optimum values of the mean square errors of the estimators $T_{ij}(i, j = 1, 2)$ with respect to φ_{ij} as well as $\mu_{ij}(i, j = 1, 2)$ which are given as

$$M(T_{11})^*_{opt.} = \frac{\left[\mu_{11}^{(0)} C_1 - C_2\right]}{n\left[\mu_{11}^{(0)2} A_4 - \mu_{11}^{(0)} C_3 - A_1\right]} \tag{30}$$

$$M(T_{12})^*_{opt.} = \frac{\left[\mu_{12}^{(0)} C_4 - C_5\right]}{n\left[\mu_{12}^{(0)2} A_6 - \mu_{12}^{(0)} C_6 - A_1\right]} \tag{31}$$

$$M(T_{21})^*_{opt.} = \frac{\left[\mu_{21}^{(0)} C_7 - C_8\right]}{n\left[\mu_{21}^{(0)2} A_4 - \mu_{21}^{(0)} C_9 - A_2\right]} \tag{32}$$

$$M(T_{22})^*_{opt.} = \frac{\left[\mu_{22}^{(0)} C_{10} - C_{11}\right]}{n\left[\mu_{22}^{(0)2} A_6 - \mu_{22}^{(0)} C_{12} - A_2\right]} \tag{33}$$

7. Efficiency Comparison

To evaluate the performance of the proposed estimators, the estimators $T_{ij}\left(i, j = 1, 2\right)$ at optimum conditions are compared with respect to (i) the sample median estimator $\hat{M}_y(n)$, when there is no matching from previous occasion and (ii) the ratio type estimator Δ proposed by Singh et al.(2007) for second quantile, where no additional auxiliary information was used at any occasion and is given by

$$\Delta = \psi \hat{M}_y(u) + (1 - \psi)\hat{M}_x(n)\left(\frac{\hat{M}_y(m)}{\hat{M}_x(m)}\right) \tag{34}$$

where ψ is an unknown constant to be determined so as to minimise the mean square error of the estimator Δ. Since, $\hat{M}_y(n)$ is unbiased and Δ is biased for population median, so variance of $\hat{M}_y(n)$ and mean square error of the estimator Δ at optimum conditions are given as

$$V\left[\hat{M}_y(n)\right] = \frac{1}{n}\frac{\left\{f_y(M_y)\right\}^{-2}}{4} \tag{35}$$

and $$M(\Delta)^*_{opt.} = \frac{\left[\mu_\Delta J_1 - J_2\right]}{n\left[\mu_\Delta^2 I_3 - \mu_\Delta J_3 - I_1\right]} \tag{36}$$

where

$$\mu_\Delta = \frac{H_2 \pm \sqrt{H_2^2 - H_1 H_3}}{H_1},\ H_1 = J_1 I_3,\ H_2 = J_2 I_3,\ H_3 = I_1 J_1 + J_2 J_3,\ J_1 = I_1 I_3,\ J_2 = I_1\left(I_2 + I_3\right),\ J_3 = I_2 + I_3 - I_1,$$

$$I_1 = \frac{\left[f_y(M_y)\right]^{-2}}{4},\ I_2 = \frac{\left[f_y(M_y)\right]^{-2}}{4} + \frac{\left[f_x(M_x)\right]^{-2} M_y^2}{4 M_x^2} - \frac{\left(4 P_{xy} - 1\right)\left[f_x(M_x)\right]^{-1}\left[f_y(M_y)\right]^{-1} M_y}{2 M_x}$$

and $I_3 = \dfrac{\left(4\,P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1} M_y}{2\,M_x} - \dfrac{\left[f_x\left(M_x\right)\right]^{-2} M_y^2}{4\,M_x^2}$.

The percent relative efficiencies $E_{ij}^{(1)}$ and $E_{ij}^{(2)}$ of the estimators $T_{ij}\left(i, j = 1, 2\right)$ (under their respective optimum conditions) with respect to $\hat{M}_y\left(n\right)$ and Δ are respectively given by

$$E_{ij}^{(1)} = \frac{V\left[\hat{M}_y\left(n\right)\right]}{M\left(T_{ij}\right)^*_{opt.}} \times 100 \quad \text{and} \quad E_{ij}^{(2)} = \frac{M\left(\Delta\right)^*_{opt.}}{M\left(T_{ij}\right)^*_{opt.}} \times 100\,;\ (i, j=1, 2)\,. \qquad (37)$$

8. Empirical Illustrations and Monte Carlo Simulation

Empirical validation can be carried out by Monte Carlo Simulation. Real life situations of two completely known finite populations have been considered.

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

The first population comprise of N = 51 states of United States. Let y_i represent the number of abortions during 2007 in the i^{th} state of U. S., x_i be the number of abortions during 2005 in the i^{th} state of U. S. and z_i denote the number of abortions during 2004 in the i^{th} state of U. S. The data are presented in Figure 8.1.

Similarly, the second population consists of N=41 corn producing states of United States. We assume y_i the production of corn (in million bushels) during 2009 in the i^{th} state of U. S., x_i be the production of corn (in million bushels) during 2008 in the i^{th} state of U. S. and z_i denote the production of corn (in million bushels) during 2007 in the i^{th} state of U. S. The data are represented by means of graph in Figure 8.2.

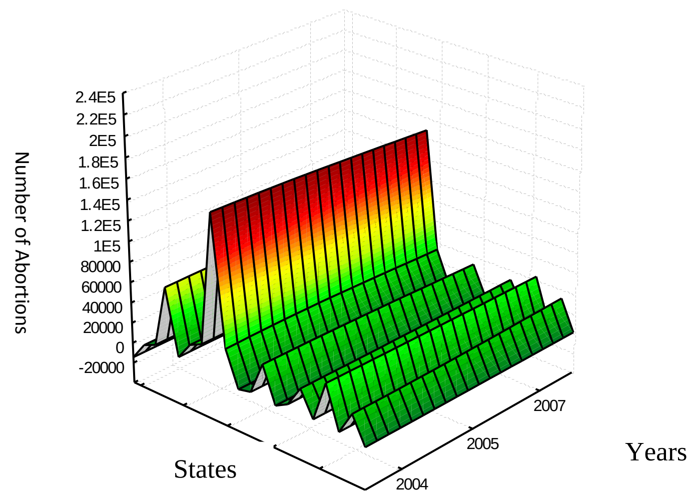


Figure 8.1: Number of abortions during 2004, 2005 and 2007 versus different states of US

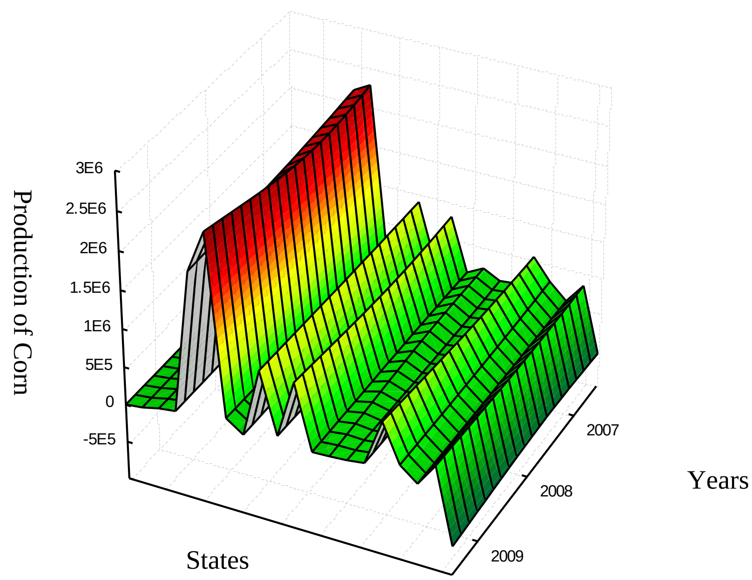


Figure 8.2: Production of corn during 2007, 2008 and 2009 versus different states of US

The graphs in Figure 8.1 and Figure 8.2 show that the number of abortions and the production of corn in different states are skewed towards right. One reason of skewness for the population-I may be the distribution of population in different states, that is the states having larger population are expected to have larger number of abortion cases.

Similarly for population-II, the states having larger area for farming are expected to have larger production of corn. Thus skewness of data indicates that the use of median may be a good measure of central location than mean in these situations.

For both the considered population-I and population-II, the optimum values of μ_{ij} (i, j = 1, 2) defined in equation (26) to (29) and percent relative efficiencies $E_{ij}^{(1)}$ and $E_{ij}^{(2)}$ defined in equation (37) of T_{ij} (i, j =1, 2) (under their respective optimality conditions) with respect to $M_y(n)$ and Δ have been computed and are presented in Table - 1.

To validate the above empirical results, Monte Carlo simulation have also been performed for Population-I. 5000 samples of n=20 states were selected using simple random sampling without replacement in the year 2005. The sample medians $\hat{M}_{x|k}(n)$ and $\hat{M}_{z|k}(n)$, k =1, 2,- - -,5000 were computed. From each one of the selected samples, m=17 states were retained and new u=3 states were selected out of N – n =51 – 20 = 31 states using simple random sampling without replacement in the year 2007. From the m units retained in the sample at the current occasion, the sample medians $\hat{M}_{x|k}(m)$, $\hat{M}_{y|k}(m)$ and $\hat{M}_{z|k}(m)$, k = 1, 2,- - -,5000 were computed. From the new unmatched units selected on the current occasion the sample medians $\hat{M}_{y|k}(u)$ and $\hat{M}_{z|k}(u)$, k = 1, 2,- - -,5000 were also calculated. The parameters φ and ψ are selected between 0.1 and 0.9 with a step of 0.1.

The percent relative efficiencies of the proposed estimators T_{ij} with respect to $M_y(n)$ and Δ are obtained as a result of above simulation and are respectively given as:

$$E_{ij}(1) = \frac{\sum_{k=1}^{5000}\left[\hat{M}_{y|k}(n) - M_y\right]^2}{\sum_{k=1}^{5000}\left[T_{ijk} - M_y\right]^2} \times 100 \quad \text{and} \quad E_{ij}(2) = \frac{\sum_{k=1}^{5000}\left[\Delta_k - M_y\right]^2}{\sum_{k=1}^{5000}\left[T_{ijk} - M_y\right]^2} \times 100\,; \quad (i, j=1, 2)$$

For better analysis, the above simulation experiments were repeated for different choices of μ.

For convenience the different choices of μ are considered as different sets for the considered Population-I which is shown below.

Set I: n=20, $\mu = 0.15$, (m =17, u = 3), **Set II:** n=20, $\mu = 0.50$, (m = 10, u = 10).

The simulation results obtained are presented in Table-2 to Table-6.

Table 1: Comparison of the proposed estimators T_{ij} (at optimum conditions) with respect to the estimators $\hat{M}_y(n)$ and Δ (at their respective optimum conditions)

	Population-I	**Population-II**
$\mu_{11}^{(0)}$	*	*
$\mu_{12}^{(0)}$	*	*
$\mu_{21}^{(0)}$	0.4114	0.4838
$\mu_{22}^{(0)}$	0.5120	0.6140
$E_{11}^{(1)}$	-	-
$E_{12}^{(1)}$	-	-
$E_{21}^{(1)}$	276.78	301.34
$E_{22}^{(1)}$	344.45	382.46
$E_{11}^{(2)}$	-	-
$E_{12}^{(2)}$	-	-
$E_{21}^{(2)}$	210.30	204.34
$E_{22}^{(2)}$	261.72	259.35

Note: '*' indicates that $\mu_{ij}^{(0)}$; $(i, j = 1, 2)$ do not exist.

Table 2: Monte Carlo Simulation results when the proposed estimator T_{ij} is compared to $\hat{M}_y(n)$ for population-I

φ	**Set**	$E_{11}(1)$	$E_{12}(1)$	$E_{21}(1)$	$E_{22}(1)$
0.1	**I**	146.57	182.32	147.48	183.81
	II	152.15	277.81	158.07	290.87
0.2	**I**	157.43	202.46	157.47	202.78
	II	176.69	314.47	185.67	331.56
0.3	**I**	174.39	222.62	172.98	220.84
	II	194.98	340.08	205.89	357.25
0.4	**I**	196.19	249.36	192.56	244.13
	II	212.07	360.70	221.46	368.06
0.5	**I**	216.80	275.64	208.53	263.21
	II	227.82	371.52	232.33	362.74
0.6	**I**	273.65	301.04	222.86	278.43
	II	238.87	371.67	235.29	342.79
0.7	**I**	258.19	324.78	232.99	286.78
	II	246.92	364.42	229.20	310.11
0.8	**I**	279.56	348.26	241.00	291.20
	II	247.70	345.58	216.75	275.50
0.9	**I**	299.13	368.28	244.26	289.30
	II	239.84	316.58	197.88	238.58

Table 3: Monte Carlo Simulation results when the proposed estimator T_{11} is compared to the estimator Δ

φ	ψ / SET	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	**I**	275.77	739.50	1352.8	2741.3	3748.7	6016.6	1259.1	1627.7	2732.2
	II	122.06	**	**	**	114.67	152.51	231.31	285.69	409.32
0.2	**I**	311.36	800.20	1550.8	2970.1	4415.9	6193.2	1301.5	1660.5	2696.9
	II	136.60	108.52	**	110.30	130.21	177.99	249.99	319.31	445.29
0.3	**I**	336.01	869.20	1704.0	3157.5	4685.7	6430.9	1295.6	1713.7	2500.3
	II	150.41	121.14	109.03	121.10	147.42	199.15	271.42	355.96	480.84
0.4	**I**	341.73	922.40	1776.5	3149.9	4884.8	6744.8	1250.5	1660.5	2175.2
	II	165.02	131.26	120.39	132.37	160.10	218.71	295.47	384.84	522.78
0.5	**I**	349.01	934.0	1803.6	3196.6	4899.7	6821.9	1147.7	1517.8	1846.1
	II	178.11	143.36	128.95	142.49	171.74	234.26	317.05	415.93	561.37
0.6	**I**	346.37	910.2	1770.9	3149.9	4895.8	6740.3	1045.2	1370.2	1518.8
	II	186.39	148.9	135.35	150.51	183.80	248.66	332.14	439.36	591.39
0.7	**I**	344.45	877.40	1712.2	3043.8	4731.7	6643.1	876.79	1168.1	1250.3
	II	193.21	154.98	140.0	155.37	190.07	256.22	340.55	451.85	613.96
0.8	**I**	331.02	842.68	1641.7	2983.0	4589.9	6355.5	746.15	1021.1	1043.5
	II	191.43	155.42	140.31	157.36	189.59	255.43	343.47	452.69	612.08
0.9	**I**	304.36	787.20	1517.9	2829.0	4274.4	5960.8	642.61	888.44	827.20
	II	188.71	153.23	136.99	152.92	186.03	249.23	335.03	439.10	601.28

Note: "**" indicates no gain.

Table 4: Monte Carlo Simulation results when the proposed estimator T_{12} is compared to the estimator Δ

φ	ψ / SET	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	**I**	382.79	977.38	1938.7	3615.9	5212.8	7650.9	1562.9	2071.2	604.32
	II	207.40	168.47	150.66	164.49	195.34	254.96	390.45	481.73	677.77
0.2	**I**	412.37	1044.8	2155.3	3865.4	5836.8	7910.1	1580.8	2058.2	694.83
	II	230.42	148.10	165.20	181.41	218.38	295.04	420.24	531.12	743.81
0.3	**I**	431.53	1116.1	2285.3	4034.5	6118.8	8107.3	1559.1	2091.9	788.21
	II	249.33	201.20	178.36	196.95	242.61	327.34	449.90	581.06	785.87
0.4	**I**	435.27	1162.7	2308.9	3957.7	6234.6	8373.5	1458.0	1981.8	882.83
	II	266.67	211.15	192.85	210.25	256.66	351.61	476.85	613.96	832.84
0.5	**I**	433.57	1152.5	2279.8	3914.5	6083.6	8240.9	1300.8	1750.7	982.72
	II	278.24	223.06	198.85	219.49	266.40	363.80	495.06	640.37	867.85
0.6	**I**	419.40	1093.1	2159.9	3752.4	5906.8	7933.0	1157.3	1533.3	1096.9
	II	277.30	221.06	199.88	221.69	273.84	370.25	496.44	648.57	872.83
0.7	**I**	406.49	1023.7	2030.3	3531.3	5536.4	7651.1	946.56	1268.9	1221.1
	II	274.81	220.32	197.97	218.46	269.83	364.06	483.84	634.79	868.93
0.8	**I**	380.53	960.60	1893.1	3386.1	5250.5	7155.5	788.85	1088.7	1353.2
	II	258.62	210.05	189.20	211.88	256.75	345.21	465.70	605.58	824.24
0.9	**I**	340.93	876.70	1705.9	3146.9	4777.4	6575.1	670.72	933.49	1483.8
	II	244.40	198.25	176.68	196.80	240.61	321.56	433.21	560.55	775.15

Table 5: Monte Carlo Simulation results when the proposed estimator T_{21} is compared to the estimator Δ

φ	SET \ ψ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	**I**	221.37	583.99	1112.3	2154.2	3005.4	4778.5	1255.8	1635.1	452.25
	II	126.34	100.97	**	100.74	118.81	158.55	240.05	297.65	424.27
0.2	**I**	188.97	477.50	953.30	1756.7	2630.2	3784.2	1277.9	1641.7	499.59
	II	143.05	114.12	104.22	115.63	136.20	187.11	261.74	335.76	467.23
0.3	**I**	146.01	369.10	748.0	1330.6	2014.4	2890.2	1228.8	1610.8	540.34
	II	157.92	127.07	115.03	127.71	154.32	210.52	285.71	373.76	505.36
0.4	**I**	111.99	280.0	572.80	1615.1	1541.8	2182.2	1124.0	1472.6	571.9
	II	170.85	136.89	125.02	138.10	166.62	227.65	308.78	400.97	543.70
0.5	**I**	**	215.50	440.60	776.0	1193.1	1679.2	987.70	1292.8	588.54
	II	180.13	144.24	131.19	145.11	174.93	238.57	322.73	421.57	570.46
0.6	**I**	**	169.20	344.70	604.30	928.50	1314.1	852.30	1114.4	596.70
	II	182.88	144.67	132.92	146.87	179.14	243.65	326.73	430.52	578.66
0.7	**I**	**	134.20	278.20	477.50	728.30	1046.2	712.42	934.0	595.40
	II	178.82	142.23	130.10	143.33	175.35	237.99	318.48	421.67	568.63
0.8	**I**	**	109.17	222.80	383.80	529.23	843.70	591.94	795.30	584.30
	II	168.53	135.68	122.22	135.66	166.62	224.0	301.69	397.37	534.44
0.9	**I**	**	**	183.70	317.10	490.4	690.70	501.36	672.26	563.0
	II	154.47	124.30	111.88	123.98	152.43	204.79	276.71	364.81	490.16

Table 6: Monte Carlo Simulation results when the proposed estimator T_{22} is compared to the estimator Δ

φ	SET \ ψ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	**I**	282.55	739.11	1493.6	2686.1	3936.3	5856.1	1563.1	2084.7	584.02
	II	216.27	176.93	157.94	171.11	204.79	266.78	407.01	505.99	710.33
0.2	**I**	223.58	565.10	1173.2	2068.3	3108.2	4426.1	1553.2	2031.1	645.12
	II	242.70	194.04	173.74	191.06	230.51	311.53	441.30	561.46	783.68
0.3	**I**	162.02	413.50	856.10	1484.5	2269.7	3233.5	1465.3	1938.9	692.13
	II	260.65	209.21	187.46	207.06	253.99	344.35	471.97	607.73	823.34
0.4	**I**	119.31	302.50	628.0	1098.8	1676.9	2360.8	1291.9	1723.7	721.17
	II	270.23	215.79	195.32	215.45	263.26	358.45	489.17	628.66	850.93
0.5	**I**	**	227.60	470.0	819.30	1268.8	1771.2	1100.9	1457.5	730.73
	II	269.61	214.41	194.81	214.42	261.90	356.15	484.27	625.35	848.60
0.6	**I**	**	176.10	360.50	628.0	969.30	1364.6	926.0	1220.3	726.10
	II	257.16	201.60	185.27	203.87	251.98	341.73	460.65	601.75	806.63
0.7	**I**	**	138.10	287.50	491.10	750.50	1074.9	757.63	997.30	709.60
	II	235.97	186.87	170.83	187.14	231.13	313.82	421.27	553.21	747.28
0.8	**I**	**	111.49	228.30	391.50	605.50	860.60	618.86	835.80	681.10
	II	210.30	168.86	151.63	167.97	208.35	278.88	377.03	492.73	663.67
0.9	**I**	**	**	187.0	321.90	498.60	700.80	518.57	697.84	642.20
	II	184.01	147.63	132.59	146.75	181.48	243.25	329.65	432.17	581.38

Note: "**" indicates no gain.

9. Mutual Comparison of the Proposed Estimators T_{ij} (i, j =1, 2)

The performances of the proposed estimators T_{ij} (i, j =1, 2) have been elaborated empirically as well as through simulation studies in above section 8 and the results obtained are presented in Table 2 to Table 6. In this section the mutual comparison of the four proposed estimators has been elaborated through different graphs given in Figure 9.1 to Figure 9.4.

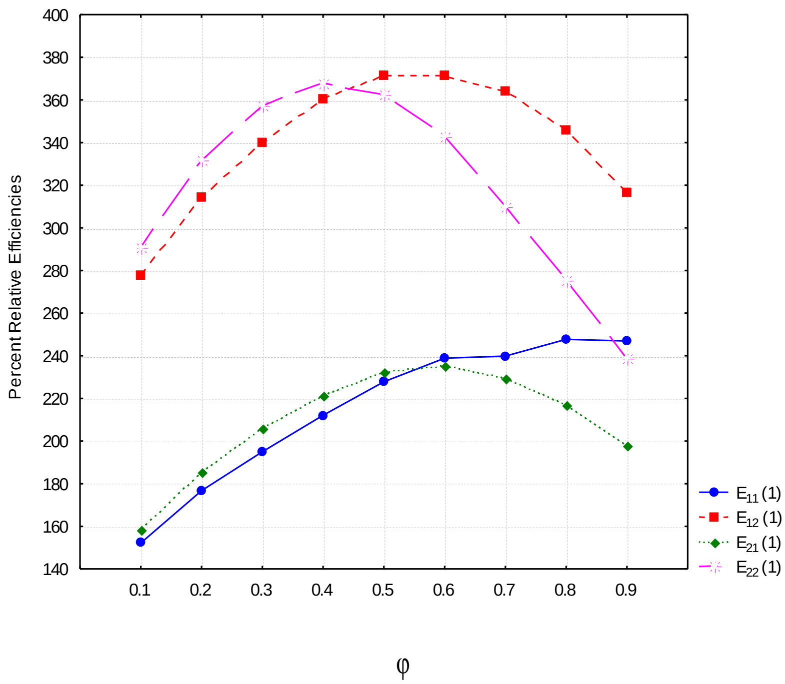


Figure 9.1: Mutual Comparison of Proposed Estimator T_{ij} (i, j =1, 2) when compared with the estimator $\hat{M}_y$ for set-II.

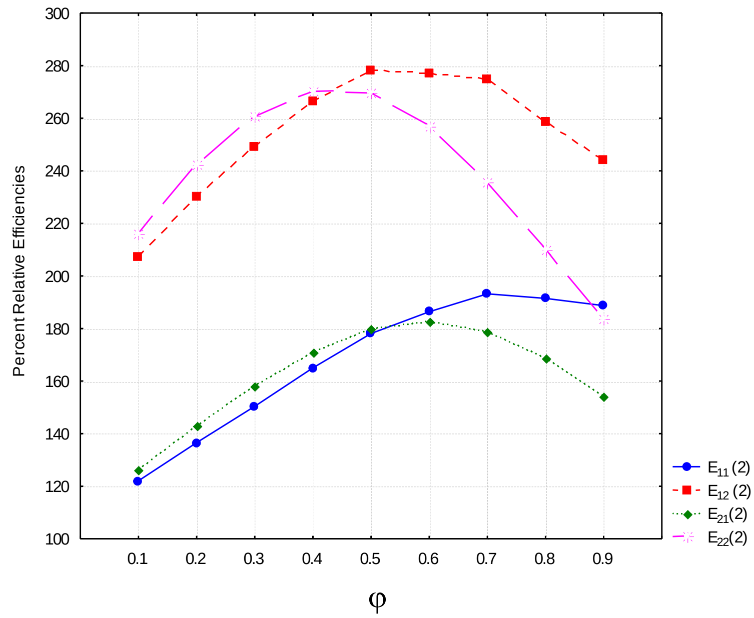


Figure 9.2: Mutual Comparison of Proposed Estimators T_{ij} (i, j =1, 2) when compared with the estimator Δ for $\psi = 0.1$ for set-II.

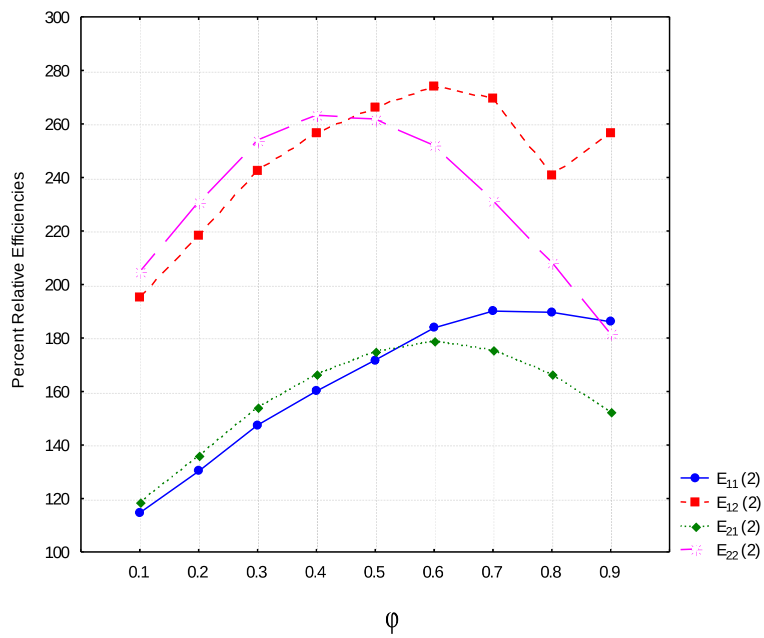


Figure 9.3: Mutual Comparison of Proposed Estimators T_{ij} (i, j =1, 2) when compared with the estimator Δ for $\psi = 0.5$ for set-II.

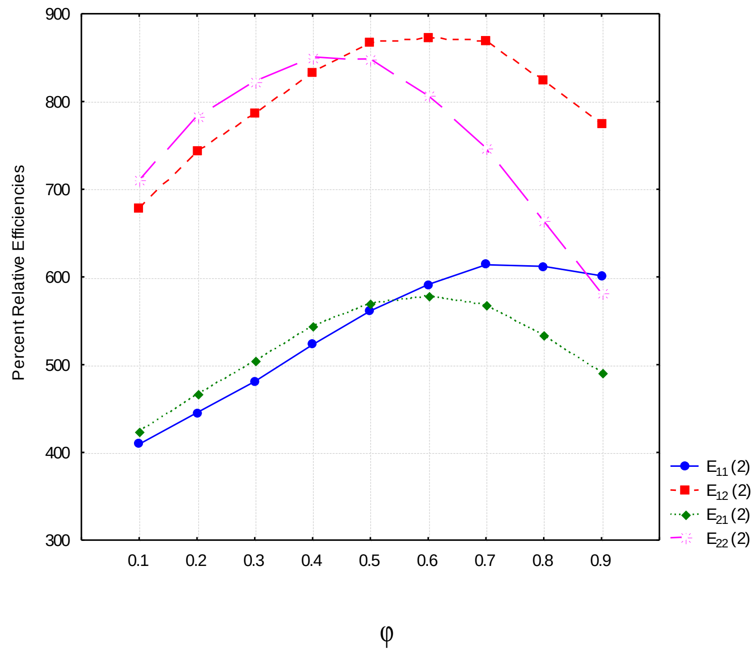


Figure 9 .4: Mutual Comparison of Proposed Estimators T_{ij} $(i, j = 1, 2)$ when compared with the estimator Δ for $\psi = 0.9$ for set-II.

11. Interpretation of Results

The following interpretation can be drawn from Tables 1 - 6 and Figure 9.1 - 9.4:

(1) From Table-1, it is observed that

(a) Optimum values $\mu_{21}^{(0)}$ and $\mu_{22}^{(0)}$ for the estimators T_{21} and T_{22} exist for both the considered Populations which justifies the applicability of the proposed estimators T_{21} and T_{22} at optimum conditions. However, the optimum values $\mu_{11}^{(0)}$ and $\mu_{12}^{(0)}$ for the estimators T_{11} and T_{12} do not exist for both the considered populations.

(b) Appreciable gain is observed in terms of precision indicating the proposed estimators T_{21} and T_{22} (at their respective optimal conditions) are preferable over the estimators $\hat{M}_y(n)$ and Δ (at optimal conditions). This result justifies the use of additional auxiliary information at both occasions which is stable over time in two occasion successive sampling.

(c) The values for $E_{11}^{(1)}, E_{12}^{(1)}, E_{11}^{(2)}$ and $E_{12}^{(2)}$ cannot be calculated as optimum values $\mu_{11}^{(0)}$ and $\mu_{12}^{(0)}$ do not exist but simulation study vindicated in Tables 2 - 6 magnify the applicability of proposed estimators T_{11} and T_{12} over sample median estimator $\hat{M}_y$ and the estimator Δ.

(2) From Table-2, it can be seen that, when T_{ij} (i, j =1, 2) is compared with sample median Estimator $\hat{M}_y(n)$.

(a) The value of $E_{11}(1)$ increases as φ increases except for set I.

(b) $E_{12}(1)$ increases as φ increases except for set II.

(c) As φ increases the value of $E_{21}(1)$ and $E_{22}(1)$ increases for set I but for set II, first it increases as φ increases and then it decreases.

(3) From Table-3, when T_{11} is compared with the estimator Δ, we infer that

(a) $E_{11}(2)$ first increases and then decreases as φ increases for all choices of ψ and for first set.

(b) For fixed choices of φ, as ψ increases the value of $E_{11}(2)$ first increases and then decreases.

(4) From Table-4, when T_{12} is compared with the estimator Δ, we observe that

(a) For set I, $E_{12}(2)$ first increases and then decreases as φ increases for all value of ψ except for few choices.

(b) For set II, $E_{12}(2)$ first increases and then decreases as φ increases for all choices of ψ.

(5) From Table-5, when T_{21} is compared with the estimator Δ, it can be seen that

(a) $E_{21}(2)$ decreases as φ increases for set I as ψ varies from 0.1 to 0.8 while for $\psi = 0.9$ it increases and then decreases as φ increases except for some combinations of φ and ψ.

(b) For set II, $E_{21}(2)$ first increases as φ increases and then decreases for all choices of ψ.

(c) $E_{21}(2)$ increases as ψ increases for all choices of φ for set I while for set II $E_{21}(2)$ first decreases and then increases as ψ increases for all choices of φ.

(6) From Table-6, it can be concluded that

(a) $E_{22}(2)$ decreases as φ increases for different choices of ψ for set I.

(b) For set II $E_{22}(2)$ first increases and then decreases as φ increases for all choices of ψ

(c) For fixed φ, $E_{22}(2)$ increases as ψ increases except for $\psi = 0.9$ for set I while for II, $E_{22}(2)$ first decreases and then increases as ψ increases for all choices of φ.

(7) The mutual comparison of the four proposed estimator T_{ij} ;(i, j = 1, 2) in Figure 9.1 to Figure 9.4 show that the estimator T_{22} comes out to be the best estimator amongst all the four proposed estimators, since, it possess largest gain over other proposed estimators. Also the estimator T_{22} has a considerably consistent nature for all combinations of φ, ψ and μ. It has also been found that the percent relative efficiency of the estimator T_{22} increases as the fraction of sample drawn at current occasion decreases and vice versa which exactly justifies the basic principles of sampling on successive occasions.

12. Conclusion

From the preceding interpretations, it may be concluded that the use of exponential ratio type estimators for the estimation of population median at current occasion in two occasion successive sampling is highly appreciable as vindicated through empirical and simulation results. The use of highly correlated auxiliary information which is stable over time is highly rewarding in terms of precision. The mutual comparison of the proposed estimators indicates that the estimators utilizing more exponential ratio type structures perform much better. It has also been observed that the estimator T_{22} in which maximum utilization of exponential ratio type structures have been considered turned out to be the most efficient among all the four proposed estimators. Hence, the proposed estimators especially the estimator T_{22} may be recommended for their practical use by survey practitioners.

CHAPTER – 4*

Multivariate Analysis of Longitudinal Surveys for Population Median

* Following is the publication based on the work of this chapter:--

1. Priyanka, K. and Mittal, R. (2016): Multivariate Analysis of Longitudinal Surveys for Population Median. Journal of Applied Statistics, (Communicated).

Multivariate Analysis of Longitudinal Surveys for Population Median

1. Introduction

In order to understand the dynamics of economic and social process which are changing over time, single time survey and their analysis do not serve the purpose. For these situations longitudinal surveys, in which the same units are investigated on several occasions, over extensive period of time becomes important. In such situations, the same population is sampled repeatedly and the same study variable is measured at each occasion, so that development over time can be followed. For example, in many countries, labour-force surveys are conducted monthly to estimate the number of employed and the rate of unemployment. Other examples are monthly surveys in which the data on price of goods are collected to determine a consumer price index, and political opinion surveys conducted at regular intervals to measure voter preferences. These longitudinal surveys in which the sampling is done on successive occasions (over years or seasons or months) according to a specified rule, with partial replacement of units, is called successive (rotation) sampling. Successive sampling provides a strong tool for generating the reliable estimates at different occasions. In this case the survey estimates are time specific, For example, the unemployment rate is a key economic indicator that varies over time, the rate may change from one month to the next because of a change in the economy (with business laying off or recruiting new employees).

The problem of sampling on two successive occasions was first considered by Jessen (1942) and latter this idea was extended, see, for example, Patterson (1950), Narain (1953), Eckler (1955), Gordon (1983), Arnab and Okafar (1992), Feng and Zou (1997), Singh and Singh (2001), Singh and Priyanka (2008a), Singh et al. (2013a), Bandhopadhyay and Singh (2014) and many others. All the above efforts were devoted to the estimation of population mean or variance on two or more occasion successive sampling.

When a distribution concerned with longitudinal survey is skewed, when end-values are not known, or when one requires reduced importance to be attached to outliers because they may be measurement errors, median can be used as a measure of central location. Median is defined on ordered one-dimensional data, and is independent of any distance metric so it can be seen as a better indication of central tendency (less susceptible to the exceptionally large value in data) than the arithmetic mean.

Very few researchers see, for example, Martinez et al. (2005), Singh et al. (2007) and Rueda et al. (2008) have proposed estimators for population median in successive sampling. Singh and Priyanka (2008b) have proposed estimator to estimate population median in two-occasion successive sampling assuming that a guess value of the population median is known. In all the above quoted papers, related to the study of median, they have assumed that the density functions appearing in the results are known. But, in general being a population parameter they are not known. Hence, using the information on additional stable auxiliary variable available on both the occasions, see Priyanka and Mittal (2014, 2016) proposed estimators for population median in successive sampling. In these papers they have also estimated the unknown density functions by using the method of generalized nearest neighbour density estimator related to kernel estimator.

Sometimes, information on several auxiliary variables may be readily available or may be made easily available by diverting a small amount of fund available for the survey. For example, to study the social evil such as number (or rate) of abortions, many factors like availability of medical facilities, income of households, level of education can be taken as additional auxiliary information. Likewise, suppose for Asian countries, one may be interested in estimating the military expenditure then the gross national product of the said countries, average export, average import etc. may be considered as additional auxiliary information.

Following Olkin (1958), technique of weighted ratio-type estimator, the objective of the present study is to develop more effective and relevant estimator using exponential ratio type estimators for population median on current occasion in two occasion successive sampling embedding information on p – additional auxiliary variates $(p \geq 1)$, which are stable over time. Properties of the proposed estimator are discussed. Optimum replacement strategies are elaborated. Proposed estimator is compared with the estimator when information on single auxiliary variable (p = 1) is available on both the occasions and also with the sample median estimator when there is no matching from the previous occasion. The dominance of the proposed estimator is justified by empirical interpretations. The results are validated by the means of simulation studies.

2. Sample Structure and Notations

Let $U = \left(U_1, U_2, \ldots, U_N\right)$ be the finite population of N units, which has been sampled over two occasions. It is assumed that size of the population remains unchanged but values of units change over two occasions. The character under study be denoted by x (y) on the first (second) occasions respectively. It is assumed that information on p - additional auxiliary variables $z_1, z_2, \ldots, z_p$, whose population median is known and stable over occasions, are readily available on both the occasions and positively correlated to x and y respectively. Simple random sample (without replacement) of n units is taken on the first occasion. A random subsample of $m = n\lambda$ units is retained for use on the second occasion. Now at the current occasion a simple random sample (without replacement) of u= (n-m) = nμ units is drawn afresh from the remaining (N - n) units of the population so that the sample size on the second occasion is also n. Let the fractions of fresh and matched samples at the second (current) occasion be μ and $\lambda\left(\mu + \lambda = 1\right)$ respectively, where $0 \leq \mu, \lambda \leq 1$. The following notations are considered for the further use:

M_i : Population median of the variable i; $i \in \left\{x, y, z_1, z_2, \ldots, z_p\right\}$.

$\hat{M}_i(u)$: Sample median of variable i; $i \in \left\{y, z_1, z_2, \ldots, z_p\right\}$ based on the sample size u.

$\hat{M}_i\left(m\right)$: Sample median of variable i; $i \in \left\{x, y, z_1, z_2, \ldots, z_p\right\}$ based on the sample size m.

$\hat{M}_i(n)$: Sample medians of variable i; $i \in \left\{x, z_1, z_2, \ldots, z_p\right\}$ based on the sample size n.

$f_i\left(M_i\right)$: The marginal densities of variable i; $i \in \left\{x, y, z_1, z_2, \ldots, z_p\right\}$.

3. Proposed Estimator T

To estimate the population median M_y on the current (second) occasion, utilizing p-additional auxiliary information which are stable over time and are readily available on both the occasions, a multivariate weighted estimator T_u based on sample of the size u= nμ drawn afresh on the current (second) occasion is proposed as

$$T_u = \mathbf{W}_u' \, \mathbf{T}_{exp}(u) \tag{1}$$

where $\mathbf{W_u}$ is a column vector of p-weights given by $\mathbf{W_u} = [w_{u_1} \quad w_{u_2} \cdot \quad \cdot \quad \cdot \quad w_{u_p}]'$

and $\mathbf{T}_{exp}(u) = \begin{bmatrix} T(1,u) \\ T(2,u) \\ \vdots \\ T(p,u) \end{bmatrix}$, where $T(i, u) = \hat{M}_y(u) \exp\left(\dfrac{M_{z_i} - \hat{M}_{z_i}(u)}{M_{z_i} + \hat{M}_{z_i}(u)} \right)$ for i = 1, 2, 3, …, p

such that $\mathbf{1}'\mathbf{W_u} = 1$, where $\mathbf{1}$ is a column vector of order p.

The second estimator T_m is also proposed as weighted multivariate chain type ratio to exponential ratio estimator based on sample size $m = n\lambda$ common to the both occasions and is given by

$$T_m = \mathbf{W}_m' \, \mathbf{T}_{exp}(m, n) \tag{2}$$

where $\mathbf{W_m}$ is a column vector of p-weights as $\mathbf{W_m} = [w_{m_1} \quad w_{m_2} \cdot \quad \cdot \quad \cdot \quad w_{m_p}]'$

and $\mathbf{T}_{exp}(m, n) = \begin{bmatrix} T(1,m,n) \\ T(2,m,n) \\ \vdots \\ T(p,m,n \end{bmatrix}$, where $T(i, m, n) = \left(\dfrac{\hat{M}_y^*(i, m)}{\hat{M}_x^*(i, m)} \hat{M}_x^*(i, n) \right)$

where $\hat{M}_y^*(i, m) = \hat{M}_y(m) \exp\left(\dfrac{M_{z_i} - \hat{M}_{z_i}(m)}{M_{z_i} + \hat{M}_{z_i}(m)} \right)$, $\hat{M}_x^*(i, m) = \hat{M}_x(m) \exp\left(\dfrac{M_{z_i} - \hat{M}_{z_i}(m)}{M_{z_i} + \hat{M}_{z_i}(m)} \right)$

and $\hat{M}_x^*(i, n) = \hat{M}_x(n) \exp\left(\dfrac{M_{z_i} - \hat{M}_{z_i}(n)}{M_{z_i} + \hat{M}_{z_i}(n)} \right)$ for i=1, 2, 3, ..., p.

Such that $\mathbf{1}'\mathbf{W_m} = 1$, where $\mathbf{1}$ is a column vector of order *p*.

The optimum weights $\mathbf{W_u}$ and $\mathbf{W_m}$ in $\mathbf{T_u}$ and $\mathbf{T_m}$ are chosen by minimizing their mean square errors respectively.

Considering the convex linear combination of the two estimators T_u and T_m, we have the final estimator of population median M_y on the current occasion as

$$T = \varphi\, T_u + (1 - \varphi)\, T_m \tag{3}$$

where φ is an unknown constant to be determined so as to minimise the mean square error of the estimator T.

Remark 3.1: For estimating the median on each occasion, the estimators T_u is suitable, which implies that more belief on T_u could be shown by choosing φ as 1 (or close to 1), while for estimating the change from occasion to occasion, the estimators T_m could be more useful so φ might be chosen as 0 (or close to 0). For asserting both problems simultaneously, the suitable (optimum) choices of φ are desired.

4. Properties of the Proposed Estimator T

4.1. Assumptions

The properties of the proposed estimators T are derived under the following assumptions:

(i) Population size is sufficiently large (i.e. $N \to \infty$), therefore finite population corrections are ignored.

(ii) As $N \to \infty$, the distribution of the bivariate variable (a, b) where $a \text{ and } b \in \{x, y, z_1, z_2, ..., z_p\}$ and $a \neq b$ approaches a continuous distribution with marginal densities $f_a(.)$ and $f_b(.)$ respectively, (see Kuk and Mak (1989)).

(iii) The marginal densities $f_x(.), f_y(.), f_{z_1}(.), f_{z_2}(.), ..., f_{z_p}(.)$ are positive.

(iv) The sample medians $\hat{M}_y(u), \hat{M}_y(m), \hat{M}_x(m), \hat{M}_x(n), \hat{M}_{z_i}(u), \hat{M}_{z_i}(m)$ and $\hat{M}_{z_i}(n)$ for i = 1, 2, 3, …, p; are consistent and asymptotically normal (see Gross (1980)).

(v) Following Kuk and Mak (1989), let P_{ab} be the proportion of elements in the population such that $a \le \hat{M}_a$ and $b \le \hat{M}_b$ where a and $b \in \{x, y, z_1, z_2, ..., z_p\}$ and $a \neq b$.

(vi) The following large sample approximations are assumed:

$$\hat{M}_y(u) = M_y(1 + e_0),\ \hat{M}_y(m) = M_y(1 + e_1),\ \hat{M}_x(m) = M_x(1 + e_2), \hat{M}_x(n) = M_x(1 + e_3)$$
$$\hat{M}_{z_i}(u) = M_{z_i}(1 + e_{4i}), \hat{M}_{z_i}(m) = M_{z_i}(1 + e_{5i}) \text{ and } \hat{M}_{z_i}(n) = M_{z_i}(1 + e_{6i})$$

such that $|e_k| < 1\ \forall$ k = 0, 1, 2, 3, 4, 5 and 6 and $|e_{ki}| < 1\ \forall$ i = 1, 2, 3, ..., p.

The values of various related expectations can be seen in Allen et al. (2002) and Singh (2003).

4.2. Bias and Mean Square Error of the Estimator T

The estimators T_u and T_m are weighted multivariate exponential ratio and chain type ratio to exponential ratio type in nature respectively. Hence they are biased for population median M_y. Therefore, the final estimator T defined in equation (3) is also biased estimator of M_y. Bias $B(.)$ and Mean square error $M(.)$ of the proposed estimator T have been derived up to first order of approximations and thus we have following theorems:

Theorem 4.2.1. Bias of the estimator T to the first order of approximations is obtained as

$$B(T) = \varphi\, B(T_u) + (1 - \varphi)\, B(T_m) \tag{4}$$

$$B(T_u) = \frac{1}{u}\mathbf{W}_u'\ \mathbf{B}_u \tag{5}$$

$$B(T_m) = \mathbf{W}_m'\ \left(\frac{1}{m}B_{m1} + \frac{1}{n}\mathbf{B}_{m2}\right) \tag{6}$$

where $\mathbf{W_u} = [w_{u_1} \quad w_{u_2} \cdot \quad \cdot \quad \cdot \quad w_{u_p}]'$, $\mathbf{B_u} = \left(B_1(u), B_2(u), \ldots, B_p(u)\right)'$,

$$B_i(u) = \frac{3\left[f_{z_i}\left(M_{z_i}\right)\right]^{-2} M_y}{32\, M_{z_i}^2} - \frac{\left(4P_{y z_i} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_i}\left(M_{z_i}\right)\right]^{-1}}{8\, M_{z_i}} \text{ for i =1, 2, 3, ..., p},$$

$\mathbf{W_m} = [w_{m_1} \quad w_{m_2} \cdot \quad \cdot \quad \cdot \quad w_{m_p}]'$,

$$B_{m1} = \frac{\left[f_x\left(M_x\right)\right]^{-2} M_y}{4\, M_x^2} - \frac{\left(4P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1}}{4\, M_x},$$

$\mathbf{B}_{m2} = \left(Bm_{21}, Bm_{22}, \ldots, Bm_{2p}\right)'$ where

$$Bm_{2i} = \left(\frac{\left(4P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1}}{4\, M_x} - \frac{\left[f_x\left(M_x\right)\right]^{-2} M_y}{4\, M_x^2} - \frac{3\left[f_{z_i}\left(M_{z_i}\right)\right]^{-2} M_y}{32\, M_{z_i}^2} - \frac{\left(4P_{y z_i} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_i}\left(M_{z_i}\right)\right]^{-1}}{8\, M_{z_i}}\right) \text{ for i = 1, 2, 3, ..., p.}$$

Proof: The bias of the estimator T is given by

$$B(T) = E\left[T - M_y\right] = \varphi\, B\left(T_u\right) + \left(1 - \varphi\right) B\left(T_m\right)$$

where $B\left(T_u\right) = E\left[T_u - M_y\right]$ and $B\left(T_m\right) = E\left[T_m - M_y\right]$

Using large sample approximations assumed in Section 4.1 and retaining terms upto the first order of approximations, the expression for $B\left(T_u\right)$ and $B\left(T_m\right)$ are obtained as in equations (5) and (6) and hence the expression for bias of the estimator T is obtained as in equation (4).

Theorem 4.2.2. Mean square error of the estimator T to the first order of approximations is obtained as

$$M(T) = \varphi^2\, M\left(T_u\right) + \left(1 - \varphi\right)^2 M\left(T_m\right) + 2\,\varphi\left(1 - \varphi\right) Cov\left(T_u, T_m\right) \tag{7}$$

$$M(T_u) = \mathbf{W}_u' \, \mathbf{D}_u \, \mathbf{W}_u \tag{8}$$

$$M(T_m) = (B)\mathbf{W}_m' \,\mathbf{E}\, \mathbf{W}_m + \mathbf{W}_m' \,\mathbf{D}_m\, \mathbf{W}_m \tag{9}$$

where $\mathbf{W_u} = [w_{u_1} \quad w_{u_2} \cdot \quad \cdot \quad \cdot \quad w_{u_p}]'$, $\mathbf{W_m} = [w_{m_1} \quad w_{m_2} \cdot \quad \cdot \quad \cdot \quad w_{m_p}]'$, $\mathbf{E}$ is a unit matrix of order $p \times p$, $\mathbf{D}_u = \left(\frac{1}{u} - \frac{1}{N}\right)\mathbf{D}_{u*}$, $\mathbf{D}_m = \left(\frac{1}{n} - \frac{1}{N}\right)\mathbf{D}_{m*}$ where

$$\mathbf{D}_{u*} = \begin{bmatrix} du_{11} & du_{12} & . & . & . & du_{1p} \\ du_{21} & du_{22} & . & . & . & du_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ du_{p1} & du_{p2} & . & . & . & du_{pp} \end{bmatrix}_{p \times p} \quad \text{and} \quad \mathbf{D}_{m*} = \begin{bmatrix} dm_{11} & dm_{12} & . & . & . & dm_{1p} \\ dm_{21} & dm_{22} & . & . & . & dm_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ dm_{p1} & dm_{p2} & . & . & . & dm_{pp} \end{bmatrix}_{p \times p}$$

where $B = \left(\frac{1}{m} - \frac{1}{N}\right) B_1$,

$$B_1 = \frac{\left[f_y(M_y)\right]^{-2}}{4} + \frac{\left[f_x(M_x)\right]^{-2} M_y^2}{4\,M_x^2} - \frac{(4P_{xy} - 1)\left[f_x(M_x)\right]^{-1}\left[f_y(M_y)\right]^{-1} M_y}{2\,M_x},$$

$$du_{ii} = \frac{\left[f_y(M_y)\right]^{-2}}{4} + \frac{\left[f_{z_i}(M_{z_i})\right]^{-2} M_y^2}{16\,M_{z_i}^2} - \frac{(4P_{y\,z_i} - 1)\left[f_y(M_y)\right]^{-1}\left[f_{z_i}(M_{z_i})\right]^{-1} M_y}{4\,M_{z_i}},$$

$$du_{ij} = \frac{\left[f_y(M_y)\right]^{-2}}{4} - \frac{(4P_{y\,z_i} - 1)\left[f_y(M_y)\right]^{-1}\left[f_{z_i}(M_{z_i})\right]^{-1} M_y}{8\,M_{z_i}} - \frac{(4P_{y\,z_j} - 1)\left[f_y(M_y)\right]^{-1}\left[f_{z_j}(M_{z_j})\right]^{-1} M_y}{8\,M_{z_j}}$$
$$+ \frac{(4P_{z_i\,z_j} - 1)\left[f_{z_i}(M_{z_i})\right]^{-1}\left[f_{z_j}(M_{z_j})\right]^{-1} M_y^2}{16\,M_{z_i} M_{z_j}},$$

$$dm_{ii} = -\frac{\left[f_x(M_x)\right]^{-2} M_y^2}{4\,M_x^2} + \frac{\left[f_{z_i}(M_{z_i})\right]^{-2} M_y^2}{16\,M_{z_i}^2} + \frac{(4P_{xy} - 1)\left[f_x(M_x)\right]^{-1}\left[f_y(M_y)\right]^{-1} M_y}{2\,M_x}$$
$$- \frac{(4P_{y\,z_i} - 1)\left[f_y(M_y)\right]^{-1}\left[f_{z_i}(M_{z_i})\right]^{-1} M_y}{4\,M_{z_i}}$$

$$dm_{ij} = -\frac{\left[f_x(M_x)\right]^{-2} M_y^2}{4\,M_x^2} + \frac{(4P_{xy}-1)\left[f_x(M_x)\right]^{-1}\left[f_y(M_y)\right]^{-1} M_y}{2\,M_x}$$

$$-\frac{(4P_{yz_i}-1)\left[f_y(M_y)\right]^{-1}\left[f_{z_i}(M_{z_i})\right]^{-1} M_y}{8\,M_{z_i}} - \frac{(4P_{yz_j}-1)\left[f_y(M_y)\right]^{-1}\left[f_{z_j}(M_{z_j})\right]^{-1} M_y}{8\,M_{z_j}}$$

$$+\frac{(4P_{z_iz_j}-1)\left[f_{z_i}(M_{z_i})\right]^{-1}\left[f_{z_j}(M_{z_j})\right]^{-1} M_y^2}{16\,M_{z_i}M_{z_j}} \text{ for i=1, 2, 3,..., p.}$$

Proof: The mean square errors of the estimator T is given by

$$M(T) = E\left[T - M_y\right]^2 = E\left[\varphi\left(T_u - M_y\right) + (1-\varphi)\left(T_m - M_y\right)\right]^2$$

$$= \varphi^2\, M(T_u) + (1-\varphi)^2 M\left[T_m\right] + 2\,\varphi\,(1-\varphi)\,\mathrm{Cov}\left(T_u, T_m\right)$$

where $M(T_u) = E\left[T_u - M_y\right]^2$ and $M(T_m) = E\left[T_m - M_y\right]^2$;

The estimators T_u and T_m are based on two independent samples of sizes u and m respectively, hence $\mathrm{Cov}\left(T_u, T_m\right) = 0$.Using large sample approximations assumed in section 4.1 and retaining terms upto the first order of approximations, the expression for $M(T_u)$ and $M(T_m)$ are obtained as given in equations (8) and (9) and hence the expressions for mean square error of estimator T is obtained.

Remark 4.2.1: The mean square error of the estimator T in equation (7) depends on the population parameters P_{xy}, P_{yz_i}, P_{xz_i}, $P_{z_iz_j}$, $f_x(M_x)$, $f_y(M_y)$ and $f_{z_i}(M_{z_i})$; (for i=1, 2, 3, .., p). If these parameters are known, the properties of proposed estimator can be easily studied. Otherwise, which is the most often situation in practice, the unknown population parameters are replaced by their sample estimates. The population proportions P_{xy}, P_{yz_i}, P_{xz_i} and $P_{z_iz_j}$ can be replaced by the sample estimate $\hat{P}_{xy}$, $\hat{P}_{yz_i}$, $\hat{P}_{xz_i}$ and $\hat{P}_{z_iz_j}$ and the marginal densities $f_y(M_y)$, $f_x(M_x)$ and $f_{z_i}(M_{z_i})$; (i = 1, 2, 3, ..., p) can be substituted by their kernel estimator or nearest neighbour density estimator or generalized nearest neighbour density estimator related to the kernel estimator (see [4]). Here, the marginal densities $f_y(M_y), f_x(M_x)$ and $f_{z_i}(M_{z_i})$ are replaced by $\hat{f}_y(\hat{M}_y(m)), \hat{f}_x(\hat{M}_x(n))$ and $\hat{f}_{z_i}(\hat{M}_{z_i}(n))$; $(i = 1, 2, 3, ..., p)$ respectively, which are obtained by method of generalized nearest neighbour density estimation related to kernel estimator.

To estimate $f_y(M_y), f_x(M_x)$ and $f_{z_i}(M_{z_i}); (i = 1, 2, 3, ..., p)$, by generalized nearest neighbour density estimator related to the kernel estimator, following procedure has been adopted:

Choose an integer $h \approx n^{1/2}$ and define the distance $\delta(x_1, x_2)$ between two points on the line to be $|x_1 - x_2|$.

For $\hat{M}_x(n)$, define $\delta_1(\hat{M}_x(n)) \leq \delta_2(\hat{M}_x(n)) \leq --- \leq \delta_n(\hat{M}_x(n))$ to be the distances, arranged in ascending order, from $\hat{M}_x(n)$ to the points of the sample.

The generalized nearest neighbour density estimate is defined by

$$\hat{f}(\hat{M}_x(n)) = \frac{1}{n\,\delta_h(\hat{M}_x(n))} \sum_{g=1}^{n} K\left[\frac{\hat{M}_x(n) - x_g}{\delta_h(\hat{M}_x(n))}\right]$$

where the kernel function K, satisfies the condition $\int_{-\infty}^{\infty} K(x)\, dx = 1$.

Here, the kernel function is chosen as Gaussian Kernel given by $K(x) = (1 / 2\pi) \exp(- ((1/2)\, x^2))$.

The estimate of $f_y(M_y)$ and $f_{z_i}(M_{z_i}); i = 1, 2, 3, ..., p$ can be obtained by the above explained procedure in similar manner.

5. Choice of Optimal Weights

To find the optimum of the weight vector $\mathbf{W}_u = [w_{u_1} \quad w_{u_2} \cdot \quad \cdot \quad \cdot \quad w_{u_p}]'$, the mean square error $M(T_u)$ given in equation (8) is minimized subject to the condition $\mathbf{1}'\mathbf{W}_u = 1$ using the method of Lagrange's Multiplier explained as under:

To find the extrema using Lagrange's Multiplier Technique, we define Q_1 as

$$Q_1 = \mathbf{W}_u' \mathbf{D}_u \mathbf{W}_u - \lambda_u(\mathbf{1}'\mathbf{W}_u - 1), \tag{10}$$

where $\mathbf{1}$ is a unit column vector of order p and λ_u is the Lagrangian multiplier.

Now, by differentiating equation (10) partially with respect to $\mathbf{W}_u$ and equating it to zero we have

$$\frac{\partial Q_1}{\partial \mathbf{W}_u} = \frac{\partial}{\partial \mathbf{W}_u}[\mathbf{W}_u' \mathbf{D}_u \mathbf{W}_u - \lambda_u(\mathbf{1}'\mathbf{W}_u - 1)] = 0$$

This implies that, $2\,\mathbf{D}_u\mathbf{W}_u - \lambda_u\mathbf{1} = \mathbf{0}$, which yields

$$\mathbf{W}_u = \frac{\lambda_u}{2}\,\mathbf{D}_u^{-1}\,\mathbf{1} \tag{11}$$

Now pre- multiplying equation (11) by $\mathbf{1}'$, we get

$$\frac{\lambda_u}{2} = \frac{1}{\mathbf{1}'\,\mathbf{D}_u^{-1}\,\mathbf{1}} \tag{12}$$

Thus, using equation (12) in equation (11) , we obtain the optimal weight vector as

$$\mathbf{W}_{u_{opt.}} = \frac{\mathbf{D}_u^{-1}}{\mathbf{1}'\mathbf{D}_u^{-1}\mathbf{1}} \tag{13}$$

In similar manners, the optimal of the weight $\mathbf{W}_m = [w_{m_1} \quad w_{m_2} \cdot \quad \cdot \quad \cdot \quad w_{m_p}]'$ is obtained by minimizing $M(T_m)$ subject to the constraint $\mathbf{1}'\mathbf{W}_m = 1$ using the method of Lagrange's multiplier, for this we define

$$Q_2 = (B)\mathbf{W}_m' \mathbf{E}\,\mathbf{W}_m + \mathbf{W}_m' \mathbf{D}_m \mathbf{W}_m - \lambda_m(\mathbf{1}'\mathbf{W}_m - 1),$$

where λ_m is the Lagrangian multiplier.

Now, differentiating Q_2 with respect to $\mathbf{W}_m$ and equating to 0, we get

$$\mathbf{W}_{m_{opt.}} = \frac{\mathbf{D}_m^{-1}}{\mathbf{1}'\mathbf{D}_m^{-1}\mathbf{1}} \tag{14}$$

Then substituting the optimum values of $\mathbf{W}_u$ and $\mathbf{W}_m$ in equations (8) and (9) respectively, the optimum mean square errors of the estimators are obtained as:

$$M(T_u)_{opt.} = \left(\frac{1}{u} - \frac{1}{N}\right)\frac{1}{\mathbf{1}'\,\mathbf{D}_{u*}^{-1}\,\mathbf{1}} \tag{15}$$

$$M(T_m)_{opt.} = \left(\frac{1}{m} - \frac{1}{N}\right)B_1 \; + \; \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{\mathbf{1}'\,\mathbf{D}_{m*}^{-1}\,\mathbf{1}} \tag{16}$$

6. Minimum Mean Square Errors of the Proposed Estimator T

Since the mean square error of the estimator T given in equation (7) is a function of unknown constants φ, therefore, it is minimized with respect to φ and subsequently the optimum values of φ is obtained as

$$\varphi_{opt.} = M(T_m)_{opt.} \Big/ \left(M(T_u)_{opt.} + M(T_m)_{opt.}\right) \tag{17}$$

Now substituting the values of $\varphi_{opt.}$ in equation (7), we obtain the optimum mean square error of the estimators T as

$$M(T)^*_{opt.} = \left(M(T_u)_{opt.} \,.\, M(T_m)_{opt.}\right) \Big/ \left(M(T_u)_{opt.} + M(T_m)_{opt.}\right) \tag{18}$$

Further, substituting the optimum values of the mean square errors of the estimators given in equations (15) and (16) in equation (17) and (18) respectively, the simplified values $\varphi_{opt.}$ and $M(T)^*_{opt.}$ are obtained as

$$\varphi_{opt.} = \mu\left[\mu C - \left(B_1 + C\right)\right] \Big/ \left[\mu^2 C - \mu \left(B_1 + C - A\right) - A\right] \tag{19}$$

$$M(T)^*_{opt.} = \frac{1}{n} \frac{\left[\mu D_1 - D_2\right]}{\left[\mu^2 C - \mu D_3 - A\right]} \tag{20}$$

where

$A = 1/\mathbf{1}' \mathbf{D}_{u*}^{-1} \mathbf{1}$, $C = 1/\mathbf{1}' \mathbf{D}_{m*}^{-1} \mathbf{1}$, $D_1 = A\,C$, $D_2 = A\,B_1 + A\,C$, $D_3 = B_1 + C - A$

$B_1 = \left[f_y(M_y)\right]^{-2} \Big/ 4 + \left(\left[f_x(M_x)\right]^{-2} M_y^2\right) \Big/ 4\,M_x^2 - \left((4P_{xy} - 1)\left[f_x(M_x)\right]^{-1} \left[f_y(M_y)\right]^{-1} M_y\right) \Big/ 2\,M_x$,

and μ is the fraction of the sample drawn afresh at the current (second) occasion.

Remark 6.1: $M(T)^*_{opt.}$ derived in equation (20) is a function of μ. To estimate the population median on each occasion the better choice of μ is 1(case of no matching); however, to estimate the change in median from one occasion to other, μ should be 0(case of complete matching). But intuition suggests that an optimum choice of μ is desired to devise the amicable strategy for both the problems simultaneously.

7. Optimum Replacement Strategy for the Estimator T

The key design parameter affecting the estimates of change is the overlap between successive samples. Maintaining high overlap between repeats of a survey is operationally convenient, since many sampled units have been located and have some experience in the survey. Hence to decide about the optimum value of μ (fractions of samples to be drawn afresh on current occasion) so that M_y may be estimated with maximum precision and minimum cost, we minimize the mean square error $M(T)^*_{opt.}$ in equation (20) with respect to μ.

The optimum value of μ so obtained is one of the two roots given by

$$\mu = \left(G_2 \pm \left(G_2^2 - G_1\, G_3\right)^{1/2}\right) \Big/ G_1 \tag{21}$$

where $G_1 = C\,D_1$, $G_2 = C\,D_2$ and $G_3 = A\,D_1 + D_2\,D_3$.

The real value of μ exist, iff $G_2^2 - G_1\,G_3 \geq 0$. For any situation, which satisfies this condition, two real values of μ may be possible, hence to choose a value of μ, it should be taken care that $0 \leq \mu \leq 1$, all other values of μ are inadmissible. If both the real values of μ are admissible, the lowest one will be the best choice as it reduces the total cost of the survey. Substituting the admissible value of μ say μ_0 from (21) in to the equation (20), we get the optimum value of the mean square error of the estimator T with respect to φ as well as μ which, is given as

$$M(T)^{**}_{opt.} = \frac{1}{n}\frac{\left[\mu_0\,D_1 - D_2\right]}{\left[\mu_0^2\,C - \mu_0\,D_3 - A\right]}\,. \tag{22}$$

8. Efficiency with Increased Number of Auxiliary Variables

As we know that increasing the number of auxiliary variables typically increases the precision of the estimates. In this section we verify this property for the proposed estimator as under: Let T_p and T_q be two proposed estimators based on p and q auxiliary variables respectively such that $p < q$, then $M(T_p) \geq M(T_q)$, i.e.

$$M(T_p) - M(T_q) \;\geq\; 0 \tag{23}$$

$$\frac{1}{n}\frac{\left[\mu\,A_pC_p - A_p\,(B + C_p)\right]}{\left[\mu^2\,C_p - \mu\,(B + C_p + A_p) - A_p\right]} - \frac{1}{n}\frac{\left[\mu\,A_qC_q - A_q\,(B + C_q)\right]}{\left[\mu^2\,C_q - \mu\,(B + C_q + A_q) - A_q\right]} \;\geq\; 0$$

On simplification, we get

$$\left(A_p - A_q\right)\left[(\mu - 1)^2\left(\mu\,C_pC_q + \frac{A_pA_q\left(C_p - C_q\right)}{\left(A_p - A_q\right)}\right) - \mu\,B\left(\left(C_p - C_q\right)(\mu - 1) - B\right)\right] \geq 0$$

This reduces to the condition

$$\left(A_p - A_q\right) \geq 0 \qquad (24)$$

So from section 6 above, we get

$$\frac{1}{\mathbf{1}'\,\mathbf{D}_p^{-1}\,\mathbf{1}} - \frac{1}{\mathbf{1}'\,\mathbf{D}_q^{-1}\,\mathbf{1}} \geq 0$$

$$\mathbf{1}'\,\mathbf{D}_q^{-1}\,\mathbf{1} \geq \mathbf{1}'\,\mathbf{D}_p^{-1}\,\mathbf{1}$$

Following, see Rao (2006), the matrix $\mathbf{D}_q$ can be partitioned and can be written as

$$\mathbf{D}_q = \begin{pmatrix} \mathbf{D}_p & \mathbf{F} \\ \mathbf{F}' & \mathbf{G} \end{pmatrix}$$

where $\mathbf{F}$, $\mathbf{F}'$ and $\mathbf{G}$ are matrices deduced from $\mathbf{D}_q$ such that their order never exceeds q-p and always greater than or equal to 1.Then,

$$\mathbf{D}_q^{-1} = \begin{pmatrix} \mathbf{D}_p^{-1} + \mathbf{HJH}' & -\mathbf{HJ} \\ -\mathbf{JH}' & \mathbf{J} \end{pmatrix} \qquad (25)$$

where $\mathbf{J} = \left(\mathbf{G} - \mathbf{F}'\mathbf{D}_p^{-1}\mathbf{F}\right)^{-1}$ and $\mathbf{H} = \mathbf{D}_p^{-1}\mathbf{F}$. (See Rao (2006) and Olkin (1958))

Now rewriting $\mathbf{1}'\mathbf{D}_q^{-1}\mathbf{1}$ by putting the value of $\mathbf{D}_q^{-1}$ from equation (25), we get

$$\mathbf{1}'\mathbf{D}_q^{-1}\mathbf{1} = \begin{pmatrix}\mathbf{1}_p & \mathbf{1}_{q-p}\end{pmatrix}' \begin{pmatrix} \mathbf{D}_p^{-1} + \mathbf{HJH}' & -\mathbf{HJ} \\ -\mathbf{JH}' & \mathbf{J} \end{pmatrix} \begin{pmatrix} \mathbf{1}_p \\ \mathbf{1}_{q-p} \end{pmatrix}$$

$$= \left(\mathbf{1}_p'\left(\mathbf{D}_p^{-1} + \mathbf{HJH}'\right) - \mathbf{1}_{q-p}'\mathbf{JH}' \quad - \mathbf{1}_p'\,\mathbf{HJ} + \mathbf{1}_{q-p}'\mathbf{J}\right) \begin{pmatrix} \mathbf{1}_p \\ \mathbf{1}_{q-p} \end{pmatrix}$$

$$= \mathbf{1}_p'\left(\mathbf{D}_p^{-1} + \mathbf{HJH}'\right)\mathbf{1}_p - \mathbf{1}_{q-p}'\mathbf{JH}'\,\mathbf{1}_p - \mathbf{1}_p'\,\mathbf{HJ}\,\mathbf{1}_{q-p} + \mathbf{1}_{q-p}'\mathbf{J}\,\mathbf{1}_{q-p}$$

$$\Rightarrow \mathbf{1}'\,\mathbf{D}_q^{-1}\,\mathbf{1} - \mathbf{1}_p'\,\left(\mathbf{D}_p^{-1}\right)\,\mathbf{1}_p = \mathbf{1}_p'\left(\mathbf{HJH}'\right)\mathbf{1}_p - \mathbf{1}_{q-p}'\mathbf{JH}'\,\mathbf{1}_p - \mathbf{1}_p'\,\mathbf{HJ}\,\mathbf{1}_{q-p} + \mathbf{1}_{q-p}'\mathbf{J}\,\mathbf{1}_{q-p}$$

$$\mathbf{1}'\,\mathbf{D}_q^{-1}\,\mathbf{1} - \mathbf{1}_p'\,\left(\mathbf{D}_p^{-1}\right)\,\mathbf{1}_p = \begin{pmatrix}\mathbf{1}_p & \mathbf{1}_{q-p}\end{pmatrix}' \begin{pmatrix}\mathbf{HJH}' & -\mathbf{HJ} \\ -\mathbf{JH}' & \mathbf{J}\end{pmatrix} \begin{pmatrix}\mathbf{1}_p \\ \mathbf{1}_{q-p}\end{pmatrix}$$

$$\mathbf{1}'\,\mathbf{D}_q^{-1}\,\mathbf{1} - \mathbf{1}_p'\,\left(\mathbf{D}_p^{-1}\right)\,\mathbf{1}_p = \mathbf{1}' \begin{pmatrix}\mathbf{H} \\ -\mathbf{I}\end{pmatrix} \mathbf{J}\,\begin{pmatrix}\mathbf{H} & -\mathbf{I}\end{pmatrix}\,\mathbf{1} \geq 0$$

Now latter follows since **J** is positive definite so that $\mathbf{R}'\mathbf{J}\,\mathbf{R} \geq 0$ for all **R,**

where $\mathbf{R} = \begin{pmatrix}\mathbf{H} & -\mathbf{I}\end{pmatrix}\,\mathbf{1}$.

Hence from equation (23)

$$M\left(T_p\right) - M\left(T_q\right) \geq 0$$

This leads to the result that utilizing more number of auxiliary variables provides more efficient estimates in terms of mean square error for the proposed estimator.

9. Efficiency Comparison

To evaluate the performance of the proposed estimator, the estimator T at optimum condition is compared with respect to the sample median estimator $\hat{M}_y(n)$, when there is no matching from previous occasion. For empirical investigations the proposed estimator have been considered for the cases p = 1 and p = 2.

The variance of sample median estimator $\hat{M}_y(n)$ is given as

$$V\left[\hat{M}_y(n)\right] = \left[f_y\left(M_y\right)\right]^{-2} \Big/ 4\,n \tag{26}$$

The percent relative efficiencies $E_{T|p=1}$ and $E_{T|p=2}$ of the estimator T (under their respective optimum conditions) with respect to $\hat{M}_y(n)$ are given by

$$E_{T|p=1} = \frac{V\left[\hat{M}_y(n)\right]}{M\left(T_{|p=1}\right)^{**}_{opt.}} \times 100 \quad \text{and} \quad E_{T|p=2} = \frac{V\left[\hat{M}_y(n)\right]}{M\left(T_{|p=2}\right)^{**}_{opt.}} \times 100 \tag{27}$$

10. Empirical Illustrations and Monte Carlo Simulation

Empirical validation can be carried out by Monte Carlo Simulation. Real life situations of two completely known finite populations have been considered.

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

The first population comprise of N = 40 states of United States. Let y_i represent the rate of abortions during 2008 in the i^{th} state of U. S., x_i be the rate of abortions during 2007 in the i^{th} state of U. S., z_{1i} denote the rate of abortions during 2005 in the i^{th} state of U. S. and z_{2i} denote the rate of abortions during 2004 in the i^{th} state of U. S. The data are presented in Figure 1.

Similarly, the second population consists of N=41 corn producing states of United States. We assume y_i the production of corn (in million bushels) during 2009 in the i^{th} state of U. S., x_i be the production of corn (in million bushels) during 2008 in the i^{th} state of U. S., z_{1i} denote the production of corn (in million bushels) during 2007 in the i^{th} state of U. S. and z_{2i} denote the production of corn (in million bushels) during 2006 in the i^{th} state of U. S. The data are represented by means of a histogram in Figure 2.

The graphs in Figure 1 and Figure 2 show that the rate of abortions and the production of corn in different states are almost skewed towards right. One reason of skewness for the population-I may be the distribution of population in different states, that is the states having larger population are expected to have larger rate of abortion cases. Similarly for population-II, the states having larger area for farming are expected to have larger production of corn. Thus skewness of data indicates that the use of median may be a good measure of central location than mean in these situations.

Figure 1: Rate of Abortion versus different states of U.S. during 2007

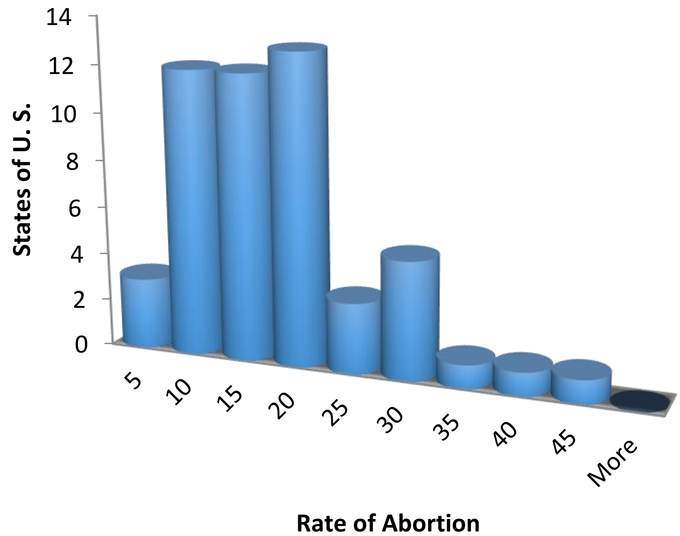
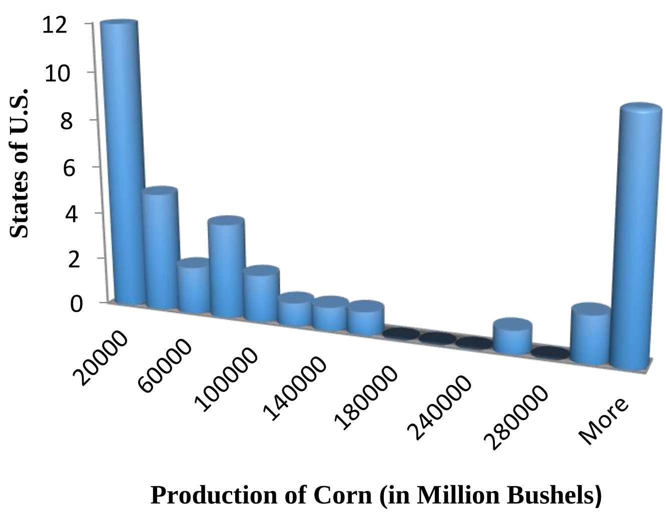


Figure 2: Production of Corn (In Million Bushels) versus different states of U.S. during 2008



For the considered population-I and population-II, the optimum values of μ defined in (21) and percent relative efficiencies $E_{T|p=1}$ and $E_{T|p=2}$ (defined in (27)) of T (for p=1 and p=2 under their optimal conditions) with respect to $\hat{M}_y(n)$ have been computed and are presented in Table-1. To validate the above empirical results, Monte Carlo simulation have also been performed for Population-I.

Simulation Algorithm

(i) Choose 5000 samples of size n=15 using simple random sampling without replacement on first occasion for both the study and auxiliary variables.

(ii) Calculate sample median $\hat{M}_{x|k}(n)$, $\hat{M}_{z_1|k}(n)$ and $\hat{M}_{z_2|k}(n)$ for k =1, 2, - - -, 5000.

(iii) Retain m = 13 units out of each n = 15 sample units of the study and auxiliary variables at the first occasion.

(iv) Calculate sample median $\hat{M}_{x|k}(m)$, $\hat{M}_{z_1|k}(m)$ and $\hat{M}_{z_2|k}(m)$ for k= 1, 2, - - -, 5000.

(v) Select u = 2 units using simple random sampling without replacement from N-n = 25 units of the population for study and auxiliary variables at second (current) occasion.

(vi) Calculate sample medians $\hat{M}_{y|k}(u)$, $\hat{M}_{z_1|k}(u)$, $\hat{M}_{z_2|k}(u)$ and $\hat{M}_{y|k}(m)$ for k = 1, 2, - - -, 5000 .

(vii) Iterate the parameter φ from 0.1 to 0.9 with a step of 0.1.

(viii) Calculate the percent relative efficiencies of the proposed estimator T with the case p = 1 and p = 2 (i.e. $T_{|p=1}$ and $T_{|p=2}$) with respect to sample median estimator $\hat{M}_y(n)$ as

$$E_1(\text{sim}) = \frac{\sum_{k=1}^{5000}\left[\hat{M}_{y|k}(n) - M_y\right]^2}{\sum_{k=1}^{5000}\left[T_{p=1|k} - M_y\right]^2} \times 100 \quad \text{and} \quad E_2(\text{sim}) = \frac{\sum_{k=1}^{5000}\left[\hat{M}_{y|k}(n) - M_y\right]^2}{\sum_{k=1}^{5000}\left[T_{p=2|k} - M_y\right]^2} \times 100$$

for k= 1, 2, - - -, 5000.

For better analysis, the above simulation experiments were repeated for different choices of μ. For convenience the different choices of μ are considered as different sets for the considered Population-I which is shown below:

Set I: n=15, $\mu = 0.10$, (m = 13, u = 2), **Set II:** n=15, $\mu = 0.20$, (m = 12, u = 3)

Set III: n=15, $\mu = 0.30$, (m = 10, u = 5), **Set IV:** n=15, $\mu = 0.40$, (m = 9, u = 6)

The simulation results obtained are presented in Table-2.

Table 1: Comparison of the proposed estimators $T_{|p=1}$ and $T_{|p=2}$ (at their respective optimum conditions) with respect to the estimator $\hat{M}_y(n)$

	Population-I	**Population-II**
$\mu_{T\,|\,p=1}$	0.5478	0.5418
$\mu_{T\,|\,p=2}$	0.5229	0.4759
$E_{T\,|\,p=1}$	171.16	136.74
$E_{T\,|\,p=2}$	199.54	322.16

Table 2: Estimated values of population Median by using the proposed estimators $T_{|p=1}$ and

$T_{|p=2}$ at their optimum conditions.

Actual / Estimated	Population-I			Population-II		
	$M_y = 15.50$			$M_y = 57$		
	n=10	**n=15**	**n=20**	**n=10**	**n=15**	**n=20**
$T_{\|p=1}$	14.83	15.10	16.16	55.74	50.97	50.21
$T_{\|p=2}$	15.01	15.47	15.98	56.31	60.69	58.72

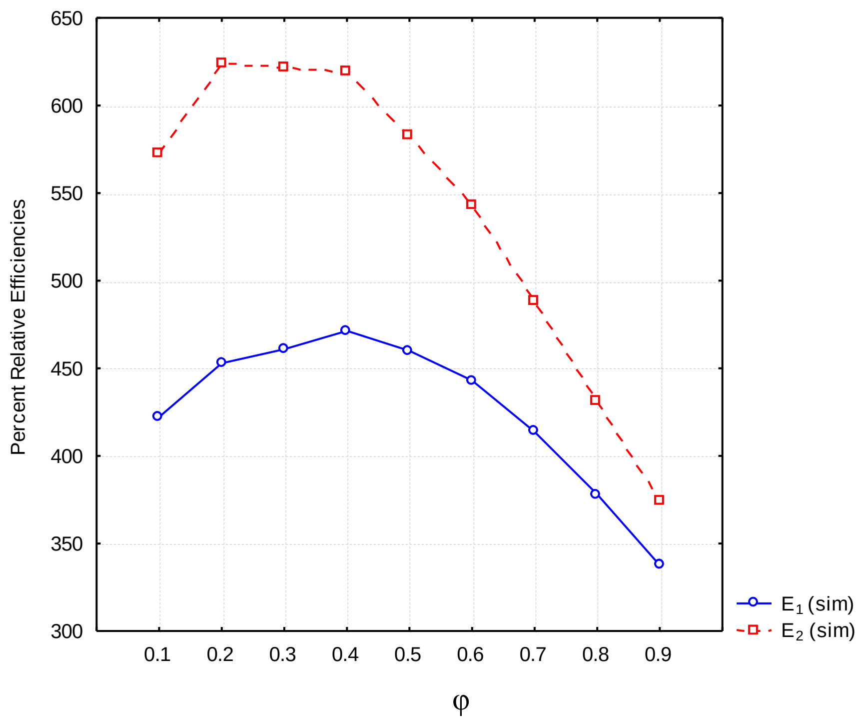
Table 3: Monte Carlo Simulation results when the proposed estimators $T_{|p=1}$ and $T_{|p=2}$ are compared to $\hat{M}_y(n)$ for population-I

φ \ Set		I	II	III	IV
0.1	E_1(sim)	307.79	536.79	316.69	422.16
	E_2(sim)	528.09	750.31	503.03	572.87
0.2	E_1(sim)	304.69	523.28	352.51	452.88
	E_2(sim)	538.70	742.33	545.61	624.11
0.3	E_1(sim)	294.64	505.08	370.03	460.89
	E_2(sim)	521.40	727.42	556.61	621.96
0.4	E_1(sim)	277.73	470.01	366.51	471.37
	E_2(sim)	480.55	669.88	521.47	618.86
0.5	E_1(sim)	260.27	426.75	355.05	459.98
	E_2(sim)	431.18	588.43	479.63	582.28
0.6	E_1(sim)	241.34	381.47	328.24	443.04
	E_2(sim)	379.41	506.80	418.77	542.37
0.7	E_1(sim)	222.24	339.33	301.81	413.75
	E_2(sim)	329.89	433.81	366.84	488.27
0.8	E_1(sim)	204.09	298.86	272.14	378.01
	E_2(sim)	285.40	366.39	316.65	431.25
0.9	E_1(sim)	184.42	263.30	239.41	337.58
	E_2(sim)	243.33	312.30	268.82	373.59

11. Mutual Comparison of the Estimators $\mathbf{T}_{|p=1}$ and $\mathbf{T}_{|p=2}$

The performances of the estimator $T_{|p=1}$ and $T_{|p=2}$ have been elaborated empirically as well as through simulation studies in above sections and the results obtained are presented in Table 1 and Table 3. The mutual comparison of the estimators for the cases when p=1 and p = 2 has been elaborated graphically and is presented in Figure 3.

Figure 3: Mutual Comparison of Proposed Estimator $T_{|p=1}$ and $T_{|p=2}$ when compared with the estimator $\hat{M}_y$ for set-IV



12. Interpretation of Results

(i) It is clear from Table 1 that optimum values of $\mu_{T|p=1}$ and $\mu_{T|p=2}$ exist for both the considered population and $\mu_{T|p=2} < \mu_{T|p=1}$. This indicates that less fraction of fresh sample is required when more number of auxiliary variables is used. Hence, total cost of survey will also get reduced when more number of additional auxiliary variables will be considered.

(ii) Table 1 also explains that the value of $E_{T|p=2} > E_{T|p=1}$, this justifies the fact that efficiency is highly increased when more numbers of auxiliary variates are taken into consideration, which also resembles in accordance with the theory.

(iii) In Table 2 we have also calculated the estimates of population median by using the proposed estimator T for p=1 and p=2 at their respective optimum conditions. We see that the estimates for population median are quite near to the original value of population median.

(iv) From simulation study in Table 3 and Figure 3, we observe that the value of $E_1(\text{sim})$ and $E_2(\text{sim})$ exists for all choices of φ and for all different sets. As φ increases the value of $E_1(\text{sim})$ and $E_2(\text{sim})$ decreases for all sets which indicates that if more weight is given to the estimator defined on current occasion, the efficiency of the estimator T get reduced, which is in accordance with see [19] results. The big difference in two lines in Figure 3 shows that the performance of estimator drastically enhances when more number of auxiliary variables is taken in to account. In real time exercise the estimates for population median are more near to the original value of population median when the numbers of auxiliary variables are increased.

(v) From Table 3 we also observe that for set II, the estimators $T_{|p=1}$ and $T_{|p=2}$ prove to be extensively better than the sample median estimator. Although no fixed pattern is observed in the efficiencies of the proposed estimators, if the value of fraction of fresh sample to be drawn on current occasion increases.

12. Conclusion

From the preceding interpretations, it may be concluded that the use of multivariate exponential ratio type estimators for the estimation of population median at current occasion in two occasion successive sampling is highly appreciable as vindicated through empirical and simulation results. The mutual comparison of the proposed estimators indicates that the estimators utilizing more auxiliary variables perform much better in terms of cost as well as precision. Hence, the proposed multivariate estimator T may be recommended for its practical use in longitudinal surveys for the estimation of population median by survey practitioners.

CHAPTER – 5*

Estimation of Population Median in Two-Occasion Rotation Sampling

* Following is the publication based on the work of this chapter:--

1. Priyanka, K. and Mittal, R. (2015): Estimation of Population Median in Two-Occasion Rotation Sampling. Journal of Statistics Applications & Probability Letters, Vol. 2, No. 3, 205-219.

Estimation of Population Median in Two-Occasion Rotation Sampling Two Occasion

1. Introduction

Survey often get repeated on many occasions for estimating same characteristics at different point of time technically called repetitive sampling or sampling over successive occasions. It has been given considerable attention by some survey statisticians, when a population is subject to change, a survey carried out on a single occasion will provide information about the characteristic of the surveyed population for the given occasion only, the survey estimates are therefore time specific. Generally, the main objective of successive surveys is to estimate the change with a view to study the effect of the forces acting upon the population as this scheme consists of selecting sample units on different occasions such that some units are common with sample drawn on previous occasions. This retention of a part of sample in periodic surveys provides efficient estimates as compared to other alternatives by eliminating some of the old elements from the sample and adding new elements to the sample each time.

The problem of sampling on two successive occasions was first considered by Jessen (1942) and latter this idea was extended, see, for example, Patterson (1950), Narain (1953), Eckler (1955), Gordon (1983), Arnab and Okafar (1992), Feng and Zou (1997), Singh and Singh (2001), Singh and Priyanka (2008a), Singh et al. (2013a), Bandhopadhyay and Singh (2014) and many others. All the above efforts were devoted to the estimation of population mean or variance on two or more occasion successive sampling.

When a distribution is skewed, when end-values are not known, or when one requires reduced importance to be attached to outliers because they may be measurement errors, median can be used as a measure of central location. Median is defined on ordered one-dimensional data, and is independent of any distance metric so it can be seen as a better indication of central tendency (less susceptible to the exceptionally large value in data) than the arithmetic mean.

Most of the studies related to median have been developed by assuming simple random sampling or its ramification in stratified random sampling considering only the variable of interest without making explicit use of auxiliary variables (Sedranks and Meyer (1978), Gross (1980), Smith and Sedranks (1983)). Some of the researchers namely Chamber and Dunstan (1986), Kuk and Mak (1989), Rao et al. (1990), Rueda et al (1998) and Allen et al. (2002) etc. have utilized the auxiliary information for the estimation of population median.

Very few researchers namely Martinez et al (2005) and Rueda and Munoz (2008) have proposed estimators for population median in successive sampling.

The work done in Singh and Priyanka (2008b) have proposed estimator to estimate population median in two-occasion successive sampling assuming that a guess value of the population median is known. In all the above quoted papers, related to the study of median, they have assumed that the density functions appearing in the results are known. But, in general being a population parameter they are not known. Hence, using the additional stable auxiliary variable available on both the occasions, Priyanka and Mittal (2014, 2016) have proposed estimators for population median in successive sampling. In these papers they have also estimated the unknown density functions by using the method of generalized nearest neighbour density estimator related to kernel estimator.

But in practice, one may find that if the gap between two successive occasions is large, the stability character of the auxiliary variate may not sustain. In addition to this, we may also find several other situations where auxiliary variate may not be stable over time, whatever is the duration between two surveys. In such situations the use of dynamic auxiliary variate (changing over time) which are readily available on different occasions, may be efficiently utilized for estimating the population median at current occasions.

Hence, focusing on the above problems in this work we have proposed more effective and relevant estimators of population median at current occasion in two occasion successive sampling using additional auxiliary information which is dynamic over time and is readily available at both the occasions. Properties of the proposed estimators are discussed. The density functions appearing in the results have been estimated by the method of generalized nearest neighbour density estimator related to kernel estimator.

Optimum replacement strategies are elaborated for the proposed estimators. Proposed estimators at optimum conditions are compared with the sample median estimator when there is no matching from the previous occasion as well as with the ratio type estimator proposed by Singh et al. (2007) for second quantile, when no additional auxiliary information was used at any occasion. The behaviours of the proposed estimators are justified by empirical interpretations and validated by the means of simulation study with the help of some natural populations.

2. Sample Structure and Notations

Let $U = \left(U_1, U_2, \ldots, U_N\right)$ be the finite population of N units, which has been sampled over two occasions. It is assumed that size of the population remains unchanged but values of units change over two occasions. The character under study be denoted by x (y) on the first (second) occasions respectively. It is assumed that information on an auxiliary variable whose population medians are known and dynamic over occasions are readily available on both the occasions and positively correlated to x and y respectively. Let z_1 be the auxiliary variable on first occasion which changes to z_2 on second (current) occasions. Simple random sample (without replacement) of n units is taken on the first occasion. A random subsample of $m = n\lambda$ units is retained for use on the second occasion. Now at the current occasion a simple random sample (without replacement) of u= (n-m) = nμ units is drawn afresh from the remaining (N-n) units of the population so that the sample size on the second occasion is also n. μ and λ, $\left(\mu + \lambda = 1\right)$ are the fractions of fresh and matched samples respectively at the second (current) occasion. The following notations are considered for the further use:

$M_x, M_y, M_{z_1}, M_{z_2}$: Population median of the variables x, y, z_1 and z_2 respectively.

$\hat{M}_y(u), \hat{M}_{z_1}(u), \hat{M}_{z_2}(u)$: Sample median of variables y, z_1 and z_2 based on the sample of size u.

$\hat{M}_x(m), \hat{M}_y(m), \hat{M}_{z_1}(m), \hat{M}_{z_2}(m)$: Sample median of variables x, y, z_1 and z_2 based on the sample of size m.

$\hat{M}_x(n), \hat{M}_{z_1}(n), \hat{M}_{z_2}(n)$: Sample medians of variables x, z_1 and z_2 based on the sample of size n.

$f_x(M_x), f_y(M_y), f_{z_1}(M_{z_1}), f_{z_2}(M_{z_2})$: The marginal densities of variables x, y, z_1 and z_2 respectively.

3. Proposed Estimator T_{ij} (i, j =1, 2)

To estimate the population median M_y on the current (second) occasion, two sets of estimators have been proposed utilizing the concept of exponential ratio type estimators. First set of estimators $\{T_{1u}, T_{2u}\}$ is based on sample of the size $u = n\mu$ drawn afresh on the current (second) occasion and the second set of estimators $\{T_{1m}, T_{2m}\}$ is based on sample size $m = n\lambda$ common to the both occasions. The two sets of the proposed estimators are given as

$$T_{1u} = M_{z_2}\left(\frac{\hat{M}_y(u)}{\hat{M}_{z_2}(u)}\right) \tag{1}$$

$$T_{2u} = \hat{M}_y(u)\ \exp\left(\frac{M_{z_2} - \hat{M}_{z_2}(u)}{M_{z_2} + \hat{M}_{z_2}(u)}\right) \tag{2}$$

$$T_{1m} = \hat{M}_x(n)\ \left(\frac{\hat{M}_y(m)}{\hat{M}_x(m)}\right)\ \exp\ \left(\frac{M_{z_2} - \hat{M}_{z_2}(m)}{M_{z_2} + \hat{M}_{z_2}(m)}\right) \tag{3}$$

$$T_{2m} = \hat{M}_x^*(n)\left(\frac{\hat{M}_y^*(m)}{\hat{M}_x^*(m)}\right) \tag{4}$$

where $\hat{M}_y^*(m) = \hat{M}_y(m)\ \exp\left(\frac{M_{z_2} - \hat{M}_{z_2}(m)}{M_{z_2} + \hat{M}_{z_2}(m)}\right), \hat{M}_x^*(m) = \hat{M}_x(m)\ \exp\left(\frac{M_{z_1} - \hat{M}_{z_1}(m)}{M_{z_1} + \hat{M}_{z_1}(m)}\right)$

and $\hat{M}_x^*(n) = \hat{M}_x(n) \exp\left(\frac{M_{z_1} - \hat{M}_{z_1}(n)}{M_{z_1} + \hat{M}_{z_1}(n)}\right).$

Considering the convex linear combination of the two sets of estimators $T_{iu}(i = 1, 2)$ and $T_{jm}(j = 1, 2)$, we have the final estimators of population median M_y on the current occasion as

$$T_{ij} = \varphi_{ij} T_{iu} + (1 - \varphi_{ij})\, T_{jm}\ ;\ (i, j = 1, 2) \tag{5}$$

where $\varphi_{ij}(i, j = 1, 2)$ are the unknown constants to be determined so as to minimise the mean square error of the estimators T_{ij}(i, j=1, 2).

Remark 3.1: For estimating the median on each occasion, the estimators $T_{iu}(i = 1, 2)$ are suitable, which implies that more belief on T_{iu} could be shown by choosing $\varphi_{ij}(i, j = 1, 2)$ as 1 (or close to 1), while for estimating the change from occasion to occasion, the estimators $T_{jm}(j=1, 2)$ could be more useful so φ_{ij} might be chosen as 0 (or close to 0). For asserting both problems simultaneously, the suitable (optimum) choices of φ_{ij} are desired.

4. Properties of the Proposed Estimators T_{ij} (i, j =1, 2)

4.1. Assumptions

The properties of the proposed estimators $T_{ij}(i, j =1, 2)$ are derived under the following assumptions:

(i) Population size is sufficiently large (i.e. $N \to \infty$), therefore finite population corrections are ignored.

(ii) As $N \to \infty$, the distribution of the bivariate variable (a, b) where $a \text{ and } b \in \{x, y, z_1, z_2\}$ and $a \neq b$ approaches a continuous distribution with marginal densities $f_a(.)$ and $f_b(.)$ respectively, (see Kuk and Mak (1989)).

(iii) The marginal densities $f_x(.), f_y(.), f_{z_1}(.)$ and $f_{z_2}(.)$ are positive.

(iv) The sample medians $\hat{M}_y(u), \hat{M}_y(m), \hat{M}_x(m), \hat{M}_x(n), \hat{M}_{z_1}(u), \hat{M}_{z_2}(u), \hat{M}_{z_1}(m)$, $\hat{M}_{z_2}(m), \hat{M}_{z_1}(n)$ and $\hat{M}_{z_2}(n)$ are consistent and asymptotically normal (see Gross (1980)).

(v) Following Kuk and Mak (1989), let P_{ab} be the proportion of elements in the population such that $a \le \hat{M}_a$ and $b \le \hat{M}_b$ where a and $b \in \{x, y, z_1, z_2\}$ and $a \neq b$.

(vi) Following large sample approximations are assumed:

$$\hat{M}_y(u) = M_y(1 + e_0),\ \hat{M}_y(m) = M_y(1 + e_1),\ \hat{M}_x(m) = M_x(1 + e_2),\ \hat{M}_x(n) = M_x(1 + e_3),$$
$$\hat{M}_{z_2}(u) = M_{z_2}(1 + e_4), \hat{M}_{z_2}(m) = M_{z_2}(1 + e_5), \hat{M}_{z_1}(m) = M_{z_1}(1 + e_6) \text{ and } \hat{M}_{z_1}(n) = M_{z_1}(1 + e_7),$$

such that $|e_i| < 1\ \forall\ i = 0, 1, 2, 3, 4, 5, 6$ and 7 .

The values of various related expectations can be seen in Allen et al. (2002) and Singh (2003).

4.2. Bias and Mean Square Errors of the Estimators T_{ij} (i, j =1, 2)

The estimators T_{iu} and $T_{jm}(i, j=1, 2)$ are ratio, exponential ratio, ratio to exponential ratio and chain type ratio to exponential ratio type in nature respectively. Hence they are biased for population median M_y. Therefore, the final estimators $T_{ij}(i, j =1, 2)$ defined in equation (5) are also biased estimators of M_y. Bias $B(.)$ and mean square errors $M(.)$ of the proposed estimators $T_{ij}(i, j =1, 2)$ are obtained up to first order of approximations and thus we have following theorems:

Theorem 4.2.1. Bias of the estimators $T_{ij}(i, j =1, 2)$ to the first order of approximations are obtained as

$$B(T_{ij}) = \varphi_{ij}\ B(T_{iu}) + (1 - \varphi_{ij})\ B(T_{jm});\ (i, j=1,2) \qquad (6)$$

where $$B\left(T_{1u}\right) = \frac{1}{u}\left\{\frac{\left[f_{z_2}\left(M_{z_2}\right)\right]^{-2} M_y}{4\,M_{z_2}^2} - \frac{\left(4\,P_{yz_2} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1}}{4\,M_{z_2}}\right\} \qquad (7)$$

$$B\left(T_{2u}\right) = \frac{1}{u}\left\{\frac{3\left[f_{z_2}\left(M_{z_2}\right)\right]^{-2} M_y}{32\,M_{z_2}^2} - \frac{\left(4\,P_{yz_2} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1}}{8\,M_{z_2}}\right\} \qquad (8)$$

$$\begin{aligned} B\left(T_{1m}\right) = \frac{1}{m} & \left\{\frac{\left[f_x\left(M_x\right)\right]^{-2} M_y}{4\,M_x^2} - \frac{\left(4P_{xy}-1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1}}{4M_x} + \frac{3\left[f_{z_2}\left(M_{z_2}\right)\right]^{-2} M_y}{32\,M_{z_2}^2}\right. \\ & \left. + \frac{\left(4\,P_{xz_2} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1} M_y}{8\,M_x M_{z_2}} - \frac{\left(4\,P_{yz_2} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1}}{8M_{z_2}}\right\} \\ & + \frac{1}{n}\left\{\frac{\left(4\,P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1}}{4\,M_x} - \frac{\left(4\,P_{xz_2} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1} M_y}{8\,M_x\,M_{z_2}}\right. \\ & \left. - \frac{\left[f_x\left(M_x\right)\right]^{-2} M_y}{4M_x^2}\right\} \end{aligned} \qquad (9)$$

$$\begin{aligned} B\left(T_{2m}\right) = \frac{1}{m} & \left\{\frac{\left[f_x\left(M_x\right)\right]^{-2}\ M_y}{4\,M_x^2} + \frac{3\left[f_{z_2}\left(M_{z_2}\right)\right]^{-2} M_y}{32M_{z_2}^2} - \frac{\left(4\,P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1}}{4\,M_x}\right. \\ & - \frac{\left(4\,P_{xz_1} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_{z_1}\left(M_{z_1}\right)\right]^{-1} M_y}{8\,M_x M_{z_1}} + \frac{\left(4\,P_{xz_2} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1} M_y}{8\,M_x M_{z_2}} \\ & + \frac{\left(4\,P_{yz_1} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_1}\left(M_{z_1}\right)\right]^{-1}}{8\,M_{z_1}} - \frac{\left(4\,P_{yz_2} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1}}{8\,M_{z_2}} \\ & \left. - \frac{\left(4\,P_{z_1 z_2} - 1\right)\left[f_{z_1}\left(M_{z_1}\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1} M_y}{16\,M_{z_1} M_{z_2}} - \frac{\left[f_{z_1}\left(M_{z_1}\right)\right]^{-2} M_y}{32M_{z_1}^2}\right\} \\ & + \frac{1}{n}\left\{\frac{\left(4\,P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1}}{4\,M_x} + \frac{\left(4\,P_{xz_1} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_{z_1}\left(M_{z_1}\right)\right]^{-1} M_y}{8\,M_x M_{z_1}}\right. \\ & - \frac{\left(4\,P_{xz_2} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1} M_y}{8\,M_x M_{z_{2.}}} - \frac{\left(4\,P_{yz_1} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_1}\left(M_{z_1}\right)\right]^{-1}}{8\,M_{z_1}} \end{aligned}$$

$$+\frac{\left(4\,P_{z_1z_2}-1\right)\left[f_{z_1}\left(M_{z_1}\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1}M_y}{16\,M_{z_1}M_{z_2}}-\frac{\left[f_x\left(M_x\right)\right]^{-2}\,M_y}{4\,M_x^2}+\frac{\left[f_{z_1}\left(M_{z_1}\right)\right]^{-2}\,M_y}{32\,M_{z_1}^2}\right\} \quad (10)$$

Proof: The bias of the estimators $T_{ij}\left(i, j = 1, 2\right)$ are given by

$$B\left(T_{ij}\right) = E\left[T_{ij} - M_y\right] = \varphi_{ij}\,B\left(T_{iu}\right) + \left(1-\varphi_{ij}\right)B\left(T_{jm}\right)$$

where $B\left(T_{iu}\right) = E\left[T_{iu} - M_y\right]$ and $B\left(T_{jm}\right) = E\left[T_{jm} - M_y\right]$

Using large sample approximations assumed in Section 4.1 and retaining terms upto the first order of approximations, the expression for $B\left(T_{iu}\right)$ and $B\left(T_{jm}\right)$ are obtained as in equations (7) - (10) and hence the expression for bias of the estimators $T_{ij}\left(i, j =1, 2\right)$ are obtained as in equation (6).

Theorem 4.2.2. Mean square errors of the estimators $T_{ij}\left(i, j =1, 2\right)$ to the first order of approximations are obtained as

$$M\left(T_{ij}\right)=\varphi_{ij}^2\,M\left(T_{iu}\right)+\left(1-\varphi_{ij}\right)^2M\left(T_{jm}\right)+2\,\varphi_{ij}\left(1-\varphi_{ij}\right)Cov\left(T_{iu},T_{jm}\right);(i, j=1,2) \quad (11)$$

Where $M\left(T_{1u}\right) = \frac{1}{u}\,A_1$ (12)

$$M\left(T_{2u}\right) = \frac{1}{u}\,A_4 \quad (13)$$

$$M\left(T_{1m}\right) = \frac{1}{m}\,A_2 + \frac{1}{n}\,A_3 \quad (14)$$

$$M\left(T_{2m}\right) = \frac{1}{m}\,A_5 + \frac{1}{n}\,A_6 \quad (15)$$

$$A_1 = \left\{ \frac{\left[f_y\left(M_y\right)\right]^{-2}}{4} + \frac{\left[f_{z_2}\left(M_{z_2}\right)\right]^{-2} M_y^2}{4\,M_{z_2}^2} - \frac{\left(4\,P_{yz_2} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1} M_y}{2\,M_{z_2}} \right\},$$

$$A_2 = \left\{ \frac{\left[f_y\left(M_y\right)\right]^{-2}}{4} + \frac{\left[f_x\left(M_x\right)\right]^{-2} M_y^2}{4\,M_x^2} + \frac{\left[f_{z_2}\left(M_{z_2}\right)\right]^{-2} M_y^2}{16\,M_{z_2}^2} - \frac{\left(4\,P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1} M_y}{2\,M_x} \right.$$
$$\left. - \frac{\left(4\,P_{yz_2} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1} M_y}{4\,M_{z_2}} + \frac{\left(4\,P_{xz_2} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1} M_y^2}{4\,M_x\,M_{z_2}} \right\},$$

$$A_3 = \left\{ \frac{\left(4\,P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1} M_y}{2\,M_x} - \frac{\left(4\,P_{xz_2} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_z\left(M_{z_2}\right)\right]^{-1} M_y^2}{4\,M_x\,M_{z_2}} \right.$$
$$\left. - \frac{\left[f_x\left(M_x\right)\right]^{-2} M_y^2}{4\,M_x^2} \right\},$$

$$A_4 = \left\{ \frac{\left[f_y\left(M_y\right)\right]^{-2}}{4} + \frac{\left[f_{z_2}\left(M_{z_2}\right)\right]^{-2} M_y^2}{16\,M_{z_2}^2} - \frac{\left(4\,P_{yz_2} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1} M_y}{4\,M_{z_2}} \right\},$$

$$A_5 = \left\{ \frac{\left[f_y\left(M_y\right)\right]^{-2}}{4} + \frac{\left[f_x\left(M_x\right)\right]^{-2} M_y^2}{4\,M_x^2} + \frac{\left[f_{z_2}\left(M_{z_2}\right)\right]^{-2} M_y^2}{16\,M_{z_2}^2} + \frac{\left[f_{z_1}\left(M_{z_1}\right)\right]^{-2} M_y^2}{16\,M_{z_1}^2} \right.$$
$$- \frac{\left(4\,P_{xy} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_y\left(M_y\right)\right]^{-1} M_y}{2\,M_x} + \frac{\left(4\,P_{yz_1} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_1}\left(M_{z_1}\right)\right]^{-1} M_y}{4\,M_{z_1}}$$
$$- \frac{\left(4\,P_{yz_2} - 1\right)\left[f_y\left(M_y\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1} M_y}{4\,M_{z_2}} - \frac{\left(4\,P_{xz_1} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_{z_1}\left(M_{z_1}\right)\right]^{-1} M_y^2}{4\,M_x\,M_{z_1}}$$
$$\left. + \frac{\left(4\,P_{xz_2} - 1\right)\left[f_x\left(M_x\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1} M_y^2}{4\,M_x\,M_{z_2}} - \frac{\left(4\,P_{z_1z_2} - 1\right)\left[f_{z_1}\left(M_{z_1}\right)\right]^{-1}\left[f_{z_2}\left(M_{z_2}\right)\right]^{-1} M_y^2}{8\,M_{z_1} M_{z_2}} \right\}$$

and

$$A_6 = \left\{ \frac{(4\,P_{xy} - 1)\left[f_x(M_x)\right]^{-1}\left[f_y(M_y)\right]^{-1} M_y}{2\,M_x} - \frac{\left[f_x(M_x)\right]^{-2} M_y^2}{4\,M_x^2} - \frac{\left[f_{z_1}(M_{z_1})\right]^{-2} M_y^2}{16\,M_{z_1}^2} \right.$$

$$- \frac{(4\,P_{yz_1} - 1)\left[f_y(M_y)\right]^{-1}\left[f_{z_1}(M_{z_1})\right]^{-1} M_y}{4\,M_{z_1}} + \frac{(4\,P_{xz_1} - 1)\left[f_x(M_x)\right]^{-1}\left[f_{z_1}(M_{z_1})\right]^{-1} M_y^2}{4\,M_x\,M_{z_1}}$$

$$\left. - \frac{(4\,P_{xz_2} - 1)\left[f_x(M_x)\right]^{-1}\left[f_{z_2}(M_{z_2})\right]^{-1} M_y^2}{4\,M_x\,M_{z_2}} + \frac{(4\,P_{z_1 z_2} - 1)\left[f_{z_1}(M_{z_1})\right]^{-1}\left[f_{z_2}(M_{z_2})\right]^{-1} M_y^2}{8\,M_{z_1} M_{z_2}} \right\}$$

Proof: The mean square errors of the estimators T_{ij} are given by

$$M(T_{ij}) = E\left[T_{ij} - M_y\right]^2 = E\left[\varphi_{ij}\,(T_{iu} - M_y) + (1 - \varphi_{ij})\{T_{jm} - M_y\}\right]^2$$

$$= \varphi_{ij}^{\,2}\, M(T_{iu}) + (1 - \varphi_{ij})^2 M\left[T_{jm}\right] + 2\,\varphi_{ij}\,(1 - \varphi_{ij})\,\mathrm{Cov}(T_{iu}, T_{jm})$$

where $M(T_{iu}) = E\left[T_{iu} - M_y\right]^2$ and $M\left[T_{jm}\right] = E\left[T_{jm} - M_y\right]^2$; (i, j=1, 2)

The estimators T_{iu} and T_{jm} are based on two independent samples of sizes u and m respectively, hence $\mathrm{Cov}(T_{iu}, T_{jm}) = 0$; $(i, j = 1, 2)$. Using large sample approximations assumed in section 4.1 and retaining terms upto the first order of approximations, the expression for $M(T_{iu})$ and $M(T_{jm})$ are obtained as given in equations (12) - (15) and hence the expressions for mean square error of estimators $T_{ij}(i, j = 1, 2)$ are obtained as in equation(11).

Remark 4.2.1: The mean square errors of the estimators T_{ij}(i, j=1, 2) in equation (11) depend on the population parameters $P_{xy}, P_{yz_1}, P_{yz_2}, P_{xz_1}, P_{xz_2}, P_{z_1 z_2}$, $f_x(M_x)$, $f_y(M_y)$, $f_{z_1}(M_{z_1})$ and $f_{z_2}(M_{z_2})$. If these parameters are known, the properties of proposed estimators can be easily studied. Otherwise, which is the most often situation in practice, the unknown population parameters are replaced by their sample estimates. The population proportions $P_{xy}, P_{yz_1}, P_{yz_2}, P_{xz_1}, P_{xz_2}$ and $P_{z_1 z_2}$ can be replaced by the sample estimate $\hat{P}_{xy}, \hat{P}_{xz_1}, \hat{P}_{xz_2}, \hat{P}_{yz_1}, \hat{P}_{yz_2}$ and $\hat{P}_{z_1 z_2}$ and the marginal densities $f_y,(M_y)$, $f_x(M_x)$, $f_{z_1}(M_{z_1})$ and

$f_{z_2}(M_{z_2})$ can be substituted by their kernel estimator or nearest neighbour density estimator or generalized nearest neighbour density estimator related to the kernel estimator Silverman (1986). Here, the marginal densities $f_y(M_y)$, $f_x(M_x)$, $f_{z_1}(M_{z_1})$ and $f_{z_2}(M_{z_2})$ are replaced by $\hat{f}_y(\hat{M}_y(m))$, $\hat{f}_x(\hat{M}_x(n))$, $\hat{f}_{z_1}(\hat{M}_{z_1}(n))$ and $\hat{f}_{z_2}(\hat{M}_{z_2}(n))$ respectively, which are obtained by method of generalized nearest neighbour density estimator related to kernel estimator.

To estimate $f_y(M_y)$, $f_x(M_x)$, $f_{z_1}(M_{z_1})$ and $f_{z_2}(M_{z_2})$, by generalized nearest neighbour density estimator related to the kernel estimator, following procedure has been adopted:

Choose an integer $h \approx n^{1/2}$ and define the distance $\delta(x_1, x_2)$ between two points on the line to be $|x_1 - x_2|$.

For $\hat{M}_x(n)$, define $\delta_1(\hat{M}_x(n)) \le \delta_2(\hat{M}_x(n)) \le --- \le \delta_n(\hat{M}_x(n))$ to be the distances, arranged in ascending order, from $\hat{M}_x(n)$ to the points of the sample.

The generalized nearest neighbour density estimate is defined by

$$\hat{f}(\hat{M}_x(n)) = \frac{1}{n\,\delta_h(\hat{M}_x(n))} \sum_{i=1}^{n} K\left[\frac{\hat{M}_x(n) - x_i}{\delta_h(\hat{M}_x(n))}\right]$$

where the kernel function K, satisfies the condition $\int_{-\infty}^{\infty} K(x)\,dx = 1$.

Here, the kernel function is chosen as Gaussian Kernel given by $K(x) = \frac{1}{2\pi} e^{-\left(\frac{1}{2}x^2\right)}$.

The estimate of $f_y(M_y)$, $f_{z_1}(M_{z_1})$ and $f_{z_2}(M_{z_2})$ can be obtained by the above explained procedure in similar manner.

5. Minimum Mean Square Errors of the Proposed Estimators T_{ij} (i, j =1, 2)

Since the mean square errors of the estimators $T_{ij}\left(i, j = 1, 2\right)$ given in equation (11) are the functions of unknown constants $\varphi_{ij}\left(i, j = 1, 2\right)$, therefore, they are minimized with respect to φ_{ij} and subsequently the optimum values of φ_{ij} are obtained as

$$\varphi_{ij_{opt.}} = \frac{M\left(T_{jm}\right)}{M\left(T_{iu}\right) + M\left(T_{jm}\right)}; \; (i, j = 1, 2) \tag{16}$$

Now substituting the values of $\varphi_{ij_{opt.}}$ in equation (11), we obtain the optimum mean square errors of the estimators $T_{ij}\left(i, j = 1, 2\right)$ as

$$M\left(T_{ij}\right)_{opt.} = \frac{M\left(T_{iu}\right) . M\left(T_{jm}\right)}{M\left(T_{iu}\right) + M\left(T_{jm}\right)}; \; \left(i, j = 1, 2\right) \tag{17}$$

Further, substituting the values of the mean square error of the estimators defined in equation (12) to equation (15) in equation (16) and (17), the simplified values $\varphi_{ij_{opt.}}$ and $M\left(T_{ij}\right)_{opt.}$ are obtained as

$$\varphi_{11_{opt.}} = \frac{\mu_{11}\left[\mu_{11} \, A_3 - \left(A_2 + A_3\right)\right]}{\left[\mu_{11}^2 \, A_3 - \mu_{11}\left(A_2 + A_3 - A_1\right) - A_1\right]} \tag{18}$$

$$\varphi_{12_{opt.}} = \frac{\mu_{12}\left[\mu_{12} \, A_6 - \left(A_5 + A_6\right)\right]}{\left[\mu_{12}^2 \, A_6 - \mu_{12}\left(A_5 + A_6 - A_1\right) - A_1\right]} \tag{19}$$

$$\varphi_{21_{opt.}} = \frac{\mu_{21}\left[\mu_{21} \, A_3 - \left(A_2 + A_3\right)\right]}{\left[\mu_{21}^2 \, A_3 - \mu_{21}\left(A_2 + A_3 - A_4\right) - A_4\right]} \tag{20}$$

$$\varphi_{22_{opt.}} = \frac{\mu_{22}\left[\mu_{22} \, A_6 - \left(A_5 + A_6\right)\right]}{\left[\mu_{22}^2 \, A_6 - \mu_{22}\left(A_5 + A_6 - A_4\right) - A_4\right]} \tag{21}$$

$$M\left(T_{11}\right)_{opt.} = \frac{1}{n}\frac{\left[\mu_{11} \, C_1 - C_2\right]}{\left[\mu_{11}^2 \, A_3 - \mu_{11} \, C_3 - A_1\right]} \tag{22}$$

$$M(T_{12})_{opt.} = \frac{1}{n}\frac{[\mu_{12}\ C_4\ - C_5]}{\left[\mu_{12}^2\ A_6\ - \mu_{12}\ C_6\ - A_1\right]} \tag{23}$$

$$M(T_{21})_{opt.} = \frac{1}{n}\frac{[\mu_{21}\ C_7\ - C_8]}{\left[\mu_{21}^2\ A_3\ - \mu_{21}\ C_9\ - A_4\right]} \tag{24}$$

$$M(T_{22})_{opt.} = \frac{1}{n}\frac{[\mu_{22}\ C_{10}\ - C_{11}]}{\left[\mu_{22}^2\ A_6\ - \mu_{22}\ C_{12}\ - A_4\right]} \tag{25}$$

where

$C_1 = A_1A_3,\ C_2 = A_1A_2\ + A_1A_3,\ C_3 = A_2\ + A_3\ - A_1,\ C_4 = A_1A_6,\ C_5 = A_1A_5\ + A_1A_6,$

$C_6 = A_5\ + A_6\ - A_1,\ C_7 = A_3A_4,\ C_8 = A_2A_4\ + A_3A_4,\ C_9 = A_2\ + A_3\ - A_4,\ C_{10} = A_4A_6$

$C_{11} = A_4A_5\ + A_4A_6$, $C_{12} = A_5\ + A_6\ - A_4$ and $\mu_{i\,j}(i, j = 1, 2)$ are the fractions of the sample drawn afresh at the current(second) occasion.

Remark 5.1: $M(T_{i\,j})_{opt.}$ derived in equation (22) - (25) are the functions of $\mu_{i\,j}(i, j = 1, 2)$. To estimate the population median on each occasion the better choices of $\mu_{i\,j}(i, j = 1, 2)$ are 1(case of no matching); however, to estimate the change in median from one occasion to other, $\mu_{i\,j}(i, j = 1, 2)$ should be 0(case of complete matching). But intuition suggests that an optimum choices of $\mu_{i\,j}(i, j = 1, 2)$ are desired to devise the amicable strategy for both the problems simultaneously.

6. Optimum Replacement Strategies for the Estimators $T_{i\,j}$ (i, j =1, 2)

The key design parameter affecting the estimates of change is the overlap between successive samples. Maintaining high overlap between repeats of a survey is operationally convenient, since many sampled units have been located and have some experience in the survey. Hence to decide about the optimum value of $\mu_{i\,j}(i, j = 1, 2)$ (fractions of samples to be drawn afresh on current occasion) so that M_y may be estimated with maximum precision and minimum cost, we minimize the mean square errors $M(T_{ij})_{opt.}\ (i, j = 1, 2)$ in equation (22) to (25) with respect to $\mu_{ij}\ (i, j=1, 2)$ respectively.

The optimum value of $\mu_{ij}\ (i, j=1, 2)$ so obtained is one of the two roots given by

$$\hat{\mu}_{11} = \frac{D_2 \pm \sqrt{D_2^2 - D_1 D_3}}{D_1} \tag{26}$$

$$\hat{\mu}_{12} = \frac{D_5 \pm \sqrt{D_5^2 - D_4 D_6}}{D_4} \tag{27}$$

$$\hat{\mu}_{21} = \frac{D_8 \pm \sqrt{D_8^2 - D_7 D_9}}{D_7} \tag{28}$$

$$\hat{\mu}_{22} = \frac{D_{11} \pm \sqrt{D_{11}^2 - D_{10} D_{12}}}{D_{10}} \tag{29}$$

where $D_1 = A_3C_1,\ D_2 = A_3C_2,\ D_3 = A_1C_1 + C_2C_3,\ D_4 = A_6C_4,\ D_5 = A_6C_5,\ D_6 = A_1C_4 + C_5C_6$ $D_7 = A_3C_7,\ D_8 = A_3C_8,\ D_9 = A_4C_7 + C_8C_9,\ D_{10} = A_6C_{10},\ D_{11} = A_6C_{11}$ and $D_{12} = A_4C_{10} + C_{11}C_{12}$.

The real values of $\hat{\mu}_{ij}\ (i, j=1, 2)$ exist, iff $D_2^2 - D_1 D_3 \geq 0,\ D_5^2 - D_4 D_6 \geq 0,\ D_8^2 - D_7 D_9 \geq 0$, and $D_{11}^2 - D_{10} D_{12} \geq 0$. For any situation, which satisfies these conditions, two real values of $\hat{\mu}_{ij}\ (i, j = 1, 2)$ may be possible , hence to choose a value of $\hat{\mu}_{ij}\ (i, j = 1, 2)$, it should be taken care of that $\hat{\mu}_{ij} \in (0,\ 1)$, all other values of $\hat{\mu}_{ij}\ (i, j = 1, 2)$ are inadmissible. If both the real values of $\hat{\mu}_{ij}\ (i, j=1, 2)$ are admissible, the lowest one will be the best choice as it reduces the total cost of the survey. Substituting the admissible value of $\hat{\mu}_{ij}$ say $\mu_{ij}^{(0)}\ (i, j=1, 2)$ from equation (26) to (29) in equation (22) to (25) respectively , we get the optimum values of the mean square errors of the estimators $T_{ij}\ (i, j=1, 2)$ with respect to φ_{ij} as well as $\mu_{ij}\ (i, j=1, 2)$ which are given as

$$M(T_{11})^{*}_{opt.} = \frac{\left[\mu_{11}^{(0)}\ C_1\ -\ C_2\right]}{n\left[\mu_{11}^{(0)\,2}\ A_3\ -\ \mu_{11}^{(0)}\ C_3\ -\ A_1\right]} \tag{30}$$

$$M(T_{12})^{*}_{opt.} = \frac{\left[\mu_{12}^{(0)}\ C_4\ -\ C_5\right]}{n\left[\mu_{12}^{(0)\,2}\ A_6\ -\ \mu_{12}^{(0)}\ C_6\ -\ A_1\right]} \tag{31}$$

$$M(T_{21})^{*}_{opt.} = \frac{\left[\mu_{21}^{(0)}\ C_7\ -\ C_8\right]}{n\left[\mu_{21}^{(0)\,2}\ A_3\ -\ \mu_{21}^{(0)}\ C_9\ -\ A_4\right]} \tag{32}$$

$$M(T_{22})^{*}_{opt.} = \frac{\left[\mu_{22}^{(0)}\ C_{10}\ -\ C_{11}\right]}{n\left[\mu_{22}^{(0)\,2}\ A_6\ -\ \mu_{22}^{(0)}\ C_{12}\ -\ A_4\right]} \tag{33}$$

7. Efficiency Comparison

To evaluate the performance of the proposed estimators, the estimators $T_{ij}\,(i, j = 1, 2)$ at optimum conditions are compared with respect to (i) the sample median estimator $\hat{M}_y(n)$, when there is no matching from previous occasion and (ii) the ratio type estimator Δ proposed by Singh et al. (2007) for second quantile, where no additional auxiliary information was used at any occasion and is given by

$$\Delta = \psi\,\hat{M}_y(u) + (1 - \psi)\hat{M}_x(n)\left(\frac{\hat{M}_y(m)}{\hat{M}_x(m)}\right) \tag{34}$$

where ψ is an unknown constant to be determined so as to minimise the mean square error of the estimator Δ. Since, $\hat{M}_y(n)$ is unbiased and Δ is biased for population median, so

variance of $\hat{M}_y(n)$ and mean square error of the estimator Δ at optimum conditions are given as

$$V\left(\hat{M}_y(n)\right) = \frac{1}{n}\frac{\left[f_y\left(M_y\right)\right]^{-2}}{4} \tag{35}$$

and $$M(\Delta)^{*}_{opt.} = \frac{[\mu_{\Delta} J_1 - J_2]}{n[\mu_{\Delta}^{2} I_3 - \mu_{\Delta} J_3 - I_1]} \tag{36}$$

where $\mu_{\Delta} = \frac{H_2 \pm \sqrt{H_2^2 - H_1 H_3}}{H_1}$, $H_1 = J_1 I_3$, $H_2 = J_2 I_3$, $H_3 = I_1 J_1 + J_2 J_3$, $J_1 = I_1 I_3$, $J_2 = I_1 (I_2 + I_3)$,

$J_3 = I_2 + I_3 - I_1$, $I_1 = \frac{[f_y(M_y)]^{-2}}{4}$, $I_2 = \frac{[f_y(M_y)]^{-2}}{4} + \frac{[f_x(M_x)]^{-2} M_y^2}{4 M_x^2} - \frac{(4 P_{xy} - 1)[f_x(M_x)]^{-1}[f_y(M_y)]^{-1} M_y}{2 M_x}$

and $I_3 = \frac{(4 P_{xy} - 1)[f_x(M_x)]^{-1}[f_y(M_y)]^{-1} M_y}{2 M_x} - \frac{[f_x(M_x)]^{-2} M_y^2}{4 M_x^2}$.

The percent relative efficiencies $E_{ij}^{(1)}$ and $E_{ij}^{(2)}$ of the estimators $T_{ij}(i, j = 1, 2)$ (under their respective optimum conditions) with respect to $\hat{M}_y(n)$ and Δ are respectively given by

$$E_{ij}^{(1)} = \frac{V(\hat{M}_y(n))}{M(T_{ij})^{*}_{opt.}} \times 100 \quad \text{and} \quad E_{ij}^{(2)} = \frac{M(\Delta)^{*}_{opt.}}{M(T_{ij})^{*}_{opt.}} \times 100; \ (i, j=1, 2) \tag{37}$$

8. Empirical Illustrations and Monte Carlo Simulation

Empirical validation can be carried out by Monte Carlo Simulation. Real life situation of completely known finite population has been considered.

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

The population comprise of N = 51 states of United States. Let x_i be the Percentage of Advanced Degree Holders or More during 1990 in the i^{th} state of U. S., y_i represent the Percentage of Advanced Degree Holders or More during 2009 in the i^{th} state of U. S., z_{1_i} denote Percentage of Bachelor Degree Holders or More during 1990 in the i^{th} state of U. S. and z_{2_i} denote the Percentage of Bachelor Degree Holders or More during 2009 in the i^{th} state of U. S and The data are presented in Figure 1.

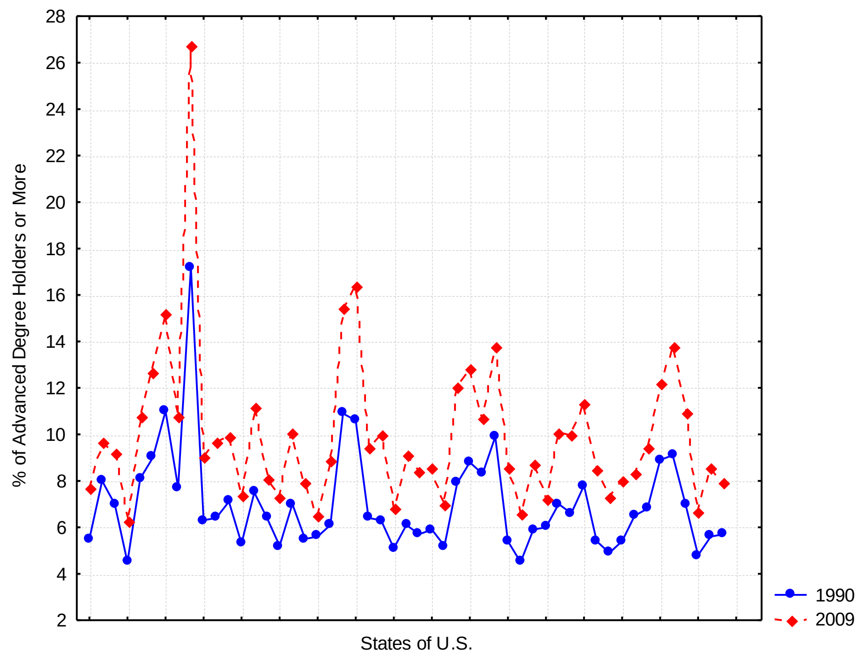


Figure 1: Percentage Advanced Degree Holders or More during 1990 and 2009 versus different states of United States.

For the considered population, the optimum values of μ_{ij} (i, j = 1, 2) defined in equation (26) to (29) and percent relative efficiencies $E_{ij}^{(1)}$ and $E_{ij}^{(2)}$ defined in equation (37) of T_{ij}(i, j =1, 2) (under their respective optimality conditions) with respect to $\hat{M}_y(n)$ and Δ have been computed and are presented in Table-2.

To validate the empirical results quoted in Table 2, Monte Carlo simulation have also been performed.5000 samples of size n=20 states are selected using simple random sampling without replacement in the year 1990. The sample medians $\hat{M}_{x|k}(n)$ and $\hat{M}_{z_1|k}(n)$, k =1, 2, - - -,5000 are computed. From each one of the selected samples, m=17 states are retained and new u=3 states are selected out of N – n =51 – 20 = 31 states of U.S. using simple random sampling without replacement in the year 2009. From the m units retained in the sample at the current occasion, the sample medians $\hat{M}_{x|k}(m)$, $\hat{M}_{y|k}(m)$, $\hat{M}_{z_1|k}(m)$ and $\hat{M}_{z_2|k}(m)$, k = 1, 2,- - -,5000 are computed. From the new unmatched units selected on the current occasion the sample medians $\hat{M}_{y|k}(u)$ and $\hat{M}_{z_2|k}(u)$, k = 1, 2,- - - ,5000 are also calculated. The parameters φ and ψ are selected between 0.1 and 0.9 with a step of 0.1.

The percent relative efficiencies of the proposed estimators T_{ij} with respect to $\hat{M}_y(n)$ and Δ are obtained as a result of above simulation and are respectively given as:

$$E_{ij}(1) = \frac{\sum_{k=1}^{5000}\left[\hat{M}_{y|k}(n) - M_y\right]^2}{\sum_{k=1}^{5000}\left[T_{ijk} - M_y\right]^2} \times 100 \quad \text{and} \quad E_{ij}(2) = \frac{\sum_{k=1}^{5000}\left[\Delta_k - M_y\right]^2}{\sum_{k=1}^{5000}\left[T_{ijk} - M_y\right]^2} \times 100\,; \quad (i, j=1, 2)$$

For better analysis, the above simulation experiments were repeated for different choices of μ. For convenience the different choices of μ are considered as different sets for the considered Population which is shown below:

Sets	Population
I	n=20, $\mu = 0.15$, (m =17, u = 3)
II	n=20, $\mu = 0.20$, (m = 16, u = 4)
III	n=20, $\mu = 0.35$, (m = 13, u = 7)
IV	n=20, $\mu = 0.50$, (m = 10, u = 10)

The simulation results obtained are presented in Table-3 to Table-7.

Table 1: Descriptive statistics for the population considered

	% of Advanced Degree Holders or More(1990) (x)	**% of Advanced Degree Holders or More(2009) (y)**	**% of Bachelor's Degree or More (1990) (z_1)**	**% of Bachelor's Degree or More(2009) (z_2)**
Mean	5.7	10.00	20.00	27.40
Median	6.40	7.90	19.30	26.30
Standard deviation	4.70	11.23	16.98	30.46
Kurtosis	8.43	11.04	0.79	2.70
Skewness	2.34	2.69	0.70	1.09
Minimum	5.7	6.30	12.30	17.1
Maximum	17.2	26.7	33.37	48.2
Count	51	51	51	51

Table 2: Comparison of the proposed estimators T_{ij} (at optimum conditions) with respect to the estimators $\hat{M}_y(n)$ and Δ (at their respective optimum conditions)

$\mu_{11}^{(0)}$	*
$\mu_{12}^{(0)}$	0.8389
$\mu_{21}^{(0)}$	0.5278
$\mu_{22}^{(0)}$	0.5603
$E_{11}^{(1)}$	-
$E_{12}^{(1)}$	200.75
$E_{21}^{(1)}$	155.52
$E_{22}^{(1)}$	165.02
$E_{11}^{(2)}$	-
$E_{12}^{(2)}$	171.79
$E_{21}^{(2)}$	133.08
$E_{22}^{(2)}$	141.21

Note: '*' indicates that $\mu_{ij}^{(0)}$; $(i, j = 1, 2)$ do not exist.

Table 3: Monte Carlo Simulation results when the proposed estimator T_{ij} is compared to $\hat{M}_y(n)$.

φ \ SET		I	II	III	IV
0.1	$E_{11}(1)$	157.64	136.42	139.61	104.70
	$E_{12}(1)$	139.56	135.09	144.90	148.55
	$E_{21}(1)$	155.56	137.23	142.14	106.44
	$E_{22}(1)$	137.61	135.68	146.87	151.28
0.2	$E_{11}(1)$	161.27	145.61	148.14	119.39
	$E_{12}(1)$	147.08	144.78	153.24	167.79
	$E_{21}(1)$	161.27	144.25	148.66	121.37
	$E_{22}(1)$	142.42	143.43	152.85	170.72
0.3	$E_{11}(1)$	171.41	152.14	152.99	133.24
	$E_{12}(1)$	152.23	151.62	157.13	185.83
	$E_{21}(1)$	161.68	146.36	147.98	134.36
	$E_{22}(1)$	143.81	145.93	150.68	187.01
0.4	$E_{11}(1)$	172.75	151.39	153.37	146.51
	$E_{12}(1)$	153.29	151.85	157.52	202.15
	$E_{21}(1)$	157.17	138.96	141.53	145.79
	$E_{22}(1)$	140.00	139.43	143.68	199.08
0.5	$E_{11}(1)$	169.68	148.53	148.19	159.08
	$E_{12}(1)$	151.22	148.99	151.80	215.89
	$E_{21}(1)$	148.70	129.43	127.97	154.54
	$E_{22}(1)$	133.39	129.74	129.45	205.79
0.6	$E_{11}(1)$	162.03	140.99	138.28	171.10
	$E_{12}(1)$	145.87	141.54	141.20	227.47
	$E_{21}(1)$	136.36	115.71	112.57	160.04
	$E_{22}(1)$	123.84	116.09	113.30	206.50
0.7	$E_{11}(1)$	154.69	131.88	124.56	179.50
	$E_{12}(1)$	140.61	132.26	126.70	232.64
	$E_{21}(1)$	125.89	103.21	**	160.75
	$E_{22}(1)$	115.62	103.44	**	200.20
0.8	$E_{11}(1)$	144.46	119.52	112.15	182.42
	$E_{12}(1)$	132.79	119.77	113.77	229.90
	$E_{21}(1)$	113.79	**	**	156.07
	$E_{22}(1)$	105.73	**	**	187.95
0.9	$E_{11}(1)$	133.55	107.93	**	180.42
	$E_{12}(1)$	124.30	108.12	**	220.39
	$E_{21}(1)$	102.91	**	**	147.02
	$E_{22}(1)$	**	**	**	171.21

Note: "" indicates no gain.**

Table 4: Monte Carlo Simulation results when the proposed estimator T_{11} is compared to the estimator Δ

φ	ψ / SET	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	**I**	166.50	164.58	172.38	188.62	228.93	277.81	333.46	422.99	479.54
	II	140.62	135.80	149.73	184.74	244.06	347.10	391.73	556.52	731.76
	III	134.98	126.52	138.99	197.52	258.06	362.92	511.86	668.66	895.18
	IV	116.54	100.35	**	**	**	**	123.84	160.06	211.31
0.2	**I**	174.84	170.32	179.46	198.15	246.31	293.62	336.28	414.78	507.51
	II	147.81	141.46	157.48	196.01	256.57	350.54	433.21	581.35	756.12
	III	144.43	134.92	152.62	210.36	281.48	393.28	540.90	716.93	913.98
	IV	131.84	104.09	**	**	**	107.11	134.64	183.01	233.14
0.3	**I**	178.86	174.28	183.33	202.86	246.40	300.03	354.10	434.50	518.11
	II	152.95	145.17	161.03	200.44	263.20	355.79	454.44	603.50	776.33
	III	151.01	140.19	159.32	218.27	293.69	408.04	566.13	715.27	942.02
	IV	148.55	116.42	**	**	100.93	118.78	154.26	199.88	259.18
0.4	**I**	179.43	175.46	183.29	202.99	248.29	299.47	353.28	432.78	515.88
	II	152.22	145.54	160.58	200.91	265.35	357.71	460.92	610.86	760.58
	III	151.43	139.34	158.79	216.11	295.25	409.66	564.60	742.28	941.09
	IV	163.16	129.19	107.79	100.14	110.62	131.32	170.91	220.06	282.55
0.5	**I**	175.36	172.12	179.58	199.54	242.51	291.57	345.79	420.66	515.67
	II	149.34	142.89	157.15	197.21	261.35	352.07	452.17	607.09	749.93
	III	145.63	133.76	153.37	206.83	284.12	393.77	537.64	713.06	907.43
	IV	177.53	139.85	116.86	108.35	120.59	143.48	187.29	239.67	310.80
0.6	**I**	167.19	164.42	172.61	191.70	232.03	278.84	333.16	405.10	492.13
	II	141.98	136.24	149.74	187.96	246.67	333.07	429.17	569.91	709.42
	III	136.23	124.27	143.46	192.75	265.70	368.16	501.30	661.34	848.74
	IV	190.07	149.10	124.74	116.23	128.75	152.66	199.81	257.62	332.90
0.7	**I**	159.37	155.34	162.37	181.70	219.28	263.40	313.41	387.07	462.57
	II	132.92	125.95	138.83	174.34	229.98	308.49	397.17	528.32	661.19
	III	123.11	112.21	129.17	173.54	240.65	337.10	453.64	604.41	775.34
	IV	199.78	155.47	130.26	121.68	134.63	160.18	209.69	270.21	346.17
0.8	**I**	148.49	144.04	151.56	169.90	204.19	245.76	292.35	357.20	431.24
	II	120.56	114.86	126.15	160.31	210.25	284.17	360.20	477.84	601.11
	III	110.85	100.36	115.34	154.33	214.13	300.74	403.23	540.34	688.10
	IV	203.38	157.98	132.34	124.54	137.06	162.68	212.57	275.04	352.95
0.9	**I**	137.22	132.63	139.82	155.98	188.41	224.81	268.21	327.37	397.74
	II	108.66	104.07	114.83	145.37	189.19	255.82	325.72	431.0	544.49
	III	**	**	101.74	13525	187.25	265.51	353.08	470.35	600.77
	IV	201.06	157.02	131.81	123.53	135.94	161.63	211.07	272.07	346.66

Note: "" indicates no gain.**

Table 5: Monte Carlo Simulation results when the proposed estimator T_{12} is compared to the estimator Δ

φ	SET \ ψ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	**I**	147.41	145.42	152.84	167.96	205.15	245.85	287.97	379.09	429.76
	II	139.26	130.52	147.81	182.62	247.25	348.99	396.14	552.83	737.42
	III	140.10	125.66	142.88	200.32	266.60	371.37	520.99	675.04	907.49
	IV	165.36	131.09	111.73	105.05	110.30	135.83	168.97	217.91	293.82
0.2	**I**	154.34	150.08	158.57	176.72	218.11	257.30	298.15	368.28	453.74
	II	146.96	138.61	155.49	196.26	260.63	353.46	438.36	579.23	761.42
	III	149.40	134.87	155.77	213.36	289.19	404.90	556.13	728.97	923.68
	IV	185.30	146.82	124.64	116.52	124.85	149.48	189.38	248.68	324.44
0.3	**I**	158.85	153.55	162.09	180.36	217.48	263.04	313.32	385.31	458.56
	II	152.43	143.57	159.98	201.01	266.61	357.11	456.77	602.15	780.56
	III	155.10	141.28	162.21	220.98	301.04	420.15	579.05	770.08	952.36
	IV	207.19	163.58	137.49	128.45	139.17	164.48	214.91	273.60	359.43
0.4	**I**	159.22	154.82	162.16	180.33	219.78	262.40	314.13	348.73	456.35
	II	152.68	144.84	159.77	201.72	267.74	359.65	462.40	608.93	764.93
	III	155.53	140.76	161.10	218.47	302.35	419.97	576.67	759.17	956.37
	IV	225.13	180.10	149.75	139.74	151.07	179.92	235.54	298.40	388.24
0.5	**I**	156.28	153.07	160.21	178.21	215.58	257.46	308.89	374.79	460.02
	II	149.80	142.50	156.87	198.03	262.93	354.07	452.77	604.91	755.09
	III	149.17	135.50	155.26	209.19	289.74	401.92	548.37	726.64	921.76
	IV	240.93	191.53	159.71	149.10	162.46	193.60	253.76	320.98	420.27
0.6	**I**	150.51	147.50	155.47	172.77	208.13	248.55	299.85	363.74	442.23
	II	142.53	135.85	149.68	188.86	247.72	335.34	429.69	568.05	713.58
	III	139.11	125.74	144.98	194.42	269.92	374.53	511.26	672.23	861.67
	IV	252.69	200.10	166.61	156.52	170.26	201.90	264.86	338.34	441.53
0.7	**I**	144.87	140.89	147.95	165.22	198.88	263.40	284.91	350.90	419.95
	II	133.32	125.80	138.93	175.14	230.76	310.13	398.21	527.50	665.34
	III	125.23	113.28	130.46	174.93	244.27	342.04	460.79	612.47	785.09
	IV	258.92	202.87	169.77	159.50	173.88	206.66	271.39	346.66	448.02
0.8	**I**	136.50	132.28	139.52	156.15	187.55	245.76	268.76	327.92	396.09
	II	120.82	114.81	127.12	160.96	210.85	285.41	360.90	477.31	604.09
	III	112.45	101.14	116.26	155.42	216.76	304.15	408.39	546.84	695.73
	IV	256.31	200.14	167.41	158.42	172.14	204.04	267.17	343.12	444.21
0.9	**I**	127.70	123.33	130.28	145.20	175.21	208.21	249.65	304.26	369.99
	II	108.85	104.07	114.96	145.06	189.64	255.82	326.32	430.68	546.88
	III	**	**	102.40	136.05	189.25	267.93	357.08	474.84	605.96
	IV	245.56	192.90	161.89	152.22	165.80	196.86	258.07	329.86	423.49

Note: "" indicates no gain.**

Table 6: Monte Carlo Simulation results when the proposed estimator T_{21} is compared to the estimator Δ

φ	ψ / SET	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	**I**	164.30	162.04	170.17	186.41	225.19	273.63	318.27	416.48	473.41
	II	141.30	137.16	151.29	185.73	247.77	351.46	396.87	563.10	741.51
	III	137.43	128.55	141.92	200.86	263.88	369.83	522.15	679.07	910.72
	IV	118.48	105.56	**	**	**	**	125.42	162.57	214.0
0.2	**I**	169.22	164.43	173.74	191.94	237.26	283.01	325.54	401.66	492.17
	II	146.42	140.67	157.02	194.23	256.77	349.22	431.44	580.83	757.19
	III	144.94	134.99	153.20	210.47	282.25	393.85	543.97	717.48	920.49
	IV	134.04	117.32	**	**	**	108.93	136.67	185.99	236.66
0.3	**I**	168.71	163.74	172.99	191.54	232.10	281.87	333.45	407.94	489.82
	II	147.14	139.76	155.67	192.73	254.65	343.14	437.95	581.08	749.47
	III	146.06	134.74	154.03	209.81	233.18	396.35	547.58	718.77	917.21
	IV	149.80	128.47	100.11	**	101.37	120.47	155.25	201.64	261.26
0.4	**I**	163.25	159.18	167.59	183.77	224.98	271.79	321.96	393.64	473.0
	II	139.73	133.36	148.84	185.01	244.59	328.50	424.73	561.31	707.64
	III	139.74	127.01	145.41	197.85	270.95	376.30	518.17	675.94	871.26
	IV	162.36	135.92	107.72	100.04	109.68	131.46	169.16	218.55	281.58
0.5	**I**	153.68	151.0	158.66	175.83	211.75	255.99	304.54	370.36	452.13
	II	130.14	125.18	137.85	171.75	227.73	304.54	395.12	526.53	658.63
	III	125.76	114.80	131.95	178.34	246.39	341.66	465.45	612.26	789.09
	IV	172.46	140.33	114.30	105.17	116.65	140.05	180.77	232.54	301.70
0.6	**I**	140.70	138.80	146.75	162.06	195.09	236.45	282.72	343.0	417.54
	II	116.53	113.59	123.84	154.90	204.93	273.57	356.37	472.16	591.84
	III	110.90	100.53	115.38	157.30	215.80	300.35	408.67	537.16	688.69
	IV	177.79	139.94	117.65	109.17	120.37	144.22	186.37	240.64	311.57
0.7	**I**	129.70	126.10	132.55	149.58	177.75	215.08	257.43	314.25	380.18
	II	104.03	100.14	109.0	137.33	180.63	240.87	312.68	415.49	523.70
	III	**	**	**	133.60	185.24	258.53	351.47	464.06	595.42
	IV	178.91	135.46	117.81	109.46	120.76	144.66	186.96	241.53	311.63
0.8	**I**	116.88	113.24	119.85	135.13	159.69	193.70	232.06	282.62	342.43
	II	**	**	**	119.77	158.24	212.80	273.82	361.75	454.69
	III	**	**	**	113.23	156.67	219.80	298.48	397.24	506.82
	IV	174.0	140.86	113.71	106.67	117.35	140.40	181.37	234.29	302.86
0.9	**I**	105.73	101.54	107.40	120.34	143.32	173.41	207.59	253.34	306.09
	II	**	**	**	103.73	137.32	182.59	237.85	312.02	391.95
	III	**	**	**	**	132.98	187.26	250.32	336.43	430.48
	IV	163.85	128.33	107.93	100.65	110.64	132.25	171.50	221.13	285.08

Note: "" indicates no gain.**

Table 7: Monte Carlo Simulation results when the proposed estimator T_{22} is compared to the estimator Δ

φ	ψ / SET	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	**I**	145.34	142.91	150.34	165.76	201.25	241.72	283.71	372.27	422.76
	II	139.86	131.72	149.28	183.75	251.11	353.73	401.21	560.08	747.12
	III	142.01	126.68	145.84	202.37	271.64	376.86	528.24	684.40	918.32
	IV	168.40	133.02	113.61	106.71	112.35	138.09	171.15	222.02	279.44
0.2	**I**	149.44	144.84	152.99	170.90	209.80	247.94	288.51	355.91	438.58
	II	145.59	137.80	155.01	194.71	260.76	352.53	436.69	579.70	761.35
	III	149.03	133.54	155.58	211.23	288.36	402.60	553.84	723.93	922.12
	IV	188.53	148.72	126.73	118.13	126.62	152.41	192.27	253.35	328.86
0.3	**I**	150.07	144.53	152.80	170.30	204.83	247.52	295.52	361.88	432.91
	II	146.71	138.09	154.92	193.58	257.84	344.89	441.02	580.43	754.38
	III	148.73	134.21	155.45	209.84	287.19	404.09	555.0	727.42	918.78
	IV	208.51	164.05	138.74	128.89	139.03	166.56	25.66	275.54	359.87
0.4	**I**	145.42	141.02	148.49	165.50	199.88	239.20	287.11	350.73	418.86
	II	140.20	132.59	148.35	185.95	246.69	330.71	426.58	559.47	711.77
	III	141.86	126.81	146.01	197.66	273.69	381.30	523.37	683.22	875.16
	IV	221.71	177.43	148.03	137.60	148.13	178.21	231.03	293.93	381.61
0.5	**I**	137.85	135.10	142.12	157.92	189.45	227.59	273.41	331.58	405.16
	II	130.45	124.88	137.89	172.41	229.23	307.19	396.44	524.90	662.82
	III	127.21	114.85	132.21	177.97	248.29	344.66	468.77	615.55	792.22
	IV	229.67	182.27	153.0	142.04	154.08	185.10	240.28	305.89	398.49
0.6	**I**	127.78	125.60	133.10	147.23	176.64	212.63	256.40	310.51	377.76
	II	116.90	113.38	124.03	155.66	205.91	275.77	357.22	470.98	594.70
	III	111.62	100.0	115.44	156.69	216.72	302.02	411.56	539.55	691.10
	IV	229.39	182.55	152.29	142.69	154.53	185.09	240.15	307.21	399.19
0.7	**I**	119.13	115.58	121.87	137.22	163.02	196.13	236.15	287.92	348.08
	II	104.27	**	109.23	137.97	181.22	242.29	313.73	414.96	526.11
	III	**	**	**	133.30	185.96	259.46	353.0	465.10	596.38
	IV	222.81	175.30	147.60	137.83	150.17	179.11	232.76	298.37	386.84
0.8	**I**	108.68	105.16	111.41	125.46	148.49	179.11	215.52	262.33	317.59
	II	**	**	**	120.23	158.65	212.99	274.54	361.38	456.32
	III	**	**	**	113.03	157.10	220.20	299.43	398.12	507.52
	IV	209.54	163.69	137.32	129.39	141.03	168.18	217.93	279.63	363.49
0.9	**I**	100.05	101.54	101.09	113.19	134.89	162.45	195.18	237.99	287.66
	II	**	**	**	104.06	137.61	183.21	238.41	311.82	393.15
	III	**	**	**	109.51	133.25	187.47	250.79	337.01	430.48
	IV	190.81	150.13	126.23	100.65	128.73	153.38	199.50	255.99	331.80

Note: "" indicates no gain.**

9. Mutual Comparison of the Proposed Estimators T_{ij} (i, j =1, 2)

The performances of the proposed estimators T_{ij} (i, j =1, 2) have been elaborated empirically as well as through simulation studies in above Section 8 and the results obtained are presented in Table 2 to Table 7. In this section the mutual comparison of the four proposed estimators have been elaborated though different graphs given in Figure 2 to Figure 5.

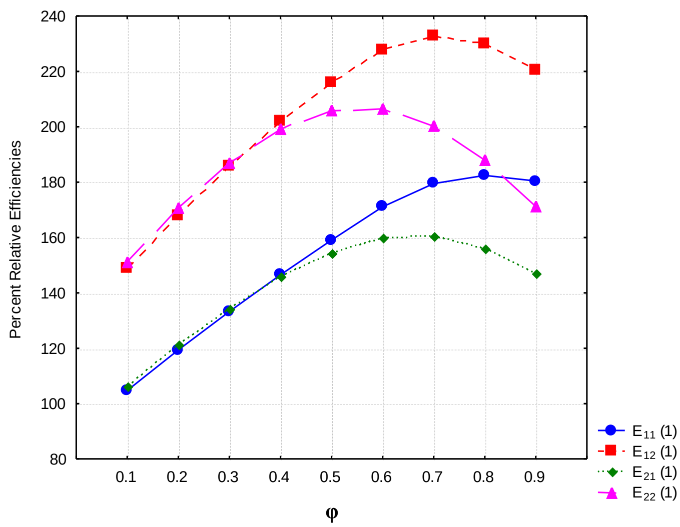


Figure 2: Mutual Comparison of Proposed Estimator T_{ij} (i, j =1, 2) **when compared with the estimator** $\hat{M}_y(n)$ **for set-IV.**

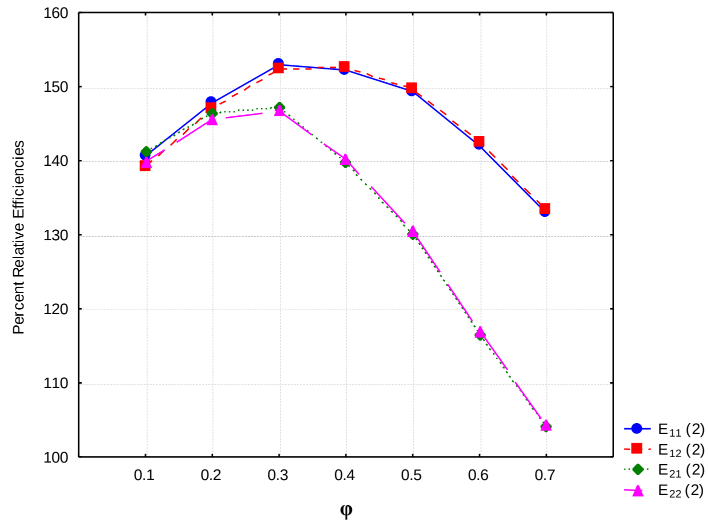


Figure 3: Mutual Comparison of Proposed Estimators T_{ij} $(i, j = 1, 2)$ when compared with the estimator Δ for $\psi = 0.1$ for set-II.

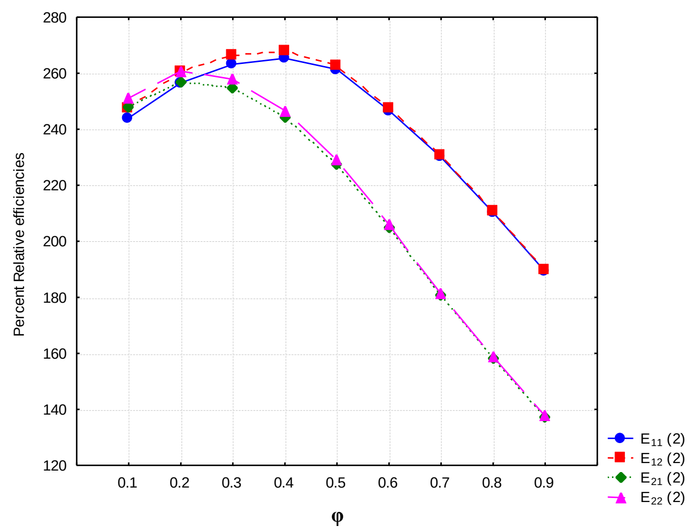


Figure 4: Mutual Comparison of Proposed Estimators T_{ij} $(i, j = 1, 2)$ when compared with the estimator Δ for $\psi = 0.5$ for set-II.

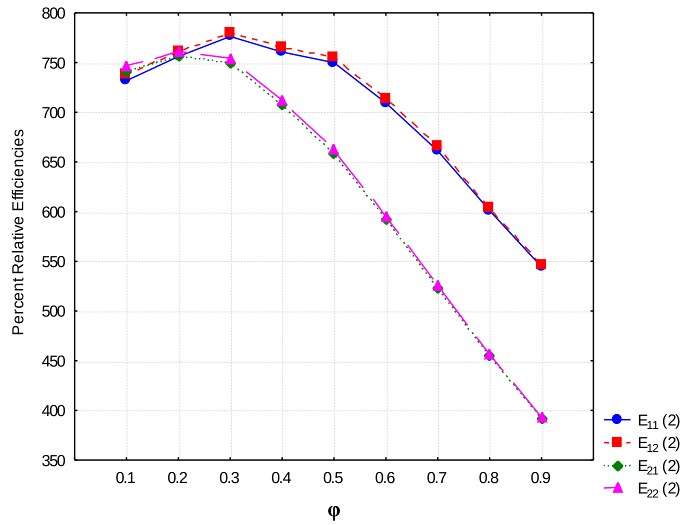


Figure 5: Mutual Comparison of Proposed Estimators T_{ij} (i, j =1, 2) **when compared with the estimator** Δ **for** $\psi = 0.9$ **for set-II.**

10. Interpretation of Results

The following interpretation can be drawn from Tables 2-7 and Figure 2-5:

(1) From Table- 2, it is observed that

(a) Optimum values $\mu_{12}^{(0)}$, $\mu_{21}^{(0)}$ and $\mu_{22}^{(0)}$ for the estimators T_{12}, T_{21} and T_{22} respectively exist for the considered population which justifies the applicability of the proposed estimators T_{12}, T_{21} and T_{22} at optimum conditions. However, the optimum value $\mu_{11}^{(0)}$ for the estimator T_{11} does not exist for the considered population.

(b) Appreciable gain is observed in terms of precision indicating the proposed estimators T_{12}, T_{21}, T_{22} (at their respective optimal conditions) are preferable over the estimator $\hat{M}_y(n)$ and Δ (at optimal conditions). This result justifies the use of additional auxiliary information at both occasions which is dynamic over time in two occasion successive sampling.

(c) The values for $E_{11}^{(1)}$ and $E_{11}^{(2)}$ cannot be calculated as optimum value $\mu_{11}^{(0)}$ does not exist but simulation study vindicated in Tables 3-7 magnify the applicability of proposed estimator T_{11} over sample median estimator $\hat{M}_y(n)$ and the estimator Δ.

(2) From Table-3, it can be seen that, when T_{ij} (i, j =1, 2) is compared with sample median estimator $\hat{M}_y(n)$

(a) $E_{11}(1)$, $E_{12}(1)$, $E_{21}(1)$, $E_{22}(1)$ first increase and then decrease as φ increases for all sets.

(b) For fixed value of φ, $E_{11}(1)$ and $E_{21}(1)$ show no fixed behaviour as the value of μ is increased.

(c) $E_{12}(1)$ and $E_{22}(1)$ increase as μ increases.

(3) From Table-4, when T_{11} is compared with the estimator Δ, we see that

(a) $E_{11}(2)$ increases as φ increases for all choices of ψ.

(b) For fixed choices of φ as ψ increases the value of $E_{11}(2)$ increases.

(c) As μ is increased $E_{11}(2)$ decreases.

(4) From Table-5, when T_{12} is compared with the estimator Δ, we observe that

(a) $E_{12}(2)$ increases for all the sets as φ increases for all choices of ψ.

(b) As ψ increases $E_{12}(2)$ also increases for all sets except for some of the combinations of φ and ψ.

(c) No fixed pattern is observed for $E_{12}(2)$ as μ is increased.

(5) From Table-6, when T_{21} is compared with the estimator Δ, it can be seen that

(a) For all choices of ψ the value of $E_{21}(2)$ first increases and then decreases as φ increases for all sets except for set IV.

(b) For different choices of φ as ψ increases, the value of $E_{21}(2)$ also increases for set I, II and III.

(c) For set IV, $E_{21}(2)$ first decreases as ψ increases and then increases for all choices of φ.

(d) As for all choices of φ and ψ as μ increases, the value of $E_{21}(2)$ decreases.

(6) From Table-7, it can be concluded that

(a) $E_{22}(2)$ first increases as φ increases and then decreases for different choices of ψ for all the four sets.

(b) As ψ increases $E_{22}(2)$ also increases for all sets and for all choices of φ.

(c) For set IV $E_{22}(2)$ first decreases and then increases as ψ increases for all choices of φ .

(d) No fixed behaviour is observed for $E_{22}(2)$ as portion of sample drawn afresh at current occasion increases.

(7) The mutual comparison of the four proposed estimators T_{ij} (i, j=1, 2) in Figure 2 to Figure 5, show that the estimator T_{22} comes out to be the best estimator amongst all the four proposed estimators when they are compared with sample median estimator $\hat{M}_y(n)$, since it is the most consistent and having greater precision but when T_{ij} (i, j =1, 2) are compared with estimator Δ, T_{12} comes out be the best as it possess largest gain over other proposed estimators and considerably consistent in nature for all combinations of φ, ψ and μ. It has also been found that the percent relative efficiency of the estimator T_{12} increases as the fraction of sample drawn at current occasion decreases and vice versa which exactly justifies the basic principles of sampling over successive occasions.

10. Conclusion

From the preceding interpretations, it may be concluded that the use of exponential ratio type estimators for the estimation of population median at current occasion in two occasion successive sampling is quite feasible as vindicated through empirical and simulation results. The use of highly correlated auxiliary information which is dynamic over time is highly rewarding in terms of precision. The mutual comparison of the proposed estimators indicates that the estimators utilizing more exponential ratio type structures perform better. It has also been observed that the estimator T_{22} in which maximum utilization of exponential ratio type structures have been considered, has turned out to be the most efficient among all the four proposed estimators when comparison is made with sample median estimator and T_{12} is most suitable amongst all when they are compared with the estimator Δ. Hence, when a highly positively correlated auxiliary information which is dynamic over time is used, the proposed estimators may be recommended for their practical use by survey practitioners.

UNIT - II

SEARCH OF GOOD ROTATION PATTERNS
FOR
ESTIMATION OF POPULATION MEAN
AT CURRENT OCCASION

CHAPTER – 6*

Longitudinal Analysis of Population Mean on Successive Occasions

* Following is the publication based on the work of this chapter:--

1. Priyanka, K. and Mittal, R. (2016): Longitudinal Analysis of Population Mean on Successive Occasions. International Journal of Mathematics and Statistics, Vol. 17, No. 2, 47-63.

Longitudinal Analysis of Population Mean on Successive Occasions

1. Introduction

Single time survey and their analysis do not serve the purpose in understanding the dynamics of economic and social process which are changing over time. For these situations longitudinal surveys, in which the same units are investigated on several occasions, over extensive period of time becomes important. In recent years, longitudinal surveys are now being used increasingly for longitudinal analysis and in many cases; longitudinal surveys are carefully designed to permit the derivation of sophisticated analysis of the long dynamics of social and economic processes. In this case, the same population is sampled repeatedly and the same study variable is measured at each occasion, so that development over time can be followed. For example, in many countries, labour-force surveys are conducted monthly to estimate the number of employed and the rate of unemployment. Other examples are monthly surveys in which the data on price of goods are collected to determine a consumer price index, and political opinion surveys conducted at regular intervals to measure voter preferences. These longitudinal surveys in which the sampling is done on successive occasions (over years or seasons or months) according to a specified rule, with partial replacement of units, is called successive (rotation) sampling. Successive sampling provides a strong tool for generating the reliable estimates at different occasions. In this case the survey estimates are time specific, For example, the unemployment rate is a key economic indicator that varies over time, the rate may change from one month to the next because of a change in the economy (with business laying off or recruiting new employees).

To cite one may refer the papers by Jesson (1942), Patterson (1950), Rao and Graham (1964), Gupta (1979), Das (1982) and Chaturvedi and Tripathi (1983) etc.

Sometimes, the information on auxiliary variables, which are correlated to the study variable, is available so that their population means are known. The question arises that whether it is possible to utilize the information on the auxiliary variables, which are available on both the occasions, to increase the precision for estimating the population mean on current occasion. For example in agriculture surveys, the crop infestation due to pest or disease during a week, in a particular area, may be associated with infestation and ancillary factors such as rainfall, temperature and humidity during the preceding week. Similarly, the yield of crop during a season in a farm is known to depend to a great extent on the climate factors, prevailing during the previous season. In biological populations we may be interested to estimate the kill of birds during season by hunter in locality, which is known to be related to the hunter's kill and his disposable income during the previous season. Utilizing the auxiliary information on both the occasions Sen (1971), Singh et al. (1991), Feng and Zou (1997), Biradar and Singh (2001), Singh and Singh (2001), Singh (2005) have successfully given some of the very literature in the field of sample surveys. Singh and Priyanka (2006, 2007a, 2008a), Singh and Karna (2009), Singh and Prasad (2010) have proposed a variety of estimators for estimating the population mean on current (second) occasion in two occasions successive sampling.

It has been theoretically established that, in general, the linear regression estimator is more efficient than the ratio estimator except when the regression line y on x passes through the neighbourhood of the origin; in this case the efficiencies of these estimators are almost equal. Also in many practical situations where the regression line does not pass through the neighbourhood of the origin, in such cases the ratio estimator does not perform as good as the linear regression estimator. Motivated with this argument the present work attempts to develop more efficient estimators to estimate population mean using the concept of exponential type estimators in two occasion successive sampling. Here we have also tried to amalgamate the auxiliary variate with different type of exponential type of estimators at different occasions to increase the efficiency of the proposed estimators. The amalgamation of auxiliary variable has been fruitfully justified when the proposed estimators are compared with sample mean estimator and general successive sampling estimator due to Jessen (1942). The proposed estimators are also compared mutually. The reasonability of using the new proposed estimators has been shown through empirical results and validated by means of Monte Carlo simulation based on some natural population.

2. Sample Structure and Notations

Let $U = (U_1, U_2, \ldots, U_N)$ be the finite population of N units, which has been sampled over two occasions. It is assumed that size of the population remains unchanged but values of units change over two occasions. The character under study be denoted by x (y) on the first (second) occasions respectively. It is assumed that information on an auxiliary variable z, whose population mean $(\bar{Z})$ is completely known and stable over occasions is readily available on both the occasions and positively correlated to x and y respectively. Simple random sample (without replacement) of n units is taken on the first occasion. A random subsample of $m = n\lambda$ units is retained for use on the second occasion. Now at the current occasion a simple random sample (without replacement) of u= (n-m) = nμ units is drawn afresh from the remaining (N-n) units of the population so that the sample size on the second occasion is also n. Let μ and $\lambda(\mu + \lambda = 1)$ are the fractions of fresh and matched samples respectively at the second (current) occasion. The following notations are considered for the further use:

$\bar{X}, \bar{Y}, \bar{Z}$: Population means of the variables x, y and z respectively.

$\bar{y}_u, \bar{z}_u, \bar{x}_m, \bar{y}_m, \bar{z}_m, \bar{x}_n, \bar{z}_n$: Sample means of respective variates based on the sample sizes shown in suffice.

$\rho_{yx}, \rho_{xz}, \rho_{yz}$: Correlation coefficient between the variables shown in suffices.

S_x^2, S_y^2, S_z^2 : Population mean square of variables x, y and z respectively.

3. Proposed Estimators T_{ij} (i, j =1, 2)

To estimate the population mean $\bar{Y}$ on the current (second) occasion, two sets of estimators have been proposed utilizing the concept of exponential ratio type estimators. First set of estimators $\{T_{1u}, T_{2u}\}$ is based on sample of the size u= nμ drawn afresh on the current (second) occasion and the second set of estimators $\{T_{1m}, T_{2m}\}$ is based on sample size $m = n\lambda$ common to the both occasions. The two sets of the proposed estimators are given as

$$T_{1u} = \bar{Z}\left(\frac{\bar{y}_u}{\bar{z}_u}\right) \tag{1}$$

$$T_{2u} = \bar{y}_u \exp\left(\frac{\bar{Z} - \bar{z}_u}{\bar{Z} + \bar{z}_u}\right) \tag{2}$$

$$T_{1m} = \bar{x}_n \left(\frac{\bar{y}_m}{\bar{x}_m}\right) \exp\left(\frac{\bar{Z} - \bar{z}_m}{\bar{Z} + \bar{z}_m}\right) \tag{3}$$

$$T_{2m} = \bar{x}_n^* \left(\frac{\bar{y}_m^*}{\bar{x}_m^*}\right) \tag{4}$$

where $\bar{y}_m^* = \bar{y}_m \exp\left(\frac{\bar{Z} - \bar{z}_m}{\bar{Z} + \bar{z}_m}\right)$, $\bar{x}_m^* = \bar{x}_m \exp\left(\frac{\bar{Z} - \bar{z}_m}{\bar{Z} + \bar{z}_m}\right)$ and $\bar{x}_n^* = \bar{x}_n \exp\left(\frac{\bar{Z} - \bar{z}_n}{\bar{Z} + \bar{z}_n}\right)$.

Considering the convex linear combination of the two sets of estimators $T_{iu}\,(i = 1, 2)$ and $T_{jm}\,(j = 1, 2)$, we have the final estimators of population mean $\bar{Y}$ on the current occasion as

$$T_{ij} = \varphi_{ij} T_{iu} + \left(1 - \varphi_{ij}\right) T_{jm}\ ;\ (i, j = 1, 2) \tag{5}$$

where $\varphi_{ij}\left(0 \le \varphi_{ij} \le 1;\ i, j = 1, 2\right)$ are the unknown constants to be determined so as to minimise the mean square error of the estimators T_{ij}(i, j=1, 2).

Therefore, following four estimators are possible namely

(i) $T_{11} = \varphi_{11}\ T_{1u} + \left(1 - \varphi_{11}\right)T_{1m}$, (ii) $T_{12} = \varphi_{12}\ T_{1u} + \left(1 - \varphi_{12}\right)T_{2m}$, (iii) $T_{21} = \varphi_{21}\ T_{2u} + \left(1 - \varphi_{21}\right)T_{1m}$ and (iv) $T_{22} = \varphi_{22}\ T_{2u} + \left(1 - \varphi_{22}\right)T_{2m}$.

Remark 3.1: For estimating the mean on each occasion, the estimators $T_{iu}\,(i = 1, 2)$ are suitable, which implies that more belief on T_{iu} could be shown by choosing $\varphi_{ij}\,(i, j = 1, 2)$ as 1 (or close to 1), while for estimating the change from occasion to occasion, the estimators $T_{jm}\,(j=1, 2)$ could be more useful so φ_{ij} might be chosen as 0 (or close to 0). For asserting both problems simultaneously, the suitable (optimum) choices of φ_{ij} are desired.

4. Properties of the Proposed Estimators T_{ij} (i, j =1, 2)

The properties of the proposed estimators $T_{ij}\,(i, j =1, 2)$ are derived under the following large sample approximations

$$\bar{y}_u = \bar{Y}(1 + e_0),\ \bar{y}_m = \bar{Y}(1 + e_1),\ \bar{x}_m = \bar{X}(1 + e_2),\ \bar{x}_n = \bar{X}(1 + e_3),\ \bar{z}_u = \bar{Z}(1 + e_4),$$
$$\bar{z}_m = \bar{Z}(1 + e_5) \text{ and } \bar{z}_n = \bar{Z}(1 + e_6) \text{ such that } |e_i| < 1\ \forall\ i = 0, 1, 2, 3, 4, 5 \text{ and } 6\,.$$

Remark 4.1: The expansion of $(1+x)^n$ for negative values of n is feasible only when $|x|<1$. The properties of proposed work have been studied under large sample approximations and we need to use Binomial expansion as well since error is very small. Hence to validate both we have considered magnitude of error, i.e. $|e_i|<1$.

4.1. Bias and Mean Squared Error of the Estimators T_{ij} (i, j =1, 2)

The estimators T_{iu} and $T_{jm}\,(i, j=1, 2)$ are ratio, exponential ratio, ratio to exponential ratio and chain type ratio to exponential ratio type in nature respectively. Hence they are biased for population mean $\bar{Y}$. Therefore, the final estimators $T_{ij}\,(i, j =1, 2)$ defined in equation (5) are also biased estimators of $\bar{Y}$. The bias $B(.)$ and mean squared errors $M(.)$ of the proposed estimators $T_{ij}\,(i, j =1, 2)$ are obtained up to first order of approximations and thus we have following theorems:

Theorem 4.1.1. Bias of the estimators $T_{ij}\,(i, j =1, 2)$ to the first order of approximations are obtained as

$$B(T_{ij}) = \varphi_{ij}\ B(T_{iu}) + (1 - \varphi_{ij})\ B(T_{jm});\ (i, j=1, 2), \tag{6}$$

where $$B(T_{1u}) = \frac{1}{u}\bar{Y}\left(\frac{C_{002}}{\bar{Z}^2} - \frac{C_{011}}{\bar{Y}\,\bar{Z}}\right), \tag{7}$$

$$B(T_{2u}) = \frac{1}{u}\bar{Y}\left(\frac{3}{8}\frac{C_{002}}{\bar{Z}^2} - \frac{1}{2}\frac{C_{011}}{\bar{Y}\,\bar{Z}}\right), \tag{8}$$

$$B(T_{1m}) = \bar{Y}\left(\frac{1}{m}\left(\frac{C_{200}}{\bar{X}^2} + \frac{3}{8}\frac{C_{002}}{\bar{Z}^2} - \frac{C_{110}}{\bar{X}\bar{Y}} - \frac{1}{2}\frac{C_{011}}{\bar{Y}\bar{Z}} + \frac{1}{2}\frac{C_{101}}{\bar{X}\bar{Z}}\right) + \frac{1}{n}\left(\frac{C_{110}}{\bar{X}\bar{Y}} - \frac{C_{200}}{\bar{X}^2} - \frac{1}{2}\frac{C_{101}}{\bar{X}\bar{Z}}\right)\right), \tag{9}$$

and $$B(T_{2m}) = \bar{Y}\left(\frac{1}{m}\left(\frac{C_{200}}{\bar{X}^2} - \frac{C_{110}}{\bar{X}\bar{Y}}\right) + \frac{1}{n}\left(\frac{3}{8}\frac{C_{002}}{\bar{Z}^2} + \frac{C_{110}}{\bar{X}\bar{Y}} - \frac{1}{2}\frac{C_{011}}{\bar{Y}\bar{Z}} - \frac{C_{200}}{\bar{X}^2}\right)\right) \tag{10}$$

where $C_{rst} = E\left[(x_i - \bar{X})^r (y_i - \bar{Y})^s (z_i - \bar{Z})^t\right];\ (r, s, t) \geq 0$.

Proof: The bias of the estimators $T_{ij}(i, j=1, 2)$ are given by

$$B(T_{ij}) = E\left[T_{ij} - \bar{Y}\right] = \varphi_{ij}\ B(T_{iu}) + (1 - \varphi_{ij})B(T_{jm})$$

where $B(T_{iu}) = E\left[T_{iu} - \bar{Y}\right]$ and $B(T_{jm}) = E\left[T_{jm} - \bar{Y}\right]$

Using large sample approximations and retaining terms up-to the first order of approximations, the expression for $B(T_{iu})$ and $B(T_{jm})$ are obtained as in equations (7) - (10) and hence the expression for bias of the estimators $T_{ij}(i, j=1, 2)$ are obtained as in equation (6).

Theorem 4.1.2. Mean squared errors of the estimators $T_{ij}(i, j=1, 2)$ to the first order of approximations are obtained as

$$M(T_{ij}) = \varphi_{ij}^2\ M(T_{iu}) + (1 - \varphi_{ij})^2 M(T_{jm}) + 2\,\varphi_{ij}(1 - \varphi_{ij})\mathrm{Cov}(T_{iu}, T_{jm});\ (i, j=1, 2) \tag{11}$$

where $M(T_{1u}) = \frac{1}{u} A_1 S_y^2$ (12)

$$M(T_{2u}) = \frac{1}{u} A_2 S_y^2 \quad (13)$$

$$M(T_{1m}) = \left(\frac{1}{m} A_3 + \frac{1}{n} A_4\right) S_y^2 \quad (14)$$

$$M(T_{2m}) = \left(\frac{1}{m} A_5 + \frac{1}{n} A_6\right) S_y^2 \quad (15)$$

$A_1 = 2(1 - \rho_{yz})$, $A_2 = \frac{5}{4} - \rho_{yz}$, $A_3 = \frac{9}{4} - 2\rho_{yx} - \rho_{yz} + \rho_{xz}$, $A_4 = 2\rho_{yx} - \rho_{xz} - 1$,

$A_5 = 2(1 - \rho_{yx})$ and $A_6 = 2\rho_{yx} - \rho_{yz} - \frac{3}{4}$.

Proof: The mean squared errors of the estimators T_{ij} are given by

$$M(T_{ij}) = E\left[T_{ij} - \bar{Y}\right]^2 = E\left[\varphi_{ij}\left(T_{iu} - \bar{Y}\right) + \left(1 - \varphi_{ij}\right)\left(T_{jm} - \bar{Y}\right)\right]^2$$

$$= \varphi_{ij}^{\;2}\, M(T_{iu}) + \left(1 - \varphi_{ij}\right)^2 M\left[T_{jm}\right] + 2\,\varphi_{ij}\left(1 - \varphi_{ij}\right) Cov\left(T_{iu}, T_{jm}\right)$$

where $M(T_{iu}) = E\left[T_{iu} - \bar{Y}\right]^2$ and $M\left[T_{jm}\right] = E\left[T_{jm} - \bar{Y}\right]^2$; (i, j=1, 2)

Since x and y denote the same study character over two occasions and z being completely known auxiliary variate positively correlated to x and y, therefore, looking at the stability nature (see Reddy (1978)) of the coefficient of variation and following Cochran (1977) and Feng and Zou (1997), the coefficient of variation of x, y and z are considered to be approximately same which is given by $C_y = \frac{S_y}{\bar{Y}}$.

The estimators T_{iu} and T_{jm} are based on two independent samples of sizes u and m respectively, hence $Cov\left(T_{iu}, T_{jm}\right) = 0$; (i, j = 1, 2). Considering population size is sufficiently large (i.e. $N \to \infty$), therefore finite population corrections are ignored and using large sample approximations and retaining terms upto the first order of approximations, the expression for $M(T_{iu})$ and $M(T_{jm})$ are obtained as given in equations (12) - (15) and hence the expressions for mean squared errors of estimators T_{ij} (i, j =1, 2) are obtained as in equation (11).

5. Minimum Mean Squared Errors of the Proposed Estimators T_{ij} (i, j =1, 2)

Since the mean squared errors of the estimators $T_{ij}\left(i, j = 1, 2\right)$ given in equation (11) are the functions of unknown constants $\varphi_{ij}\left(i, j = 1, 2\right)$, therefore, they are minimized with respect to φ_{ij} and subsequently the optimum values of φ_{ij} are obtained as

$$\varphi_{ij_{opt.}} = \frac{M\left(T_{jm}\right)}{M\left(T_{iu}\right) + M\left(T_{jm}\right)}; \ (i, j = 1, 2) \tag{16}$$

Now substituting the values of $\varphi_{ij_{opt.}}$ in equation (11), we obtain the optimum mean squared errors of the estimators $T_{ij}\left(i, j = 1, 2\right)$ as

$$M\left(T_{ij}\right)_{opt.} = \frac{M\left(T_{iu}\right) . M\left(T_{jm}\right)}{M\left(T_{iu}\right) + M\left(T_{jm}\right)}; \ \left(i, j = 1, 2\right) \tag{17}$$

Further, substituting the values of the mean squared error of the estimators defined in equations (12) to (15) in equation (16) and (17), the simplified values of $\varphi_{ij_{opt.}}$ and $M\left(T_{ij}\right)_{opt.}$ are obtained as

$$\varphi_{11_{opt.}} = \frac{\mu_{11}\left[\mu_{11}\ A_4 - \left(A_3 + A_4\right)\right]}{\left[\mu_{11}^2\ A_4 - \mu_{11}\left(A_3 + A_4 - A_1\right) - A_1\right]} \tag{18}$$

$$\varphi_{12_{opt.}} = \frac{\mu_{12}\left[\mu_{12}\ A_6 - \left(A_5 + A_6\right)\right]}{\left[\mu_{12}^2\ A_6 - \mu_{12}\left(A_5 + A_6 - A_1\right) - A_1\right]} \tag{19}$$

$$\varphi_{21_{opt.}} = \frac{\mu_{21}\left[\mu_{21}\ A_4 - \left(A_3 + A_4\right)\right]}{\left[\mu_{21}^2\ A_4 - \mu_{21}\left(A_3 + A_4 - A_2\right) - A_2\right]} \tag{20}$$

$$\varphi_{22_{opt.}} = \frac{\mu_{22}\left[\mu_{22}\ A_6 - \left(A_5 + A_6\right)\right]}{\left[\mu_{22}^2\ A_6 - \mu_{22}\left(A_5 + A_6 - A_2\right) - A_2\right]} \tag{21}$$

$$M(T_{11})_{opt.} = \frac{1}{n}\frac{[\mu_{11}\ B_1\ - B_2]\ S_y^2}{[\mu_{11}^2\ A_4\ - \mu_{11}\ B_3\ - A_1]} \tag{22}$$

$$M(T_{12})_{opt.} = \frac{1}{n}\frac{[\mu_{12}\ B_4\ - B_5]S_y^2}{[\mu_{12}^2\ A_6\ - \mu_{12}\ B_6\ - A_1]} \tag{23}$$

$$M(T_{21})_{opt.} = \frac{1}{n}\frac{[\mu_{21}\ B_7\ - B_8]S_y^2}{[\mu_{21}^2\ A_4\ - \mu_{21}\ B_9\ - A_2]} \tag{24}$$

$$M(T_{22})_{opt.} = \frac{1}{n}\frac{[\mu_{22}\ B_{10}\ - B_{11}]S_y^2}{[\mu_{22}^2\ A_6\ - \mu_{22}\ B_{12}\ - A_2]} \tag{25}$$

where

$B_1 = A_1A_4$, $B_2 = A_1A_3 + A_1A_4$, $B_3 = A_3 + A_4 - A_1$, $B_4 = A_1A_6$, $B_5 = A_1A_5 + A_1A_6$, $B_6 = A_5 + A_6 - A_1$, $B_7 = A_2A_4$, $B_8 = A_2A_3 + A_2A_4$, $B_9 = A_3 + A_4 - A_2$, $B_{10} = A_2A_6$ $B_{11} = A_2A_5 + A_2A_6$, $B_{12} = A_5 + A_6 - A_2$ and $\mu_{ij}(i, j = 1, 2)$ are the fractions of the sample drawn afresh at the current(second) occasion.

Remark 5.1: $M(T_{ij})_{opt.}$ derived in equation (22) - (25) are the functions of $\mu_{ij}(i, j = 1, 2)$. To estimate the population mean on each occasion the better choices of $\mu_{ij}(i, j = 1, 2)$ are 1(case of no matching); however, to estimate the change in mean from one occasion to other, $\mu_{ij}(i, j = 1, 2)$ should be 0(case of complete matching). But intuition suggests that the optimum choices of $\mu_{ij}(i, j = 1, 2)$ are desired to devise the amicable strategy for both the problems simultaneously.

6. Optimum Replacement Strategies for the Estimators T_{ij} (i, j =1, 2)

The key design parameter affecting the estimates of change is the overlap between successive samples. Maintaining high overlap between repeats of a survey is operationally convenient, since many sampled units have been located and have some experience in the survey. Hence to decide about the optimum value of $\mu_{ij}(i, j = 1, 2)$ (fractions of samples to be drawn afresh on current occasion) so that $\bar{Y}$ may be estimated with maximum precision and minimum cost, we minimize the mean square errors $M(T_{ij})_{opt.}(i, j = 1, 2)$ in equation (22) to (25) with respect to $\mu_{ij}(i, j = 1, 2)$ respectively.

The optimum value of $\mu_{ij}(i, j = 1, 2)$ so obtained is one of the two roots given by

$$\hat{\mu}_{11} = \frac{C_2 \pm \sqrt{C_2^2 - C_1 C_3}}{C_1} \tag{26}$$

$$\hat{\mu}_{12} = \frac{C_5 \pm \sqrt{C_5^2 - C_4 C_6}}{C_4} \tag{27}$$

$$\hat{\mu}_{21} = \frac{C_8 \pm \sqrt{C_8^2 - C_7 C_9}}{C_7} \tag{28}$$

$$\hat{\mu}_{22} = \frac{C_{11} \pm \sqrt{C_{11}^2 - C_{10} C_{12}}}{C_{10}} \tag{29}$$

where

$C_1 = A_4B_1$, $C_2 = A_4B_2$, $C_3 = A_1B_1 + B_2B_3$, $C_4 = A_6B_4$, $C_5 = A_6B_5$, $C_6 = A_1B_4 + B_5B_6$

$C_7 = A_4B_7$, $C_8 = A_4B_8$, $C_9 = A_2B_7 + B_8B_9$, $C_{10} = A_6B_{10}$, $C_{11} = A_6B_{11}$ and $C_{12} = A_2B_{10} + B_{11}B_{12}$.

The real values of $\hat{\mu}_{ij}(i, j = 1, 2)$ exist, iff $C_2^2 - C_1 C_3 \geq 0$, $C_5^2 - C_4 C_6 \geq 0$, $C_8^2 - C_7 C_9 \geq 0$, and $C_{11}^2 - C_{10} C_{12} \geq 0$ respectively. For any situation, which satisfies these conditions, two real values of $\hat{\mu}_{ij}(i, j = 1, 2)$ may be possible , hence to choose a value of $\hat{\mu}_{ij}(i, j = 1, 2)$, it should be taken care of that $0 \leq \hat{\mu}_{ij} \leq 1$, all other values of $\hat{\mu}_{ij}(i, j = 1, 2)$ are inadmissible. If both the real values of $\hat{\mu}_{ij}(i, j = 1, 2)$ are admissible, the lowest one will be the best choice as it reduces the total cost of the survey. Substituting the admissible value of $\hat{\mu}_{ij}$ say $\hat{\mu}_{ij}^{(0)}(i, j = 1, 2)$ from equation (26) to (29) in equation

(22) to (25) respectively , we get the optimum values of the mean square errors of the estimators $T_{ij}(i, j = 1, 2)$ with respect to φ_{ij} as well as $\mu_{ij}(i, j = 1, 2)$ which are given as

$$M(T_{11})^*_{opt.} = \frac{\left[\mu_{11}^{(0)} B_1 - B_2\right] S_y^2}{n\left[\mu_{11}^{(0)2} A_4 - \mu_{11}^{(0)} B_3 - A_1\right]} \tag{30}$$

$$M(T_{12})^*_{opt.} = \frac{\left[\mu_{12}^{(0)} B_4 - B_5\right] S_y^2}{n\left[\mu_{12}^{(0)2} A_6 - \mu_{12}^{(0)} B_6 - A_1\right]} \tag{31}$$

$$M(T_{21})^*_{opt.} = \frac{\left[\mu_{21}^{(0)} B_7 - B_8\right] S_y^2}{n\left[\mu_{21}^{(0)2} A_4 - \mu_{21}^{(0)} B_9 - A_2\right]} \tag{32}$$

$$M(T_{22})^*_{opt.} = \frac{\left[\mu_{22}^{(0)} B_{10} - B_{11}\right] S_y^2}{n\left[\mu_{22}^{(0)2} A_6 - \mu_{22}^{(0)} B_{12} - A_2\right]} \tag{33}$$

7. Efficiency Comparison

To evaluate the performance of the proposed estimators, the estimators $T_{ij}(i, j = 1, 2)$ at optimum conditions are compared with (i) the sample mean estimator $\bar{y}_n$, when there is no matching from previous occasion and (ii) the general successive sampling estimator $\hat{\hat{Y}}$ due to Jessen (1942)

$$\hat{\hat{Y}} = \psi\, \bar{y}_u + (1 - \psi)\bar{y}_m{}', \tag{34}$$

where $\bar{y}_m{}' = \bar{y}_m + \beta_{yx}(\bar{x}_n - \bar{x}_m)$, β_{yx} is the population regression coefficient of y on x and ψ is an unknown constant to be determined so as to minimise the mean squared error of the estimator $\hat{\hat{Y}}$. Here both $\bar{y}_n$ and $\hat{\hat{Y}}$ are unbiased for population mean, so variance of the estimator $\bar{y}_n$ and $\hat{\hat{Y}}$ at optimum conditions are given as

$$V(\bar{y}_n) = \frac{1}{n} S_y^2, \tag{35}$$

$$V\left(\hat{\hat{Y}}\right)^{*}_{opt.} = \left(\frac{1}{2}\left(1 + \sqrt{1 - \rho^2_{yx}}\right)\right)\frac{S^2_y}{n}, \tag{36}$$

and the fraction of sample to be drawn afresh provided by the estimators due Jessen (1942) is given by

$$\mu_J = \frac{1}{1 + \sqrt{1 - \rho^2_{yx}}} \tag{37}$$

The percent relative efficiencies $E_{ij}(M)$ and $E_{ij}(JS)$ of the estimator T_{ij} (under optimum conditions) with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$ are respectively given by

$$E_{ij}(M) = \frac{V(\bar{y}_n)}{M\left(T_{ij}\right)^{*}_{opt.}} \times 100 \quad \text{and} \quad E_{ij}(JS) = \frac{V\left(\hat{\hat{Y}}\right)^{*}_{opt.}}{M\left(T_{ij}\right)^{*}_{opt.}} \times 100 \; (i, j=1, 2). \tag{38}$$

8. Empirical Illustrations and Monte Carlo Simulation

Empirical validation can be carried out by Monte Carlo Simulation. Real life situation of completely known finite population has been considered.

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

The population comprise of N = 51 states of United States. Let y_i be the total energy consumption during 2008 in the i^{th} state of U. S., x_i be the total energy consumption during 2003 in the i^{th} state of U. S. and z_i denote the total energy consumption during 2001 in the i^{th} state of U. S.

For the considered population, the values of $\mu^{(0)}_{ij}$ defined in equation (26) to (29) and the percent relative efficiencies $E_{ij}(M)$ and $E_{ij}(JS)$ defined in equation (38) of $T_{ij}\left(i, j=1, 2\right)$ with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$ have been computed and are presented in Table 1. The optimum bias of the estimators $T_{ij}\left(i, j=1, 2\right)$ has been computed utilizing $\varphi_{ij}\left(i, j=1, 2\right)$ from equation (18) to (21) and the $\mu^{(0)}_{ij}\left(i, j=1, 2\right)$ from equation (26) to (29) and are shown in the Table 2.

To validate the above empirical results, Monte Carlo simulation has also been performed for the considered population.

8.1 Simulation Algorithm

(i) Choose 5000 samples of size n=20 using simple random sampling without replacement on first occasion for both the study and auxiliary variable.

(ii) Calculate sample mean $\bar{x}_{n\,|\,k}$ and $\bar{z}_{n\,|\,k}$ for k =1, 2, - - -, 5000.

(iii) Retain m=17 units out of each n=20 sample units of the study and auxiliary variables at the first occasion.

(iv) Calculate sample mean $\bar{x}_{m\,|\,k}$ and $\bar{z}_{m\,|\,k}$ for k= 1, 2, - - -, 5000.

(v) Select u=3 units using simple random sampling without replacement from N-n=31 units of the population for study and auxiliary variables at second (current) occasion.

(vi) Calculate sample mean $\bar{y}_{u\,|\,k}$, $\bar{y}_{m\,|\,k}$ and $\bar{z}_{u\,|\,k}$ for k = 1, 2, - - -, 5000.

(vii) Iterate the parameter $\varphi_{i\,j}\left(i, j=1, 2\right)$ from 0.1 to 0.9 with a step of 0.1.

(viii) Iterate ψ from 0.1 to 0.9 with a step of 0.2 within (vii).

(ix) Calculate the percent relative efficiencies of the proposed estimators $T_{i\,j}\left(i, j=1, 2\right)$ with respect to estimators respect to $\bar{y}_n$ and $\hat{\bar{Y}}$ as

$$E_{i\,j}\left(1\right) = \frac{\sum_{k=1}^{5000}\left[\bar{y}_{n|k} - \bar{Y}\right]^2}{\sum_{k=1}^{5000}\left[T_{i\,j\,|\,k} - \bar{Y}\right]^2} \times 100 \text{ and } E_{i\,j}\left(2\right) = \frac{\sum_{k=1}^{5000}\left[\hat{\bar{Y}}_k - \bar{Y}\right]^2}{\sum_{k=1}^{5000}\left[T_{i\,j\,|\,k} - \bar{Y}\right]^2} \times 100, \quad k=1, 2, ..., 5000.$$

Table 1: Empirical Comparison of the proposed estimators T_{ij} with respect to the estimators $\bar{y}_n$ and $\hat{\bar{Y}}$.

Optimum Value of $\varphi_{ij}(i, j=1, 2)$		Optimum Value of $\mu_{ij}^{(0)}(i, j=1, 2)$		Percent Relative Efficiencies with respect to $\bar{y}_n$		Percent Relative Efficiencies with respect to $\hat{\bar{Y}}$	
		μ_J	0.5362				
φ_{11}	0.5487	$\mu_{11}^{(0)}$	0.4396	$E_{11}(M)$	122.22	$E_{11}(JS)$	130.59
φ_{12}	0.5389	$\mu_{12}^{(0)}$	0.4070	$E_{12}(M)$	119.34	$E_{12}(JS)$	127.51
φ_{21}	0.50	$\mu_{21}^{(0)}$	0.3772	$E_{21}(M)$	115.20	$E_{21}(JS)$	123.01
φ_{22}	0.50	$\mu_{22}^{(0)}$	**0.3554**	$E_{22}(M)$	112.49	$E_{22}(JS)$	120.12

Table 2: Optimum Bias of the proposed estimators T_{ij} for the choices of sample size n.

Optimum Bias	n=15	n=20	n=25
$B(T_{11})$	41.87	31.40	25.12
$B(T_{12})$	38.36	28.77	23.01
$B(T_{21})$	27.69	20.77	16.61
$B(T_{22})$	**24.56**	**18.42**	**14.73**

Table 3: Monte Carlo Simulation results when the proposed estimators T_{ij} are compared to $\bar{y}_n$.

φ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$E_{11}(1)$	462.33	405.02	375.06	368.48	364.25	384.56	431.24	488.90	556.41
$E_{12}(1)$	477.58	415.00	384.20	375.98	371.26	390.83	436.56	493.61	560.69
$E_{21}(1)$	449.59	379.28	336.78	302.91	274.72	253.47	236.82	227.25	222.72
$E_{22}(1)$	**564.35**	**488.88**	**446.47**	**410.75**	**381.46**	**359.95**	**342.48**	**332.37**	**327.29**

9. Mutual Comparison of the Proposed Estimators T_{ij} (i, j = 1, 2)

The performances of the proposed estimators T_{ij} $(i, j = 1, 2)$ have been elaborated empirically as well as through simulation studies in above section 8 and the results obtained are presented in Table 1 to Table 4. In this section the mutual comparison of the four proposed estimators has been elaborated through different graphs given in Figure 9.1 to Figure 9.3.

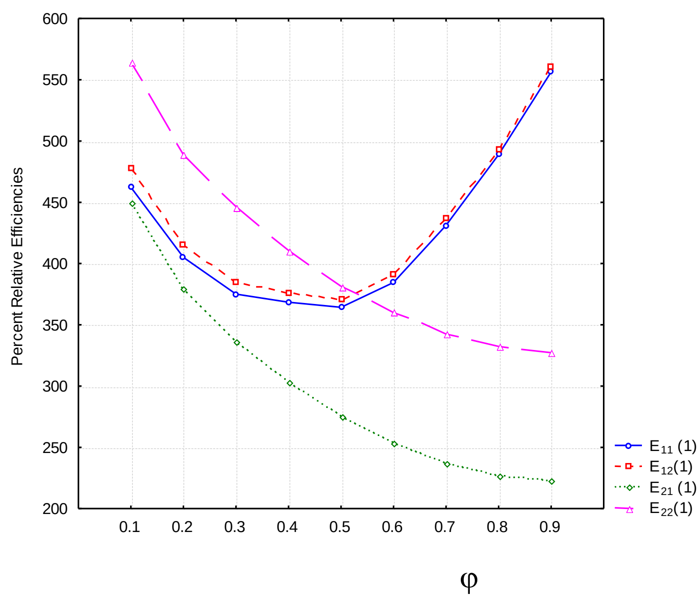


Figure 9.1: Mutual Comparison of Proposed Estimator T_{ij} $(i, j = 1, 2)$ when compared with the estimator $\bar{y}_n$.

Table 4: Monte Carlo Simulation results when the proposed estimators T_{ij} are compared to $\hat{\bar{Y}}$.

φ \ ψ		**0.1**	**0.3**	**0.5**	**0.7**	**0.9**
0.1	$E_{11}(2)$	446.13	574.50	601.84	274.17	386.37
	$E_{12}(2)$	529.93	672.20	624.14	300.44	356.37
	$E_{21}(2)$	448.55	576.43	605.09	274.80	388.70
	$E_{22}(2)$	534.19	677.14	628.52	302.17	358.75
0.2	$E_{11}(2)$	390.02	529.49	516.24	247.17	345.70
	$E_{12}(2)$	463.09	615.95	537.18	269.91	320.53
	$E_{21}(2)$	393.90	531.26	519.80	248.12	348.97
	$E_{22}(2)$	469.50	621.60	542.37	272.32	323.94
0.3	$E_{11}(2)$	351.29	469.57	463.87	221.79	303.70
	$E_{12}(2)$	415.52	545.64	481.55	241.62	282.05
	$E_{21}(2)$	356.42	471.06	468.57	222.57	307.64
	$E_{22}(2)$	423.69	552.03	488.09	244.14	286.08
0.4	$E_{11}(2)$	317.06	421.74	414.95	197.24	273.65
	$E_{12}(2)$	374.60	489.30	429.51	214.48	254.47
	$E_{21}(2)$	322.46	423.22	420.38	197.89	278.20
	$E_{22}(2)$	383.45	496.64	437.01	217.21	259.11
0.5	$E_{11}(2)$	285.50	382.09	373.44	178.19	246.45
	$E_{12}(2)$	336.00	441.38	385.90	193.55	229.61
	$E_{21}(2)$	290.94	383.39	379.11	178.40	250.93
	$E_{22}(2)$	345.27	449.03	393.81	195.99	234.18
0.6	$E_{11}(2)$	259.66	347.76	338.52	163.18	222.14
	$E_{12}(2)$	304.25	400.62	349.09	176.60	207.15
	$E_{21}(2)$	265.00	348.45	343.80	163.10	226.48
	$E_{22}(2)$	313.65	407.86	356.74	178.87	211.58
0.7	$E_{11}(2)$	236.64	318.91	312.24	150.54	200.79
	$E_{12}(2)$	275.99	365.42	321.32	162.35	187.41
	$E_{21}(2)$	241.43	319.25	316.18	150.11	205.02
	$E_{22}(2)$	284.80	372.36	327.66	164.32	191.75
0.8	$E_{11}(2)$	217.64	298.02	290.02	141.70	183.07
	$E_{12}(2)$	252.38	339.25	297.96	152.22	171.15
	$E_{21}(2)$	221.83	297.48	292.59	140.44	186.98
	$E_{22}(2)$	260.48	345.14	302.82	153.29	175.15
0.9	$E_{11}(2)$	204.19	283.51	273.36	135.44	168.58
	$E_{12}(2)$	235.14	320.21	280.43	144.79	157.93
	$E_{21}(2)$	207.48	281.09	274.59	133.58	171.96
	$E_{22}(2)$	242.09	323.79	290.79	145.11	161.40

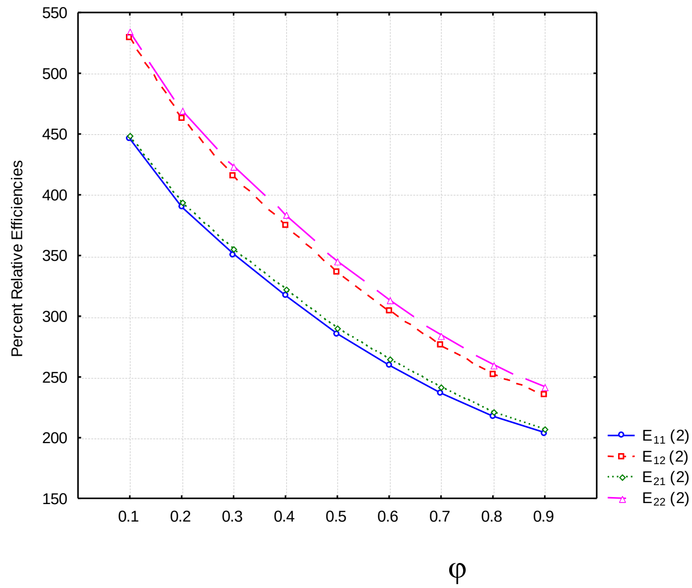


Figure 9.2: Mutual Comparison of Proposed Estimator T_{ij} (i, j =1, 2) **when compared with the estimator** $\hat{\hat{Y}}$ **for** ψ=0.1 .

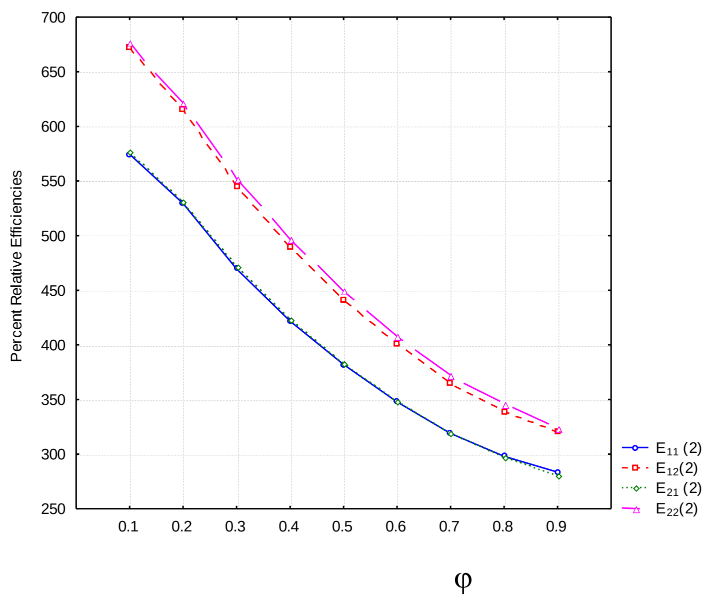


Figure 9.3: Mutual Comparison of Proposed Estimator T_{ij} (i, j =1, 2) **when compared with the estimator** $\hat{\hat{Y}}$ for ψ=0.3 .

10. Interpretation of results

The performance of an estimator in successive sampling is generally judged on the basis of percent relative efficiency and in terms of optimum values of fraction of fresh sample drawn on current(second) occasion which in turns is directly associated to the cost of survey. Here the following interpretation can be drawn from Tables 1 - 4 and Figure 9.1 - 9.3,

(1) From Table-1, it is observed that

(a) Optimum values $\mu_{11}^{(0)}, \mu_{12}^{(0)}, \mu_{21}^{(0)}$ and $\mu_{22}^{(0)}$ for the estimators $T_{11}, T_{12}, T_{21}, T_{22}$ respectively exist for the considered Population and $\mu_{22}^{(0)} < \mu_{21}^{(0)} < \mu_{12}^{(0)} < \mu_{11}^{(0)} < \mu_J$, which justifies the applicability of the proposed estimators $T_{11}, T_{12}, T_{21}, T_{22}$ at optimum conditions. The value for $\mu_{22}^{(0)}$ is lowest amongst all other which leads the results that the estimator T_{22} is most favourable in terms of cost amongst the other proposed estimators. However, at optimum conditions for the considered population the estimator T_{11} performs better in terms of efficiencies with respect to sample mean estimator as well as with estimator $\hat{\bar{Y}}$ due to Jessen (1942) in terms of precision only.

(b) Appreciable gain is observed in terms of precision indicating the proposed estimators $T_{ij}(i, j=1, 2)$ (at their respective optimal conditions) are preferable over the estimators $\bar{y}_n$ and $\hat{\bar{Y}}$ (at optimal conditions). This result justifies the use of additional auxiliary information at both occasions which is stable over time in two occasion successive sampling.

(2) In Table-2, we see that the optimum bias of the estimators $T_{ij}(i, j=1, 2)$ reduces as the sample size of sample is increased. The estimator T_{22} is least biased for population mean amongst all the four proposed estimators.

(3) From the simulation results presented in Table-3, where $T_{ij}(i, j=1, 2)$ are compared to the sample mean estimator $\bar{y}_n$, it can be seen that

(a) The values for $E_{11}(1)$ and $E_{12}(1)$ increases as the value of φ increases, this is in accordance with Sukhatme et al. (1984) results.

(b) The value for $E_{22}(1)$ is greatest amongst all when the estimators $T_{ij}(i, j=1, 2)$ are compared to sample mean estimator $\bar{y}_n$, This indicates that the estimator T_{22} outperforms amongst the four considered estimators.

(4) From simulation results presented in Table-4, where the estimators $T_{ij}(i, j=1, 2)$ are compared with the estimator $\hat{\bar{Y}}$, following results can be drawn

(a) The value for $E_{ij}(2);(i, j=1, 2)$ decreases as φ increases for all choices of ψ which is in accordance with the concept of successive sampling.

(b) The value of $T_{22}(2)$ is maximum amongst $E_{ij}(2);(i, j=1, 2)$, this indicates that the estimator $T_{22}(2)$ dominates the other estimators proposed.

11. Conclusion

From the preceding interpretations, it may be concluded that the use of exponential ratio type estimators for the estimation of population mean at current occasion in two occasion successive sampling is highly appreciable as vindicated through empirical and simulation results. The use of positively correlated auxiliary information which is stable over time is highly rewarding in terms of precision and reducing the cost of survey. From the mutual comparison of the estimators it is observed that all the four proposed estimators prove to be working more efficiently than sample mean and general successive sampling estimator due to Jessen (1942). Although the estimator T_{11} is most efficient over the estimators $\bar{y}_n$ and $\hat{\bar{Y}}$ in terms of precision.

The performance of an estimator in successive sampling is generally judged on the basis of percent relative efficiency and cost of survey involved, in terms of optimum values of fraction of fresh sample drawn on current(second) occasion since same is directly associated to the cost of survey. Empirically the estimator T_{22} is not best in terms of efficiency for this population but it provides the minimum fraction of sample to be drawn afresh on current occasion, now it is not as good as the others in terms of efficiency but just for sake of little more gain in efficiency, the cost of survey cannot be compromised, so for being more precise we have carried out the simulation which suggest that T_{22} is best irrespective of optimum value of μ. We see that T_{22} least biased. Hence, we conclude that the estimator T_{22} performs best out of the four proposed estimators. This leads to the result that more inclusion of exponential type estimators provides more efficient results as it utilizes the information on relationship between auxiliary and study variable most efficiently as compared to others. Hence the proposed estimators are justified and are recommended for their practical use by survey practitioners.

CHAPTER – 7*

Multivariate Rotation Design for Population Mean in Sampling on Successive Occasions

* Following is the publication based on the work of this chapter:--

1. Priyanka, K. Mittal, R. Kim, J. M. (2015): Multivariate Rotation Design for Population Mean
in Sampling on Successive Occasions. Communications for Statistical Applications and
Methods, Vol. 22, No. 5, 445–462.

Multivariate Rotation Design for Population Mean in Sampling on Successive Occasions

1. Introduction

Longitudinal surveys are correlational research studies which involve repeated observations of the same variables over long periods of time. Longitudinal studies are often used in psychology to study developmental trends across the life span, and in sociology to study life events throughout lifetimes or generations. The reason for this is that longitudinal studies track the same people, and therefore the differences observed in those people are less likely to be the result of cultural differences across generations. Because of this benefit, longitudinal studies make observing changes more accurate, and they are applied in various other fields. In medicine, the design is used to uncover predictors of certain diseases. In advertising, the design is used to identify the changes that advertising has produced in the attitudes and behaviours of those within the target audience who have seen the advertising campaign.

Many researchers have tried to take advantage of the longitudinal surveys, to cite one may refer the literature by Jessen (1942), Patterson (1950), Rao and Graham (1964), Gupta (1979), Das (1982) and Chaturvedi and Tripathi (1983) etc.

Sometimes we get to sense that different variables are related to the study character, which may be helpful in estimating the study character. For example many countries keep track of the population through total population register and it is often used as a sampling frame of individuals or households. The register contains a number of variables, some quantitative and some categorical, that may serve as auxiliary information for identifying the human development index. So age and the taxable income of individual are the quantitative auxiliary variables while sex of the individual, marital status and residential specification may be considered as categorical auxiliaries.

Hence utilizing the auxiliary information on both the occasions Sen (1972, 1973), Singh et al. (1991), Feng and Zou (1997), Biradar and Singh (2001), Singh and Singh (2001), Singh (2005) have successfully added some literature in the field of successive sampling. Singh and Priyanka (2006, 2007a, 2008a), Singh and Karna (2009), Singh and Prasad (2010) have proposed a variety of estimators for estimating the population mean on current (second) occasion in two occasions successive sampling.

It has been established that, in general, the linear regression estimator is more efficient than the ratio estimator except when the regression line y on x passes through the neighbourhood of the origin; in this case the efficiencies of these estimators are almost equal. Also there are many practical situations when the regression line does not pass through the neighbourhood of the origin, in such cases the ratio estimator does not perform as good as the linear regression estimator.

Motivated with this argument, in the proposed work an attempt has been made to utilise multi-auxiliary information which are available on both the occasions and are stable over time. The multi-auxiliary information are blended with exponential type structures and a multivariate exponential ratio type estimator has been proposed for estimation of population mean at current occasion in two occasion rotation sampling. The properties of the proposed estimator are derived upto the first order of approximation and the optimum replacement strategies are discussed. These properties have been corroborated empirically. The proposed multivariate estimator has been compared with the recent literature in rotation sampling due to Singh (2005) and Singh and Priyanka (2008a). A simulation study has been carried out which determines the working efficiency of the proposed estimator. It has been observed that the proposed exponential type structure works well even if the auxiliary variables have low correlation with the study variable.

2. Sample Structure and Notations

Let $U = (U_1, U_2, \ldots, U_N)$ be the finite population of N units, which has been sampled over two occasions. We have assumed that the size of the population remains unchanged but values of units change over two occasions. The characters under study have been denoted by x and y on the first and second occasions respectively. It has been assumed that information on p additional auxiliary variables, $z_1, z_2, \ldots, z_p$ whose population means are known, correlated to x and y, stable over the occasions and are readily available on both the occasions. Simple random sample (without replacement) of n units is taken on the first occasion. A random subsample of $m = n\lambda$ units is retained for use on the second (current) occasion. Now at the current occasion a simple random sample (without replacement) of u= (n-m) = nμ units is drawn afresh from the remaining (N-n) units of the population so that the sample size on the second occasion is also n. Let μ and $\lambda(\mu + \lambda = 1;\ 0 \le \mu, \lambda \le 1)$ are the fractions of fresh and matched samples respectively at the second (current) occasion.

3. Formulation of the Proposed Estimator $T_{|p}$

To estimate the population mean $\bar{Y}$ on the current (second) occasion, utilizing p-additional auxiliary information which are stable over time and are readily available on both the occasions, a multivariate weighted estimator T_u based on sample of the size u= nμ drawn afresh on the current (second) occasion is proposed as

$$T_u = \mathbf{W}_u' \, \mathbf{T}_{exp}(u) \tag{1}$$

where $\mathbf{W_u}$ is a column vector of p-weights given by $\mathbf{W_u} = [w_{u_1} \quad w_{u_2} \cdot \quad \cdot \quad \cdot \quad w_{u_p}]'$

and $\mathbf{T}_{exp}(u) = \begin{bmatrix} T(1,u) \\ T(2,u) \\ \vdots \\ T(p,u) \end{bmatrix}$, where $T(i, u) = \bar{y}_u \exp\left(\frac{\bar{Z}_i - \bar{z}_i(u)}{\bar{Z}_i + \bar{z}_i(u)}\right)$ for i = 1, 2, 3, …, p

such that $\mathbf{1}'\mathbf{W_u} = 1$, where $\mathbf{1}$ is a column vector of order p.

The second estimator T_m is also proposed as weighted multivariate chain type ratio to exponential ratio estimator based on sample size $m = n\lambda$ common to the both occasions and is given by

$$T_m = \mathbf{W}_m' \, \mathbf{T}_{exp}(m, n) \tag{2}$$

where $\mathbf{W_m}$ is a column vector of p-weights as $\mathbf{W_m} = [w_{m_1} \quad w_{m_2} \cdot \quad \cdot \quad \cdot \quad w_{m_p}]'$

and $\mathbf{T}_{exp}(m,n) = \begin{bmatrix} T(1,m,n) \\ T(2,m,n) \\ \vdots \\ T(p,m,n) \end{bmatrix}$, where $T(i, m, n) = \left(\frac{\bar{y}^*(i, m)}{\bar{x}^*(i, m)}\right)\bar{x}^*(i, n)$

where $\bar{y}^*(i, m) = \bar{y}_m \exp\left(\frac{\bar{Z}_i - \bar{z}_i(m)}{\bar{Z}_i + \bar{z}_i(m)}\right)$, $\bar{x}^*(i, m) = \bar{x}_m \exp\left(\frac{\bar{Z}_i - \bar{z}_i(m)}{\bar{Z}_i + \bar{z}_i(m)}\right)$

and $\bar{x}^*(i, n) = \bar{x}_n \exp\left(\frac{\bar{Z}_i - \bar{z}_i(n)}{\bar{Z}_i + \bar{z}_i(n)}\right)$ for i=1, 2, 3, ..., p.

Such that $\mathbf{1}'\mathbf{W_m} = 1$, where $\mathbf{1}$ is a column vector of order p.

The optimum weights $\mathbf{W_u}$ and $\mathbf{W_m}$ in T_u and T_m are chosen by minimizing their mean square errors respectively.

Now a convex linear combination of the two estimators T_u and T_m has been considered to define the final estimator of population mean $\bar{Y}$ on the current occasion and is given as

$$T_{|p} = \varphi\, T_u + (1 - \varphi)\, T_m \tag{3}$$

where $\varphi(0 \le \varphi \le 1)$ is an unknown constant to be determined so as to minimise the mean square error of the estimator $T_{|p}$.

4. Properties of the Proposed Estimator $T_{|p}$

The properties of the proposed estimator $T_{|p}$ are derived under following large sample approximations:

$\bar{y}_u = \bar{Y}(1 + e_0)$, $\bar{y}_m = \bar{Y}(1 + e_1)$, $\bar{x}_m = \bar{X}(1 + e_2)$, $\bar{x}_n = \bar{X}(1 + e_3)$, $\bar{z}_i(u) = \bar{Z}_i(1 + e_{4i})$,

$\bar{z}_i(m) = \bar{Z}_i(1 + e_{5i})$ and $\bar{z}_i(n) = \bar{Z}_i(1 + e_{6i})$ such that $|e_k| < 1 \;\forall\; k = 0, 1, 2, 3, 4, 5$ and 6

and $|e_{ki}| < 1 \;\forall\; i = 1, 2, 3, ..., p$.

Under the above transformations, the estimators T_u and T_m take the following forms:

$$T(i, u) = \frac{\bar{Y}}{8}\left(8 + 8e_0 - 4e_{4i} - 4e_0 e_{4i} + 3e_{4i}^2\right) \quad \text{for } i=1, 2, ..., p \tag{4}$$

$$\begin{aligned} T(i, m, n) = \frac{\bar{Y}}{8}\big(&8 + 8e_1 - 8e_2 + 8e_3 - 4e_{6i} - 8e_1e_2 + 8e_1e_3 - 4e_1e_{6i} - 8e_2e_3 \\ &+ 4e_2e_{6i} - 4e_3e_{6i} + 8e_2^2 + 3e_{6i}^2\big) \quad \text{for } i=1, 2, ..., p \end{aligned} \tag{5}$$

Thus we have the following theorems:

Theorem 4.1: The bias of the proposed estimator $T_{|p}$ to the first order of approximation is obtained as

$$B(T_{|p}) = \varphi\, B(T_u) + (1 - \varphi)\, B(T_m) \tag{6}$$

$$B(T_u) = \frac{1}{u}\mathbf{W}_u'\ \mathbf{B}_u \tag{7}$$

$$B(T_m) = \mathbf{W}_m'\ \left(\frac{1}{m}B_{m1} + \frac{1}{n}\mathbf{B}_{m2}\right) \tag{8}$$

where $\mathbf{B}_u = \left(B_1(u), B_2(u), \ldots, B_p(u)\right)'$, $B_i(u) = \frac{1}{u}\bar{Y}\left(\frac{3}{8}\frac{C_{002}}{\bar{Z}_i^2} - \frac{1}{2}\frac{C_{011}}{\bar{Y}\,\bar{Z}_i}\right)$,

for i =1, 2, 3, ..., p

$B_{m1} = \bar{Y}\left(\frac{C_{200}}{\bar{X}^2} - \frac{C_{110}}{\bar{X}\bar{Y}}\right)$, $\mathbf{B}_{m2} = \left(Bm_{21}, Bm_{22}, \ldots, Bm_{2p}\right)'$

where $$Bm_{2\,i} = \bar{Y}\left(\frac{3}{8}\frac{C_{002}}{\bar{Z}_i^2} + \frac{C_{110}}{\bar{X}\bar{Y}} - \frac{1}{2}\frac{C_{011}}{\bar{Y}\bar{Z}_i} - \frac{C_{200}}{\bar{X}^2}\right), C_{rst} = E\left[\left(x_i - \bar{X}\right)^r \left(y_i - \bar{Y}\right)^s \left(z_i - \bar{Z}\right)^t\right];$$

$(r, s, t) \geq 0$ for i = 1, 2, 3, ..., p .

Proof: The bias of the estimator $T_{|p}$ is given by

$$B\left(T_{|p}\right) = E\left[T_{|p} - \bar{Y}\right] = \varphi\, B\left(T_u\right) + \left(1 - \varphi\right) B\left(T_m\right)$$

where $B\left(T_u\right) = E\left[T_u - \bar{Y}\right]$ and $B\left(T_m\right) = E\left[T_m - \bar{Y}\right]$

Using large sample approximations assumed in Section 4 and retaining terms up-to the first order of approximations, the expression for $T\left(i, u\right)$ and $T\left(i, m, n\right)$ for i=1, 2, 3, …, p are obtained as in equation (4) and equation (5) respectively and hence using equation (4) and (5) in equation (1) and (2) respectively the expression for $B\left(T_u\right)$ and $B\left(T_m\right)$ are obtained as in equations (7) and equation (8) respectively, hence the expression for bias of the estimator $T_{|p}$ is obtained as in equation (6).

Theorem 4.2: The mean square error of the estimator $T_{|p}$ is given by

$$M\left(T_{|p}\right) = \varphi^2\, M\left(T_u\right) + \left(1 - \varphi\right)^2\, M\left(T_m\right) + 2\varphi\left(1 - \varphi\right) Cov\left(T_u, T_m\right) \quad (9)$$

$$M(T_u) = \mathbf{W}_u'\, \mathbf{K}_u\, \mathbf{W}_u \quad (10)$$

$$M(T_m) = (B)\mathbf{W}_m'\, \mathbf{E}\, \mathbf{W}_m + \mathbf{W}_m'\, \mathbf{K}_m\, \mathbf{W}_m \quad (11)$$

where $\mathbf{W_u} = [w_{u_1} \quad w_{u_2} \cdot \quad \cdot \quad \cdot \quad w_{u_p}]'$, $\mathbf{W_m} = [w_{m_1} \quad w_{m_2} \cdot \quad \cdot \quad \cdot \quad w_{m_p}]'$, $\mathbf{E}$ is a unit matrix of order $p \times p$, $\mathbf{K}_u = \left(\frac{1}{u} - \frac{1}{N}\right)\mathbf{K}_{u*}$, $\mathbf{K}_m = \left(\frac{1}{n} - \frac{1}{N}\right)\mathbf{K}_{m*}$ where

$$\mathbf{K}_{u*} = \begin{bmatrix} ku_{11} & ku_{12} & . & . & . & ku_{1p} \\ ku_{21} & ku_{22} & . & . & . & ku_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ ku_{p1} & ku_{p2} & . & . & . & ku_{pp} \end{bmatrix}_{p \times p} \quad \text{and} \quad \mathbf{K}_{m*} = \begin{bmatrix} km_{11} & km_{12} & . & . & . & km_{1p} \\ km_{21} & km_{22} & . & . & . & km_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ km_{p1} & km_{p2} & . & . & . & km_{pp} \end{bmatrix}_{p \times p}$$

where $B = \left(\frac{1}{m} - \frac{1}{N}\right) B_1$, $B_1 = 2\,\bar{Y}^2 \left(1 - \rho_{yx}\right) C_0^2$, $ku_{ii} = \bar{Y}^2 \left(C_0^2 + \frac{1}{4} C_{z_i}^2 - \rho_{yz_i} C_0 C_{z_i} \right)$,

$$ku_{ij} = \bar{Y}^2 \left(C_0^2 - \frac{1}{2} \rho_{yz_i} C_0 C_{z_i} - \frac{1}{2} \rho_{yz_j} C_0 C_{z_j} + \frac{1}{4} \rho_{z_i z_j} C_{z_i} C_{z_j} \right)$$

$$km_{ii} = \bar{Y}^2 \left(C_0^2 \left(2\rho_{yx} - 1\right) - \rho_{yz_i} C_0 C_{z_i} + \frac{1}{4} C_{z_i}^2 \right) \text{ and}$$

$$km_{ij} = \bar{Y}^2 \left(C_0^2 \left(2\rho_{yx} - 1\right) - \frac{1}{2} \rho_{yz_i} C_0 C_{z_i} - \frac{1}{2} \rho_{yz_j} C_0 C_{z_j} + \frac{1}{4} \rho_{z_i z_j} C_{z_i} C_{z_j} \right) \forall\, i \neq j = 1, 2, 3, ..., p.$$

Proof: The mean square error of the estimator $T_{|p}$ is given by

$$M\left(T_{|p}\right) = E\left[T_{|p} - \bar{Y}\right]^2 = E\left[\varphi\left(T_u - \bar{Y}\right) + (1 - \varphi)\left(T_m - \bar{Y}\right)\right]^2$$

$$= \varphi^2\, M\left(T_u\right) + \left(1 - \varphi\right)^2 M\left(T_m\right) + 2\,\varphi\,\left(1 - \varphi\right) \text{Cov}\left(T_u, T_m\right)$$

where $M\left(T_u\right) = E\left[T_u - \bar{Y}\right]^2$ and $M\left(T_m\right) = E\left[T_m - \bar{Y}\right]^2$;

The estimators T_u and T_m are based on two independent samples of sizes u and m respectively, hence $\text{Cov}\left(T_u, T_m\right) = 0$; Considering the population is sufficiently large so using large sample approximations assumed in section 4 and retaining terms upto the first order of approximations and also assuming $C_x = C_y = C_0$ (following Cochran(1977)), the expression for $M\left(T_u\right)$ and $M\left(T_m\right)$ are obtained as given in equation (10) and (11) and hence the expression for mean square error of estimator $T_{|p}$ is obtained as in equation (9).

5. Choice of Optimal Weights

To find the optimization of the weight vector $\mathbf{W}_u = [w_{u_1} \quad w_{u_2} \cdot \quad \cdot \quad \cdot \quad w_{u_p}]'$, the mean square error $M(T_u)$ given in equation (10) is minimized subject to the condition $\mathbf{1}'\mathbf{W}_u = 1$ using the method of Lagrange's Multiplier explained as:

To find the extrema using Lagrange's Multiplier Technique, we define L_u as

$$L_u = \mathbf{W}_u' \mathbf{K}_u \mathbf{W}_u - \lambda_u(\mathbf{1}'\mathbf{W}_u - 1), \quad (12)$$

where $\mathbf{1}$ is a unit column vector of order p and λ_u is the Lagrangian multiplier.

Now, by differentiating equation (12) partially with respect to $\mathbf{W}_u$ and equating it to zero we have

$$\frac{\partial L_u}{\partial \mathbf{W}_u} = \frac{\partial}{\partial \mathbf{W}_u}[\mathbf{W}_u' \mathbf{K}_u \mathbf{W}_u - \lambda_u(\mathbf{1}'\mathbf{W}_u - 1)] = 0$$

This implies that, $2\,\mathbf{K}_u\mathbf{W}_u - \lambda_u\mathbf{1} = \mathbf{0}$, which yields

$$\mathbf{W}_u = \frac{\lambda_u}{2}\,\mathbf{K}_u^{-1}\,\mathbf{1} \quad (13)$$

Now pre- multiplying equation (13) by $\mathbf{1}'$, we get

$$\frac{\lambda_u}{2} = \frac{1}{\mathbf{1}'\,\mathbf{K}_u^{-1}\,\mathbf{1}} \quad (14)$$

Thus, using equation (14) in equation (13), we obtain the optimal weight vector as

$$\mathbf{W}_{u_{opt.}} = \frac{\mathbf{K}_u^{-1}}{\mathbf{1}'\mathbf{K}_u^{-1}\mathbf{1}} \quad (15)$$

In similar manners, the optimal of the weight $\mathbf{W}_m = [w_{m_1} \quad w_{m_2} \cdot \quad \cdot \quad \cdot \quad w_{m_p}]'$ is obtained by minimizing $M(T_m)$ subject to the constraint $\mathbf{1}'\mathbf{W}_m = 1$ using the method of Lagrange's multiplier, for this we define

$$L_m = (B)\mathbf{W}_m'\,\mathbf{E}\,\mathbf{W}_m + \mathbf{W}_m'\,\mathbf{K}_m\,\mathbf{W}_m - \lambda_m(\mathbf{1}'\mathbf{W}_m - 1),$$

where λ_m is the Lagrangian multiplier.

Now, differentiating L_m with respect to $\mathbf{W}_m$ and equating to 0, we get

$$\mathbf{W}_{m_{opt.}} = \frac{\mathbf{K}_m^{-1}}{\mathbf{1}'\mathbf{K}_m^{-1}\mathbf{1}} \quad (16)$$

Then substituting the optimum values of $\mathbf{W}_u$ and $\mathbf{W}_m$ in equations (10) and (11) respectively, the optimum mean square errors of the estimators are obtained as:

$$M(T_u)_{opt.} = \left(\frac{1}{u} - \frac{1}{N}\right)\frac{1}{\mathbf{1}'\,\mathbf{K}_{u*}^{-1}\,\mathbf{1}} \quad (17)$$

$$M(T_m)_{opt.} = \left(\frac{1}{m} - \frac{1}{N}\right)B_1 + \left(\frac{1}{n} - \frac{1}{N}\right)\frac{1}{\mathbf{1}'\,\mathbf{K}_{m*}^{-1}\,\mathbf{1}} \quad (18)$$

6. Minimum Mean Square Error of the Proposed Estimator $T_{|p}$

The mean square error of the proposed estimator $T_{|p}$ is given by

$$M\left(T_{|p}\right) = \varphi^2 M\left(T_u\right)_{opt.} + \left(1 - \varphi\right)^2 M\left(T_m\right)_{opt.}$$

Minimizing $M\left(T_{|p}\right)$ with respect to φ gives the optimum value of φ as

$$\varphi_{opt.} = \frac{M\left(T_m\right)_{opt.}}{M\left(T_u\right)_{opt.} + M\left(T_m\right)_{opt.}} \tag{19}$$

Now substituting the above value of $\varphi_{opt.}$ in equation (9), we obtain the optimum mean square error of the estimators $T_{|p}$ as

$$M\left(T_{|p}\right)^*_{opt.} = \frac{M\left(T_u\right)_{opt.} \,.\, M\left(T_m\right)_{opt.}}{M\left(T_u\right)_{opt.} + M\left(T_m\right)_{opt.}} \tag{20}$$

Further, substituting the optimum values of the mean square errors of the estimators given in equations (17) and (18) in equation (19) and (20) respectively, the simplified values $\varphi_{opt.}$ and $M\left(T_{|p}\right)^*_{opt.}$ are obtained as

$$\varphi_{opt.} = \frac{\mu\left[\mu\, C - \left(B_1 + C\right)\right]}{\left[\mu^2\, C - \mu\left(B_1 + C - A\right) - A\right]} \tag{21}$$

$$M\left(T_{|p}\right)^*_{opt.} = \frac{1}{n}\frac{\left[\mu\, D_1 - D_2\right]}{\left[\mu^2\, C - \mu\, D_3 - A\right]} \tag{22}$$

where

$A = \frac{1}{\mathbf{1}'\,\mathbf{K}_{u*}^{-1}\,\mathbf{1}}$, $B_1 = 2\,\bar{Y}^2\left(1 - \rho_{yx}\right)C_0^2$, $\quad C = \frac{1}{\mathbf{1}'\mathbf{K}_{m*}^{-1}\,\mathbf{1}}$, $\quad D_1 = A\,C$, $\quad D_2 = A\,B_1 + A\,C$, $D_3 = B_1 + C - A$ and μ is the fraction of the sample drawn afresh at the current (second) occasion.

7. Optimum Replacement Strategy for the Estimator $T_{|p}$

The idea of longitudinal surveys is mainly concerned with obtaining efficient estimates with minimal cost in carrying out the survey. So it is technically convenient to maintain a high overlap between repeats of the survey which provides the advantage due to many sampled units being located and have some experience in the survey. Hence the decision of the optimum value of μ should be made (fractions of samples to be drawn afresh on the current occasion) so that $\bar{Y}$ may be estimated with maximum precision and minimum cost, we minimize the mean square error $M\left(T_{|p}\right)^{*}_{opt.}$ in equation (22) with respect to μ as:

$$\frac{\partial\left(M\left(T_{|p}\right)^{*}_{opt.}\right)}{\partial\,\mu} = 0,$$

$$\Rightarrow \qquad \mu^2 G_1 - 2\mu G_2 + G_3 = 0,$$

Thus the optimum value of μ so obtained is one of the two roots given by

$$\mu = \frac{G_2 \pm \sqrt{G_2^2 - G_1 G_3}}{G_1} \tag{23}$$

where $G_1 = C D_1$, $G_2 = C D_2$ and $G_3 = A D_1 + D_2 D_3$.

The real value of μ exist, iff $G_2^2 - G_1 G_3 \geq 0$. For any situation, which satisfies this condition, two real values of μ may be possible, hence choose a value of μ such that $0 \leq \mu \leq 1$. All other values of μ are inadmissible. If both the real values of μ are admissible, the lowest one will be the best choice as it reduces the total cost of the survey. Substituting the admissible value of μ say $\mu_{T_{|p}}$ from (23) in to the equation (22), we get the optimum value of the mean square error of the estimator $T_{|p}$ with respect to φ as well as μ which, is given as

$$M\left(T_{|p}\right)^{**}_{opt.} = \frac{1}{n}\frac{\left[\mu_{T_{|p}} D_1 - D_2\right]}{\left[\mu^2_{T_{|p}} C - \mu_{T_{|p}} D_3 - A\right]}. \tag{24}$$

8. Efficiency with Increased Number of Auxiliary Variables

As we know that increasing the number of auxiliary variables typically increases the precision of the estimates. In this section we verify this property for the proposed estimator as under: Let $T_{|p}$ and $T_{|q}$ be two proposed estimators based on p and q auxiliary variables respectively such that $p < q$, then $M(T_{|p}) \geq M(T_{|q})$, i.e.

$$M(T_{|p}) - M(T_{|q}) \geq 0 \tag{25}$$

$$\frac{1}{n}\frac{\left[\mu A_p C_p - A_p (B + C_p)\right]}{\left[\mu^2 C_p - \mu (B + C_p + A_p) - A_p\right]} - \frac{1}{n}\frac{\left[\mu A_q C_q - A_q (B + C_q)\right]}{\left[\mu^2 C_q - \mu (B + C_q + A_q) - A_q\right]} \geq 0$$

On simplification, we get

$$(A_p - A_q)\left[(\mu - 1)^2\left(\mu C_p C_q + \frac{A_p A_q (C_p - C_q)}{(A_p - A_q)}\right) - \mu B\left((C_p - C_q)(\mu - 1) - B\right)\right] \geq 0$$

This reduces to the condition

$$(A_p - A_q) \geq 0 \tag{26}$$

So from Section 6 above, we get

$$\frac{1}{\mathbf{1}' \mathbf{K}_p^{-1} \mathbf{1}} - \frac{1}{\mathbf{1}' \mathbf{K}_q^{-1} \mathbf{1}} \geq 0$$

$$\mathbf{1}' \mathbf{K}_q^{-1} \mathbf{1} \geq \mathbf{1}' \mathbf{K}_p^{-1} \mathbf{1}$$

Following Rao (2006), the matrix $\mathbf{K}_q$ can be partitioned and can be written as

$$\mathbf{K}_q = \begin{pmatrix} \mathbf{K}_p & \mathbf{F} \\ \mathbf{F}' & \mathbf{G} \end{pmatrix}$$

where $\mathbf{F}$, $\mathbf{F}'$ and $\mathbf{G}$ are matrices deduced from $\mathbf{K}_q$ such that their order never exceeds q-p and always greater than or equal to 1. Then,

$$\mathbf{K}_q^{-1} = \begin{pmatrix} \mathbf{K}_p^{-1} + \mathbf{HJH}' & -\mathbf{HJ} \\ -\mathbf{JH}' & \mathbf{J} \end{pmatrix} \tag{27}$$

where $\mathbf{J} = \left(\mathbf{G} - \mathbf{F}'\mathbf{K}_p^{-1}\mathbf{F}\right)^{-1}$ and $\mathbf{H} = \mathbf{K}_p^{-1}\mathbf{F}$. (See Rao (2006) and Olkin(1958))

Now rewriting $\mathbf{1}'\mathbf{K}_q^{-1}\mathbf{1}$ by putting the value of $\mathbf{K}_q^{-1}$ from equation (27), we get

$$\mathbf{1}'\mathbf{K}_q^{-1}\mathbf{1} = \begin{pmatrix}\mathbf{1}_p & \mathbf{1}_{q-p}\end{pmatrix}' \begin{pmatrix} \mathbf{K}_p^{-1} + \mathbf{HJH}' & -\mathbf{HJ} \\ -\mathbf{JH}' & \mathbf{J} \end{pmatrix} \begin{pmatrix} \mathbf{1}_p \\ \mathbf{1}_{q-p} \end{pmatrix}$$

$$= \begin{pmatrix}\mathbf{1}_p'\left(\mathbf{K}_p^{-1} + \mathbf{HJH}'\right) - \mathbf{1}_{q-p}'\mathbf{JH}' & -\mathbf{1}_p'\mathbf{HJ} + \mathbf{1}_{q-p}'\mathbf{J}\end{pmatrix} \begin{pmatrix} \mathbf{1}_p \\ \mathbf{1}_{q-p} \end{pmatrix}$$

$$= \mathbf{1}'_p \left(\mathbf{K}_p^{-1} + \mathbf{HJH}'\right)\mathbf{1}_p \text{ - } \mathbf{1}'_{q-p}\mathbf{JH}'\,\mathbf{1}_p \text{ - } \mathbf{1}'_p\,\mathbf{HJ}\,\mathbf{1}_{q-p} + \mathbf{1}'_{q-p}\mathbf{J}\,\mathbf{1}_{q-p}$$

$$\Rightarrow \mathbf{1}'\,\mathbf{K}_q^{-1}\,\mathbf{1} \text{ - } \mathbf{1}'_p \left(\mathbf{K}_p^{-1}\right)\mathbf{1}_p = \mathbf{1}'_p\left(\mathbf{HJH}'\right)\mathbf{1}_p \text{ - } \mathbf{1}'_{q-p}\mathbf{JH}'\,\mathbf{1}_p \text{ - } \mathbf{1}'_p\,\mathbf{HJ}\,\mathbf{1}_{q-p} + \mathbf{1}'_{q-p}\mathbf{J}\,\mathbf{1}_{q-p}$$

$$\mathbf{1}'\,\mathbf{K}_q^{-1}\,\mathbf{1} \text{ - } \mathbf{1}'_p \left(\mathbf{K}_p^{-1}\right)\mathbf{1}_p = \left(\mathbf{1}_p \quad \mathbf{1}_{q-p}\right)' \begin{pmatrix} \mathbf{HJH}' & \text{-}\,\mathbf{HJ} \\ \text{-}\mathbf{JH}' & \mathbf{J} \end{pmatrix} \begin{pmatrix} \mathbf{1}_p \\ \mathbf{1}_{q-p} \end{pmatrix}$$

$$\mathbf{1}'\,\mathbf{K}_q^{-1}\,\mathbf{1} \text{ - } \mathbf{1}'_p \left(\mathbf{K}_p^{-1}\right)\mathbf{1}_p = \mathbf{1}' \begin{pmatrix} \mathbf{H} \\ \text{-}\mathbf{I} \end{pmatrix} \mathbf{J} \left(\mathbf{H} \quad \text{-}\mathbf{I}\right) \mathbf{1} \geq 0$$

The latter follows since **J** is positive definite so that $\mathbf{R}'\mathbf{J}\,\mathbf{R} \geq 0$ for all **R,**

where $\mathbf{R} = \left(\mathbf{H} \quad \text{-}\mathbf{I}\right)\mathbf{1}$.

Hence from equation (25), we have

$$M\left(T_{|p}\right) - M\left(T_{|q}\right) \geq 0$$

This leads to the result that utilizing more auxiliary variables provides more efficient estimates in terms of mean square error for the proposed estimator.

9. Special Cases

Case 1:

There are several instances where the p-auxiliary variates are mutually uncorrelated but they are correlated to study variates for example, in survey of commercial product say the aim is to estimate the number of persons reading newspaper. Then in that case the numbers of copies produced by different newspaper companies are different and number of copies produced by a particular newspaper company is uncorrelated to the number of copies produced by another newspaper but both are correlated to study variates, i.e., number of persons reading newspaper. Similarly, in transportation survey if the aim is to estimate the number of persons traveling by air per year, then in that case the total seating capacity of different airlines may be treated as auxiliary variates. Since, the seating capacity of different airlines is different and they are mutually uncorrelated but the information on this will contribute a lot in estimation of the number of persons traveling by air. Hence, for modelling such type of situations where the p-auxiliary variates are mutually uncorrelated, i.e. $\rho_{z_i z_j} = 0 \ \forall \ i \neq j = 1, 2, ..., p$, the proposed multivariate exponential ratio type estimator is applicable and in this case optimum value of μ say μ^0 and the optimum value of the mean square error of the estimator $T_{|p}$ with respect to φ as well as μ^0 is given by

$$\mu^0 = \frac{G_2^0 \pm \sqrt{G_2^{0^2} - G_1^0 \ G_3^0}}{G_1^0} \tag{28}$$

and
$$M\left(T_{|p}\right)^{**}_{opt.} = \frac{1}{n} \frac{\left[\mu^0{}_{T_{|p}} \ D_1^0 \ - \ D_2^0\right]}{\left[\mu^{0^2}_{T_{|p}} \ C^0 \ - \mu^0{}_{T_{|p}} \ D_3^0 \ - A^0\right]} \tag{29}$$

where $G_1^0 = C^0 \ D_1^0$, $G_2^0 = C^0 \ D_2^0$, $G_3^0 = A^0 \ D_1^0 + D_2^0 \ D_3^0$, $A^0 = \frac{1}{\mathbf{1}' \mathbf{S}_{u*}^{-1} \mathbf{1}}$,

$B_1^0 = 2 \ \bar{Y}^2 \left(1 - \rho_{yx}\right) C_0^2$,

$C^0 = \frac{1}{\mathbf{1}' \mathbf{S}_{m*}^{-1} \mathbf{1}}$, $D_1^0 = A^0 \ C^0$, $D_2^0 = A^0 \ B_1^0 + A^0 \ C^0$, $D_3^0 = B_1^0 + C^0 - A^0$,

$$\mathbf{S}_{u*} = \begin{bmatrix} su_{11} & su_{12} & . & . & . & su_{1p} \\ su_{21} & su_{22} & . & . & . & su_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ su_{p1} & su_{p2} & . & . & . & su_{pp} \end{bmatrix}_{p \times p} \quad \text{and} \quad \mathbf{S}_{m*} = \begin{bmatrix} sm_{11} & sm_{12} & . & . & . & sm_{1p} \\ sm_{21} & sm_{22} & . & . & . & sm_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ sm_{p1} & sm_{p2} & . & . & . & sm_{pp} \end{bmatrix}_{p \times p}$$

where $B_1^0 = 2 \ \bar{Y}^2 \left(1 - \rho_{yx}\right) C_0^2$, $su_{ii} = \bar{Y}^2 \left(C_0^2 + \frac{1}{4} C_{z_i}^2 - \rho_{yz_i} C_0 C_{z_i} \right)$,

$su_{ij} = \bar{Y}^2 \left(C_0^2 - \frac{1}{2} \rho_{yz_i} C_0 C_{z_i} - \frac{1}{2} \rho_{yz_j} C_0 C_{z_j} \right)$, $sm_{ii} = \bar{Y}^2 \left(C_0^2 \left(2\rho_{yx} - 1\right) - \rho_{yz_i} C_0 C_{z_i} + \frac{1}{4} C_{z_i}^2 \right)$

and $sm_{ij} = \bar{Y}^2 \left(C_0^2 \left(2\rho_{yx} - 1\right) - \frac{1}{2} \rho_{yz_i} C_0 C_{z_i} - \frac{1}{2} \rho_{yz_j} C_0 C_{z_j} \right) \ \forall \ i \neq j = 1, 2, 3, ..., p$.

Case 2:

The p-auxiliary variates are mutually correlated i.e. $\rho_{z_i z_j} \neq 0 \ \forall \ i \neq j = 1, 2, ..., p$. In this case if there is high correlation between p-auxiliary variates, then such a problem can be addressed as a problem of multi collinearity in survey sampling.

10. Efficiency Comparison

In order to examine the performance of the proposed estimator with some of the recent estimators due to Singh (2005) and Singh and Priyanka (2008a) in successive sampling, same assumptions have been considered for proposed estimator for discussing the properties of estimators as that of estimators proposed by Singh (2005) and Singh and Priyanka (2008a).

Hence, following Olkin (1958), Raj (1965), Artes and Garcia (2005) and Singh et al. (2011) we consider $C_0 = C_{z_i}; \forall$ i=1, 2, 3, ..., p approximately and hence, the optimum value of μ say $\hat{\mu}^*$ and optimum value of mean square error $M\left(T_{|p}\right)^{**}_{opt.}$ of the proposed estimator $T_{|p}$ reduces to

$$\mu^* = \frac{G_2^* \pm \sqrt{G_2^{*2} - G_1^* G_3^*}}{G_1^*} \tag{30}$$

$$M\left(T_{|p}\right)^{**}_{opt.} = \frac{1}{n}\frac{\left[\mu^*_{T_{|p}} D_1^* - D_2^*\right]}{\left[\mu^{*2}_{T_{|p}} C^* - \mu^*_{T_{|p}} D_3^* - A^*\right]} \tag{31}$$

where $G_1^* = C^* D_1^*$, $G_2^* = C^* D_2^*$, $G_3^* = A^* D_1^* + D_2^* D_3^*$, $A^* = \frac{1}{\mathbf{1'}\,\mathbf{H}_{u^*}^{-1}\,\mathbf{1}}$, $C^* = \frac{1}{\mathbf{1'}\mathbf{H}_{m^*}^{-1}\,\mathbf{1}}$,

$D_1^* = A^* C^*$, $D_2^* = A^* B_1^* + A^* C^*$, $D_3^* = B_1^* + C^* - A^*$, $B_1^* = 2\left(1 - \rho_{yx}\right)S_y^2$,

$$\mathbf{H}_{u*} = \begin{bmatrix} hu_{11} & hu_{12} & . & . & . & hu_{1p} \\ hu_{21} & hu_{22} & . & . & . & hu_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ hu_{p1} & hu_{p2} & . & . & . & hu_{pp} \end{bmatrix}_{p\times p} \quad \text{and} \quad \mathbf{H}_{m*} = \begin{bmatrix} hm_{11} & hm_{12} & . & . & . & hm_{1p} \\ hm_{21} & hm_{22} & . & . & . & hm_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ hm_{p1} & hm_{p2} & . & . & . & hm_{pp} \end{bmatrix}_{p\times p}$$

$B_1^* = 2\left(1 - \rho_{yx}\right)S_y^2$, $hu_{ii} = \left(\frac{5}{4} - \rho_{yz_i}\right) S_y^2$, $hu_{ij} = \left(1 - \frac{1}{2}\rho_{yz_i} - \frac{1}{2}\rho_{yz_j} + \frac{1}{4}\rho_{z_i z_j}\right) S_y^2$,

$hm_{ii} = \left(2\rho_{yx} - \rho_{yz_i} - \frac{3}{4}\right) S_y^2$ and $hm_{ij} = \left(2\rho_{yx} - \frac{1}{2}\rho_{yz_i} - \frac{1}{2}\rho_{yz_j} + \frac{1}{4}\rho_{z_i z_j} - 1\right) S_y^2$

$\forall$ i $\neq$ j=1, 2, 3,..., p.

10.1 Comparison of the proposed Estimator $T_{|p}$ with respect to estimator T_S due to Singh (2005)

The estimator proposed by Singh (2005) is given as

$$T_S = \psi \frac{\bar{y}_u}{\bar{z}_u}\bar{Z} + (1 - \psi)\frac{\bar{y}_m}{\bar{x}_m}\frac{\bar{x}_n}{\bar{z}_n}\bar{Z}, \tag{32}$$

and the optimum mean square error of this estimator T_S is given by

$$M(T_S)_{opt.} = \frac{\left[\alpha_1^2 + \alpha_1\alpha_2\ \mu_S\right] S_y^2}{n\left[\alpha_1 + \alpha_2\ \mu_S^2\right]}$$

with $\alpha_1 = 2\ (1 - \rho_{yz})$, $\alpha_2 = 2\ (\rho_{yz} - \rho_{yx})$ and

$\mu_S = -(1 - \rho_{yz}) \pm \sqrt{(1 - \rho_{yz})(1 - \rho_{yx})} \Big/ (\rho_{yz} - \rho_{yx})$.

Hence, the percent relative efficiency of the proposed estimator with respect to T_S is given as

$$E_{T_{|p}}^{S} = \frac{M[T_s]_{opt.}}{M(T_{|p})_{opt.}^{**}} \times 100 \tag{33}$$

10.2 Comparison of the proposed estimator $T_{|p}$ with respect to estimator T_{SP} due to Singh and Priyanka (2008a)

The proposed estimator $T_{|p}$ at optimum condition is also compared with respect to the estimator T_{SP} due to Singh and Priyanka (2008a) given as

$$T_{SP} = \xi\left[\bar{y}_u + \beta_{yz}(\bar{Z} - \bar{z}_u)\right] + (1 - \xi)\left[\bar{y}_m^* + \beta_{yx}(\bar{x}_n^* - \bar{x}_m^*)\right], \tag{34}$$

where $\bar{y}_m^* = \bar{y}_m + \beta_{yz}(\bar{Z} - \bar{z}_m)$, $\bar{x}_n^* = \bar{x}_n + \beta_{xz}(\bar{Z} - \bar{z}_n)$, $\bar{x}_m^* = \bar{x}_m + \beta_{xz}(\bar{Z} - \bar{z}_m)$, β_{yz} and β_{xz} are the population regression coefficients of y on z and x on z respectively and ξ is constant so as to minimize the variance of the estimator T_{SP}.

The optimum variance of estimator T_{SP} is given as

$$V\left(T_{SP}\right)^{*}_{opt.} = \frac{\kappa\left[\kappa + \mu_{SP}\ \chi\right]}{\left[\kappa + \mu_{SP}^{2}\ \chi\right]}\frac{S_y^2}{n}$$

where $\kappa = 1 - \rho_{yz}^2$, $\chi = 2\rho_{yz}^2\rho_{yx} - \rho_{yx}^2\left(1 + \rho_{yz}^2\right)$ and $\mu_{SP} = \dfrac{-\kappa \pm \sqrt{\kappa\,(\kappa + \chi)}}{\chi}$.

Hence, the percent relative efficiency $E_{T_{|p}}^{SP}$ for (p=1, 2, 3 …) of the estimator $T_{|p}$ (under their respective optimum conditions) with respect to T_{SP} is given by

$$E_{T_{|p}}^{SP} = \frac{V\left[T_{SP}\right]^{*}_{opt.}}{M\left(T_{|p}\right)^{**}_{opt.}} \times 100\ ;\ \text{for}\ (p = 1, 2, 3\ldots). \tag{35}$$

11. Empirical Illustrations and Monte Carlo Simulation

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

For Carrying out the empirical study the population of total electric consumption in different states of United States has been considered.

For carrying out numerical illustration we have considered the case of three auxiliary information (i.e. p=3) which are stable over time and are available at both the occasions. The population comprise of N = 51 states of the United States. Let

y_i : The total energy consumption during 2007 in the i^{th} state of U. S.

x_i : The total energy consumption during 2002 in the i^{th} state of U. S.

z_{1i} : The total energy consumption during 2001 in the i^{th} state of U. S.

z_{2i} : The total energy consumption during 2000 in the i^{th} state of U. S.

z_{3i} : The total energy consumption during 1999 in the i^{th} state of U. S.

For the considered population, the values of $\mu_{T_{|p}}^{*}$ (p=1, 2 and 3) defined in equation (31) and percent relative efficiencies $E_{T_{|p}}^{S}$ and $E_{T_{|p}}^{SP}$ defined in equation (33) and (35) of $T_{|p}$ (p= 1, 2 and 3) with respect to T_S and T_{SP} have been computed and are presented in Table-2.

11.1 Simulation Algorithm

(i) Choose 5000 samples of size n=20 using simple random sampling without replacement on first occasion for both the study and auxiliary variables.

(ii) Calculate sample mean $\bar{x}_{n|k}$, $\bar{z}_{1_{n|k}}$, $\bar{z}_{2_{n|k}}$ and $\bar{z}_{3_{n|k}}$ for k =1, 2, - - -, 5000.

(iii) Retain m=17 units out of each n=20 sample units of the study and auxiliary variables at the first occasion.

(iv) Calculate sample mean $\bar{x}_{m|k}$, $\bar{z}_{1_{m|k}}$, $\bar{z}_{2_{m|k}}$ and $\bar{z}_{3_{m|k}}$ for k= 1, 2, - - -, 5000.

(v) Select u=3 units using simple random sampling without replacement from N-n=31 units of the population for study and auxiliary variables at second (current) occasion.

(vi) Calculate sample mean $\bar{y}_{u|k}$, $\bar{y}_{m|k}$, $\bar{z}_{1_{u|k}}$, $\bar{z}_{2_{u|k}}$ and $\bar{z}_{3_{u|k}}$ for k = 1, 2, - - -, 5000 .

(vii) Iterate the parameter φ from 0.1 to 0.9 with a step of 0.1.

(viii) Iterate ψ from 0.1 to 0.9 with a step of 0.1 within (vii).

(ix) Calculate the percent relative efficiencies of the proposed estimator $T_{|p}\ (p=1, 2 \text{ and } 3)$ with the case p=1, p=2 and p=3(i.e. $T_{|p=1}$, $T_{|p=2}$ and $T_{|p=3}$) with respect to estimator due to Singh (2005) and Singh and Priyanka (2008) as

$$E_p(S) = \frac{\sum_{k=1}^{5000}\left[T_{S|k} - \bar{Y}\right]^2}{\sum_{k=1}^{5000}\left[T_{|p|k} - \bar{Y}\right]^2} \times 100 \text{ and } E_p(SP) = \frac{\sum_{k=1}^{5000}\left[T_{SP|k} - \bar{Y}\right]^2}{\sum_{k=1}^{5000}\left[T_{|p|k} - \bar{Y}\right]^2} \times 100, \text{ k=1, 2, ..., 5000.}$$

Table 1: Empirical comparison of the proposed estimator $T_{|p}$ (p=1, 2 and 3) with respect to the estimators T_S and T_{SP} respectively at their optimum conditions.

		p=1	p=2	p=3
		$\mu_1^* = 0.5355$	$\mu_2^* = 0.5196$	$\mu_3^* = 0.5137$
$E^{S}_{T_{\|p}}$	$\mu_S = 0.5502$	109.55	120.77	125.17
$E^{SP}_{T_{\|p}}$	$\mu_{SP} = 0.5496$	**	101.44	105.14

Note: '**' denote estimator $T_{|\,p=1}$ does not perform better than T_{SP} in terms of efficiency.

Table 2: Monte Carlo Simulation results when the proposed estimator $T_{|\,p=1}$ is compared to T_S and T_{SP} respectively (considering $\psi = \xi$).

φ \ ψ		**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0.1	$E_1(S)$	1349. 9	5895.1	12296.0	17324.0	36619.0	57157.0	72197.0	84832.0	84170.0
	$E_1(SP)$	104.42	102.05	127.52	171.80	262.54	375.88	520.91	638.86	867.55
0.3	$E_1(S)$	1168.9	4741.7	11796.0	16766.0	32572.0	45474.0	61039.0	81957.0	97409.0
	$E_1(SP)$	**	**	113.58	159.70	237.69	329.08	469.05	577.94	764.94
0.5	$E_1(S)$	894.8	3399.7	8454.4	12852.0	22913.0	30904.0	41579.0	60708.0	70099.0
	$E_1(SP)$	**	**	**	113.93	167.23	231.62	323.22	410.61	540.09
0.7	$E_1(S)$	608.0	2258.0	5620.8	8907.0	14926.0	19794.0	27410.0	40345.0	46610.0
	$E_1(SP)$	**	**	**	**	109.21	153.80	211.71	272.91	358.30
0.9	$E_1(S)$	413.45	1493.1	3817.9	5946.8	10156.0	13715.0	18796.0	26496.0	31819.0
	$E_1(SP)$	**	**	**	**	**	105.43	141.45	184.93	242.63

Note: '**' denotes estimator $T_{|\,p=1}$ does not perform better than T_{SP} in terms of efficiency.

Table 3: Monte Carlo Simulation results when the proposed estimator $T_{|\,p=2}$ are compared to T_S and T_{SP} respectively (considering $\psi = \xi$).

φ \ ψ		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	$E_2(S)$	1335.2	5940.6	12400.0	17376.0	36883.0	57438.0	73062.0	85114.0	84732.0
	$E_2(SP)$	104.44	102.84	128.60	172.31	264.43	377.72	527.15	640.98	873.34
0.3	$E_2(S)$	1178.3	11795.1	11935.0	16892.0	32827.0	45857.0	61624.0	82646.0	98297.0
	$E_2(SP)$	**	**	114.92	160.90	239.55	331.85	473.55	582.80	771.92
0.5	$E_2(S)$	902.63	3431.4	8553.5	12956.0	23109.0	31143.0	41964.0	61237.0	70719.0
	$E_2(SP)$	**	**	**	114.85	168.66	233.41	326.21	414.19	544.87
0.7	$E_2(S)$	612.36	2276.4	5673.3	8977.4	15037.0	19933.0	27637.0	40674.0	46957.0
	$E_2(SP)$	**	**	**	**	110.02	154.88	213.46	275.13	360.96
0.9	$E_2(S)$	416.00	1502.6	3846.8	5988.2	10221.0	13801.0	18936.0	26671.0	32009
	$E_2(SP)$	**	**	**	**	**	106.09	142.50	186.15	244.08

Note: '**' denotes estimator $T_{|\,p=2}$ does not perform better than T_{SP} in terms of efficiency.

Table 4: Monte Carlo Simulation results when the proposed estimator $T_{|\,p=3}$ are compared to T_S and T_{SP} respectively (considering $\psi = \xi$).

φ \ ψ		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	$E_3(S)$	1407.7	6252.0	12967.0	18284.0	38857.0	61313.0	76382.0	89262.0	88716.0
	$E_3(SP)$	110.12	108.23	134.48	181.31	278.58	403.21	551.10	672.22	914.40
0.3	$E_3(S)$	1222.2	5606.4	12300.0	17513.0	34076.0	47645.0	63630.0	85846.0	10116.0
	$E_3(SP)$	104.40	104.34	118.44	166.82	248.67	344.79	488.95	605.36	794.41
0.5	$E_3(S)$	934.23	3549.5	8812.8	13413.0	23967.0	32399.0	43198.0	63448.0	73073.0
	$E_3(SP)$	**	**	**	118.90	128.23	242.82	335.81	429.14	563.01
0.7	$E_3(S)$	638.63	2371.6	5891.4	9397.2	15675.0	20776.0	28770.0	42331.0	490.21
	$E_3(SP)$	**	**	**	**	114.69	161.42	222.22	286.35	376.83
0.9	$E_3(S)$	436.36	1577.9	4023.4	6288.2	10711.0	14474.0	19833.0	27952.0	33593.0
	$E_3(SP)$	**	**	**	**	**	111.27	149.26	195.09	256.16

Note: '**' denotes estimator $T_{|\,p=3}$ does not perform better than T_{SP} in terms of efficiency.

11.2 Performance of proposed estimators for various choices of correlation coefficients

To validate the applicability and performance of the proposed estimator, $T_{|p}$ (for p=1, 2 and 3) has been compared with Singh (2005) and Singh and Priyanka (2008a) at their respective optimum conditions for various combinations of correlation coefficients and results are tabulated in Table 5 to Table 7.

Table 5: For p =1

ρ_{yx}	0.5					0.6					0.7				
ρ_{yz_1}	μ_1^*	μ_S	μ_{SP}	$E^S_{T_{\|p}}$	$E^{SP}_{T_{\|p}}$	μ_1^*	μ_S	μ_{SP}	$E^S_{T_{\|p}}$	$E^{SP}_{T_{\|p}}$	μ_1^*	μ_S	μ_{SP}	$E^S_{T_{\|p}}$	$E^{SP}_{T_{\|p}}$
0.1	0.51	0.57	0.53	141.36	**	0.54	0.60	0.55	142.23	**	0.58	0.63	0.58	143.34	**
0.3	0.49	0.54	0.52	134.21	**	0.52	0.56	0.54	134.93	**	0.55	0.60	0.57	135.86	**
0.6	0.44	0.47	0.49	116.35	**	0.47	0.50	0.51	116.69	**	0.51	0.53	0.53	117.12	**

Note: '**' denotes estimator $T_{|p=1}$ does not perform better than T_{SP} in terms of efficiency.

Table 6: For p =2 and $\rho_{z_1z_2} = 0$.

ρ_{yx}		0.5					0.6					0.7				
ρ_{yz_1}	ρ_{yz_2}	μ_2^*	μ_S	μ_{SP}	$E^S_{T_{\|p}}$	$E^{SP}_{T_{\|p}}$	μ_2^*	μ_S	μ_{SP}	$E^S_{T_{\|p}}$	$E^{SP}_{T_{\|p}}$	μ_2^*	μ_S	μ_{SP}	$E^S_{T_{\|p}}$	$E^{SP}_{T_{\|p}}$
0.2	0.3	0.48	0.55	0.53	158.92	100.0	0.51	0.58	0.55	160.26	102.09	0.54	0.62	0.57	161.99	104.06
	0.4	0.47	0.55	0.53	168.30	105.91	0.50	0.58	0.55	169.91	108.23	0.53	0.62	0.57	171.98	110.48
	0.5	0.46	0.55	0.53	180.81	113.78	0.48	0.58	0.55	182.79	116.44	0.52	0.62	0.57	185.35	119.07
0.3	0.3	0.47	0.54	0.52	149.03	**	0.50	0.56	0.54	150.13	101.65	0.53	0.60	0.57	151.54	103.74
	0.4	0.46	0.54	0.52	156.62	104.62	0.49	0.56	0.54	158.10	107.05	0.53	060	0.57	159.79	109.38
	0.5	0.45	0.54	0.52	167.23	111.58	0.48	0.56	0.54	168.83	114.32	0.52	0.60	0.57	170.91	117.0
0.4	0.3	0.46	0.52	0.52	139.32	**	0.49	0.55	0.53	140.19	100.22	0.53	0.58	0.56	141.30	102.44
	0.4	0.45	0.52	0.52	145.60	102.27	0.48	0.55	0.53	146.63	104.83	0.52	0.58	0.56	147.95	107.27
	0.5	0.45	0.52	0.52	154.20	108.31	0.47	0.55	0.53	155.46	111.14	0.51	0.58	0.56	157.09	113.89

Note: '**' denotes estimator $T_{|p=2}$ does not perform better than T_{SP} in terms of efficiency.

Table 7: For p =2 and $\rho_{z_1z_2} \neq 0$.

ρ_{yx}		0.5					0.6					0.7				
$\rho_{yz_1} = \rho_{z_1z_2}$	ρ_{yz_2}	μ_2^*	μ_S	μ_{SP}	$E^S_{T_{\mid p}}$	$E^{SP}_{T_{\mid p}}$	μ_2^*	μ_S	μ_{SP}	$E^S_{T_{\mid p}}$	$E^{SP}_{T_{\mid p}}$	μ_2^*	μ_S	μ_{SP}	$E^S_{T_{\mid p}}$	$E^{SP}_{T_{\mid p}}$
0.2	0.4	0.47	0.55	0.53	162.28	104.01	0.50	0.58	0.58	166.81	106.26	0.53	0.62	0.57	168.77	108.41
	0.5	0.46	0.55	0.53	178.37	112.24	0.49	0.58	0.58	180.27	114.83	0.52	0.62	0.57	182.73	117.38
0.3	0.4	0.47	0.54	0.52	151.70	101.22	0.50	0.56	0.54	152.87	103.51	0.53	0.60	0.57	154.38	105.68
	0.5	0.46	0.54	0.52	162.40	108.36	0.48	0.56	0.54	163.86	110.95	0.52	0.60	0.57	165.75	113.47
0.4	0.4	0.46	0.52	0.52	138.66	**	0.49	0.55	0.53	139.51	**	0.53	0.58	0.56	140.60	101.94
	0.5	0.45	0.52	0.52	146.88	103.13	0.48	0.55	0.53	147.89	100.73	0.52	058	0.56	149.26	108.21

Note: '**' denotes estimator $T_{\mid p=2}$ does not perform better than T_{SP} in terms of efficiency.

Table 8: For p =3 and $\rho_{z_iz_j} = 0 \ \forall \ i \neq j = 1, 2, 3$.

ρ_{yx}	ρ_{yz_1}	ρ_{yz_2}	ρ_{yz_3}	μ_3^*	μ_S	μ_{SP}	$E^S_{T_{\mid p}}$	$E^S_{T_{\mid p}}$
0.6	0.5	0.6	0.7	0.45	0.52	0.52	152.12	114.09
0.7	0.5	0.6	0.7	0.49	0.56	0.55	153.67	117.30
0.7	0.4	0.6	0.5	0.50	0.58	0.56	169.65	122.99

Table 9: For p =3 and $\rho_{z_iz_j} \neq 0 \ \forall \ i \neq j = 1, 2, 3$.

ρ_{yx}	ρ_{yz_1}	ρ_{yz_2}	ρ_{yz_3}	$\rho_{z_1z_2}$	$\rho_{z_1z_3}$	$\rho_{z_2z_3}$	μ_3^*	μ_S	μ_{SP}	$E^S_{T_{\mid p}}$	$E^S_{T_{\mid p}}$
0.6	0.2	0.5	0.3	0.5	0.5	0.4	0.46	0.55	0.53	177.71	111.83
0.7	0.4	0.7	0.6	0.3	0.3	0.4	0.48	0.58	0.56	182.62	132.40
0.7	0.3	0.7	0.5	0.3	0.4	0.2	0.48	0.60	0.57	206.51	141.37

12. Mutual Comparison of the Estimators $T_{\mid p}$ (p=1, 2 and 3)

The performances of the estimator $T_{\mid p}$ (p=1, 2 & 3) have been elaborated empirically as well as through simulation studies in above sections and the results obtained are presented in Table 1 to Table 9. In this section the mutual comparison of the estimators for the cases when p=1, p = 2 and p=3 has been elaborated graphically and is presented in Figure 12.1 and 12.2.

Figure 12.1: Mutual Comparison of Proposed Estimator $T_{|p=1}$, $T_{|p=2}$ and $T_{|p=3}$ with respect to the estimator T_S for $\psi = 0.1$

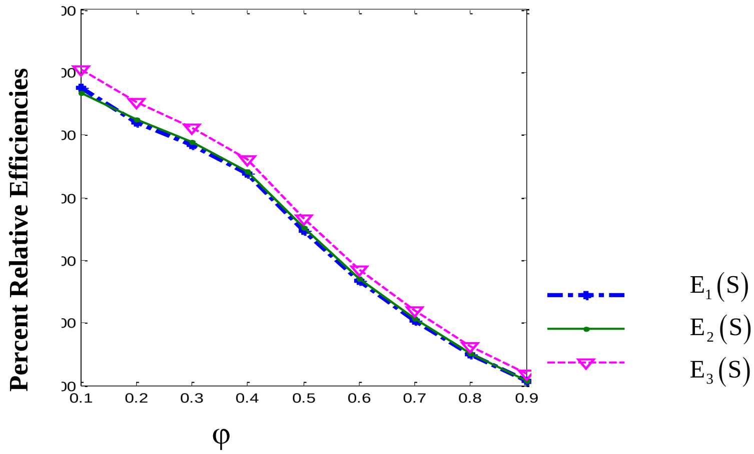
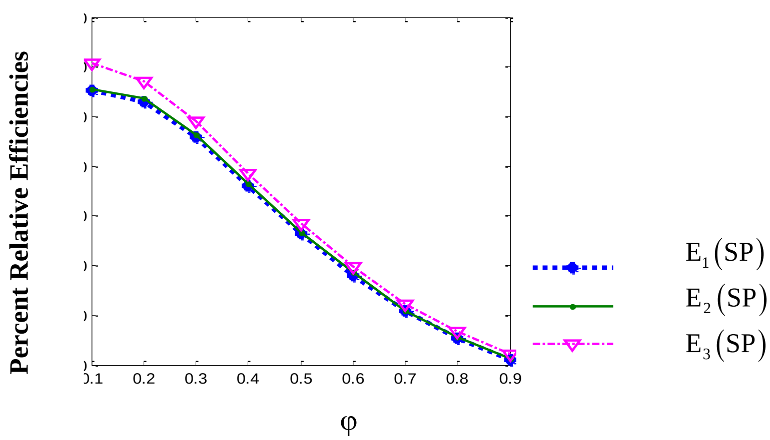


Figure 12.2: Mutual Comparison of Proposed Estimator $T_{|p=1}$, $T_{|p=2}$ and $T_{|p=3}$ with respect to the estimator T_{SP} for $\psi = 0.6$.



13. Rendition of Results

13.1 Results based on Empirical study for the considered population

1) It is clear from Table 1 that the optimum values of μ_1^*, μ_2^* and μ_3^* exist for the considered population and $\mu_3^* < \mu_2^* < \mu_1^* < \mu_{SP} < \mu_S$. This indicates that a smaller fresh sample is required when more number of auxiliary variables is used and this fraction is even lesser than the procedures given by Singh (2005) and Singh and Priyanka (2008a). Hence, total cost of survey is getting reduced.

2) The value of $E_{T|p=3} > E_{T|p=2} > E_{T|p=1}$, this justifies the fact that efficiency is highly increased when more numbers of auxiliary variates are taken into consideration and it also results peachy in terms of cost as it gets smaller on increasing the number of auxiliary variables, which also abide by Sukhatme et al.(1984).

3) In the Table 1, we see that the proposed estimator $T_{|p}$ is more efficient than the estimator T_S for all the considered values of p (i.e. p=1, 2 & 3) and $T_{|p}$ is better than the estimator T_{SP} except for p=1 as the number of auxiliary variables is increased, efficiency increases to a great extent, hence the estimator $T_{|p}$ is better than the estimator T_{SP} due to Singh and Priyanka (2008a) for p=2 onwards in terms of efficiency but in terms of cost $T_{|p}$ is better then T_S and T_{SP} for every value of p.

13.2 Results based on Simulation study

1) From simulation results in Table 2, Table 3 and Table 4 we observe that if less attention is given to φ (i.e. more attention is given to the estimator used at the first occasion) then the proposed estimators $T_{|p}$ $(p=1, 2 \text{ and } 3)$ are better than the estimator T_S and abide by the theory while keeping higher weights for ψ (i.e. more attention is given to the estimator used at the current occasion). This makes the proposed estimators $T_{|p}$ $(p=1, 2 \text{ and } 3)$ much more effective than the estimator T_S due to Singh (2005).

2) From Table 2, Table 3 and Table 4 it is vindicated that if less emphasis is supplied to φ (i.e. more attention is given to the estimator used at the first occasion) then the proposed estimators $T_{|p}$ $\left(\text{p=1, 2 and 3}\right)$ are better than the estimator T_{SP} and is in accordance with the theory while choosing a greater value for ξ (i.e. more attention is given to the estimator used at the current occasion). This makes the proposed estimators $T_{|p}$ $\left(\text{p=1, 2 and 3}\right)$ much more effective than the estimator T_{SP} due to Singh and Priyanka (2008a). As we keep on increasing the value of φ, the efficiency gets reduced for all choices of ξ.

13.3 Results extracted from General Scenario i.e. by considering different choices of correlation coefficients

1) In Table 5 we observe that for fixed value of correlation coefficient between the study variable at two occasions, if the correlation between the study and auxiliary variates is increased then the proposed estimator $T_{|p}$ for p=1 is efficient over the estimator T_S in terms of precision as well as cost but it is efficient over the estimator T_{SP} only in terms of cost. If the contribution of auxiliary variable increases the fraction of sample to be drawn on current occasion decreases.

2) From Table 6, Table 7, Table 8 and Table 9 we observe that, whether the auxiliary information utilized are mutually correlated or uncorrelated, the proposed estimator $T_{|p}$ for p=2 and p=3 is efficient over the estimators T_S and T_{SP} even for very low correlation between study variable and auxiliary variable, which is a positive point. The fraction of samples to be drawn afresh at current occasion is least for the proposed estimator than the estimators due to Singh (2005) and Singh and Priyanka (2008a) and it is getting more and more reduced as the contribution of auxiliary information increases.

12. Conclusion

The articulation of two structures (i.e. exponential ratio type and chain type ratio to exponential ratio type) is certainly beneficial as summed with multi-auxiliary information which are stable in nature, pronto and need not to be highly correlated to study variable over the two occasions. The empirical study for the considered population, simulation study and the study by taking different choices of correlation coefficient suggest that the proposed estimator is providing the lowest fraction of fresh sample drawn on the current occasion as compared to some very well-known estimators available in the literature for estimating population mean, resulting in lowering the total cost of the survey. Although the proposed estimator $T_{|p}$ is better than the estimator T_{SP} for p=1 in terms of cost only. Now as soon as we increase the number of auxiliary variables the proposed estimator comes out to be much better than the estimator T_{SP} in terms of both efficiency as well as cost. Hence, the proposed estimator may be recommended for its practical use by survey statisticians.

CHAPTER – 8*

New Approaches using Exponential Type Estimator with Cost Modelling for Population Mean on Successive Waves

* Following is the publication based on the work of this chapter:--

1. Priyanka, K. and Mittal, R. (2016): New Approaches using Exponential Type Estimator with Cost Modelling for Population Mean on Successive Waves. Statistics in Transition-new series, (Accepted for Publication).

New Approaches using Exponential Type Estimator with Cost Modelling for Population Mean on Successive Waves

1. Introduction

Real life facts always carry motleying natures which are time dependent. In such circumstances where facts change over a period of time, one time enquiry may not serve the purpose of investigation since statistics observed previously contain superannuated information which may not be good enough to be used after a long period of time. Therefore surveys are being designed sophistically to make sure no possible error gets a margin to escape at least in terms of design. For this longitudinal surveys are considered to be best since in longitudinal surveys, facts are investigated more than once i.e. over the successive waves, Also a frame is provided for reducing the cost of survey by a partial replacement of sample units in sampling over successive waves.

Jessen (1942) is considered to be the pioneer for observing dynamics of facts over a long period of time through partial replacement of sample units over successive waves. The approach of sampling over successive waves has been made more fruitful by using twisted and novel ways to consider extra information along with the study character. Enhanced literature has been made available by Patterson (1950), Narain (1953), Eckler (1955), Sen (1971, 1972, 1973), Gordon (1983), Singh et al. (1991), Arnab and Okafor (1992), Feng and Zou (1997), Biradar and Singh (2001), Singh and Singh (2001), Singh (2005), Singh and Priyanka (2006, 2007a, 2008a), Singh and Karna (2009), Singh and Prasad (2010), Singh et al. (2011), Singh et al. (2013), Bandyopadhyay and Singh (2014), Priyanka and Mittal (2014), Priyanka et al. (2015), Priyanka and Mittal (2015a, 2015b) etc.

It has been theoretically established that, in general, the linear regression estimator is more efficient than the ratio estimator except when the regression line y on x passes through the neighborhood of the origin; in this case the efficiencies of these estimators are almost equal. Also in many practical situations where the regression line does not pass through the neighborhood of the origin, in such cases the ratio estimator does not perform as good as the linear regression estimator. Here exponential type estimators play a vital role in increasing the precision of the estimates.Motivated with this idea we are aspired to develop unexampled estimators for estimating population mean over two successive waves applying the concept of exponential type ratio estimators. In this line of work, an attempt has been made to consider the dynamic nature of ancillary information also because as the time passes by, not only the nature of study variable changes but the nature of ancillary information also varies with respect to time in many real life phenomenon where time lag is very large between two successive waves. For example, in a social survey one may desire to observe the number of females human trafficked from a particular region, the number of girls child birth may serve as ancillary information which is completely dynamic over a period of 8 years of time span. Similarly in a medicinal survey one may be interested to record the number of survivors from a cancerous disease, here the number of successfully tested drugs for the disease may not sustain to be stable over a period of 10 or 20 years or in an economic survey the government may like to record the labor force, the total number of graduates in country may serve as an ancillary character to the study character but it surely inherent dynamic nature over a period of 5 or 10 years.So such situations cannot be tackled considering the ancillary character to be stable since doing so will affect the final findings of the survey. Keeping the drawback of such flaws in consideration, this work deals in bringing modern approaches for estimating population mean over two successive waves. Four estimators have been habituated with a fine amalgamation of completely known dynamic ancillary information with exponential ratio type estimators. Their properties including optimum rotation rate and a model for optimum total cost have been proposed and discussed. Also detailed empirical illustrations have been done by doing a comparison of proposed estimators with well-known existing estimators in the literature of successive sampling. Simulation algorithms have been devised to make the proposed estimators work in practical environment efficiently.

2. Survey Design and Analysis

2.1. Sample Structure and Notations

Let $U = (U_1, U_2, \ldots, U_N)$ be the finite population of N units, which has been sampled over two successive waves. It is assumed that size of the population remains unchanged but values of units change over two successive waves. The character under study be denoted by x (y) on the first (second) waves respectively. It is assumed that information on an ancillary variable $z_1(z_2)$ dynamic in nature over the successive waves with completely known population mean $\bar{Z}_1(\bar{Z}_2)$ is readily available on both the successive waves and positively correlated to x and y respectively. Simple random sample (without replacement) of n units is taken at the first wave. A random subsample of $m = n\lambda$ units is retained for use at the second wave. Now at the current wave a simple random sample (without replacement) of u= (n-m) = nμ units is drawn afresh from the remaining (N-n) units of the population so that the sample size on the second wave remains the same. Let μ and $\lambda(\mu + \lambda = 1)$ are the fractions of fresh and matched samples respectively at the second (current) successive wave. The following notations are considered here after:

$\bar{X}, \bar{Y}, \bar{Z}_1, \bar{Z}_2$: Population means of the variables x, y, z_1 and z_2 respectively.

$\bar{y}_u, \bar{z}_u, \bar{x}_m, \bar{y}_m, \bar{z}_1(m), \bar{z}_2(m), \bar{x}_n, \bar{z}_1(n), \bar{z}_2(n)$: Sample mean of respective variate based on the sample sizes shown in suffice.

$\rho_{yx}, \rho_{xz_1}, \rho_{xz_2}, \rho_{yz_1}, \rho_{yz_2}, \rho_{z_1z_2}$: Correlation coefficient between the variables shown in suffices.

$S_x^2, S_y^2, S_{z_1}^2, S_{z_2}^2$: Population mean squared of variables x, y, z_1 and z_2 respectively.

2.2 Design of the Proposed Estimators $\check{\mathcal{T}}_{ij}\,(i, j=1, 2)$

For estimating the population mean $\bar{Y}$ at the current wave, two sets of estimators have been proposed. The first set of estimators is based on sample of size u drawn afresh at current occasion and is given by

$$\check{\mathcal{T}}_u = \{t_{1u}, t_{2u}\}, \tag{1}$$

where $$t_{1u} = \bar{y}_u \left(\frac{\bar{Z}_2}{\bar{z}_2(u)} \right) \tag{2}$$

$$t_{2u} = \bar{y}_u \ \exp\left(\frac{\bar{Z}_2 - \bar{z}_2(u)}{\bar{Z}_2 + \bar{z}_2(u)} \right) \tag{3}$$

The second set of estimators is based on sample of size m common to both occasion and is

$$\check{\mathcal{T}}_m = \{t_{1m}, t_{2m}\}, \tag{4}$$

where $$t_{1m} = \bar{y}_m \left(\frac{\bar{x}_n}{\bar{x}_m} \right) \exp\left(\frac{\bar{Z}_2 - \bar{z}_2(m)}{\bar{Z}_2 + \bar{z}_2(m)} \right) \tag{5}$$

$$t_{2m} = \bar{y}_m^* \left(\frac{\bar{x}_n^*}{\bar{x}_m^*} \right) \tag{6}$$

where $\bar{y}_m^* = \bar{y}_m \ \exp\left(\frac{\bar{Z}_2 - \bar{z}_2(m)}{\bar{Z}_2 + \bar{z}_2(m)} \right)$, $\bar{x}_m^* = \bar{x}_m \ \exp\left(\frac{\bar{Z}_1 - \bar{z}_1(m)}{\bar{Z}_1 + \bar{z}_1(m)} \right)$ and $\bar{x}_n^* = \bar{x}_n \ \exp\left(\frac{\bar{Z}_1 - \bar{z}_1(n)}{\bar{Z}_1 + \bar{z}_1(n)} \right)$

.

Hence, considering the convex combination of the two sets $\check{\mathcal{T}}_u$ and $\check{\mathcal{T}}_m$, we have the final estimators of the population mean $\bar{Y}$ on the current occasion as

$$\check{\mathcal{T}}_{ij} = \varpi_{ij} t_{iu} + (1-\varpi_{ij})\, t_{jm}; (i, j=1, 2) \tag{7}$$

where $\left(t_{iu}, t_{jm}\right) \in \check{\mathcal{T}}_u \times \check{\mathcal{T}}_m$ and ϖ_{ij} are suitably chosen weights so as to minimize the mean squared error of the estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$.

2.3. Analysis of the estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$

2.3.1. Bias and Mean Squared Errors of the Proposed Estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$

The properties of the proposed estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$ are derived under the following large sample approximations

$\bar{y}_u = \bar{Y}\left(1 + e_0\right)$, $\bar{y}_m = \bar{Y}\left(1 + e_1\right)$, $\bar{x}_m = \bar{X}\left(1 + e_2\right)$, $\bar{x}_n = \bar{X}\left(1 + e_3\right)$, $\bar{z}_2\left(u\right) = \bar{Z}_2\left(1 + e_4\right)$,
$\bar{z}_2\left(m\right) = \bar{Z}_2\left(1 + e_5\right)$, $\bar{z}_1\left(m\right) = \bar{Z}_1\left(1 + e_6\right)$ and $\bar{z}_1\left(n\right) = \bar{Z}_1\left(1 + e_7\right)$ such that $|e_i| < 1 \;\forall\; i = 0,...,7$.

The estimators belonging to the sets $\check{\mathcal{T}}_u$ and $\check{\mathcal{T}}_m\left(i, j=1, 2\right)$ are ratio, exponential ratio, ratio to exponential ratio and chain type ratio to exponential ratio type in nature respectively. Hence they are biased for population mean $\bar{Y}$. Therefore, the final estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$ defined in equation (7) are also biased estimators of $\bar{Y}$. The bias $B\left(.\right)$ and mean squared errors $M\left(.\right)$ of the proposed estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$ are obtained up to first order of approximations and thus we have following theorems:

Theorem 2.3.1. Bias of the estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$ to the first order of approximations are obtained as

$$B\left(\check{\mathcal{T}}_{ij}\right) = \varpi_{ij}\, B\left(t_{iu}\right) + \left(1 - \varpi_{ij}\right) B\left(t_{jm}\right); \; (i, j=1,2), \tag{8}$$

where $$B\left(t_{1u}\right) = \frac{1}{u}\bar{Y}\left(\frac{C_{0002}}{\bar{Z}_2^2} - \frac{C_{0101}}{\bar{Y}\,\bar{Z}_2}\right), \tag{9}$$

$$B\left(t_{2u}\right) = \frac{1}{u}\bar{Y}\left(\frac{3}{8}\frac{C_{0002}}{\bar{Z}_2^2} - \frac{1}{2}\frac{C_{0101}}{\bar{Y}\,\bar{Z}_2}\right), \tag{10}$$

$$B\left(t_{1m}\right) = \bar{Y}\left(\frac{1}{m}\left(\frac{C_{2000}}{\bar{X}^2} + \frac{3}{8}\frac{C_{0002}}{\bar{Z}_2^2} - \frac{C_{1100}}{\bar{X}\bar{Y}} - \frac{1}{2}\frac{C_{0101}}{\bar{Y}\bar{Z}_2} + \frac{1}{2}\frac{C_{1001}}{\bar{X}\bar{Z}_2}\right) + \frac{1}{n}\left(\frac{C_{1100}}{\bar{X}\bar{Y}} - \frac{C_{2000}}{\bar{X}^2} - \frac{1}{2}\frac{C_{1001}}{\bar{X}\bar{Z}_2}\right)\right), \qquad (11)$$

and

$$B\left(t_{2m}\right) = \bar{Y}\left(\frac{1}{m}\left(\frac{C_{2000}}{\bar{X}^2} - \frac{1}{8}\frac{C_{0020}}{\bar{Z}_1^2} + \frac{3}{8}\frac{C_{0002}}{\bar{Z}_2^2} - \frac{C_{1100}}{\bar{X}\bar{Y}} - \frac{1}{2}\frac{C_{1010}}{\bar{X}\bar{Z}_1} + \frac{1}{2}\frac{C_{1001}}{\bar{X}\bar{Z}_2} + \frac{1}{2}\frac{C_{0110}}{\bar{Y}\bar{Z}_1} - \frac{1}{2}\frac{C_{0101}}{\bar{Y}\bar{Z}_2} - \frac{1}{4}\frac{C_{0011}}{\bar{Z}_1\bar{Z}_2}\right)\right.$$
$$\left. + \frac{1}{n}\left(\frac{1}{8}\frac{C_{002}}{\bar{Z}_1^2} - \frac{C_{2000}}{\bar{X}^2} + \frac{C_{1100}}{\bar{X}\bar{Y}} + \frac{1}{2}\frac{C_{1010}}{\bar{X}\bar{Z}_1} - \frac{1}{2}\frac{C_{1001}}{\bar{X}\bar{Z}_2} - \frac{1}{2}\frac{C_{0101}}{\bar{Y}\bar{Z}_2} + \frac{1}{4}\frac{C_{0011}}{\bar{Z}_1\bar{Z}_2}\right)\right) \qquad (12)$$

where $C_{rstq} = E\left[\left(x_i - \bar{X}\right)^r\left(y_i - \bar{Y}\right)^s\left(z_{1i} - \bar{Z}_1\right)^t\left(z_{2i} - \bar{Z}_2\right)^q\right]$; $\left(r, s, t, q\right) \geq 0$.

Theorem 2.3.2.Mean squared errors of the estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$ to the first order of approximations are obtained as

$$M\left(\check{\mathcal{T}}_{ij}\right) = \varpi_{ij}^2\, M\left(t_{iu}\right) + \left(1 - \varpi_{ij}\right)^2 M\left(t_{jm}\right) + 2\,\varpi_{ij}\left(1 - \varpi_{ij}\right)Cov\left(t_{iu}, t_{jm}\right);(i, j=1,2) \qquad (13)$$

Where $M\left(t_{1u}\right) = \frac{1}{u}\, A_1\, S_y^2$ (14)

$$M\left(t_{2u}\right) = \frac{1}{u}\, A_2\, S_y^2 \qquad (15)$$

$$M\left(t_{1m}\right) = \left(\frac{1}{m}\, A_3 + \frac{1}{n}\, A_4\right) S_y^2 \qquad (16)$$

$$M\left(t_{2m}\right) = \left(\frac{1}{m}\, A_5 + \frac{1}{n}\, A_6\right) S_y^2 \qquad (17)$$

$Cov\left(t_{iu}, t_{jm}\right)=0$, $A_1 = 2\left(1 - \rho_{yz_2}\right)$, $A_2 = \frac{5}{4} - \rho_{yz_2}$, $A_3 = \frac{9}{4} - 2\rho_{yx} - \rho_{yz_2} + \rho_{xz_2}$, $A_4 = 2\rho_{yx} - \rho_{xz_2} - 1$,

$A_5 = \frac{5}{2} - 2\rho_{yx} - \rho_{xz_1} + \rho_{xz_2} + \rho_{yz_1} - \rho_{yz_2} - \frac{1}{2}\rho_{z_1z_2}$ and $A_6 = 2\rho_{yx} + \rho_{xz_1} - \rho_{xz_2} - \rho_{yz_1} + \frac{1}{2}\rho_{z_1z_2} - \frac{5}{4}$.

2.3.2. Minimum Mean Squared Errors of the Proposed Estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$

Since the mean squared errors of the estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$ given in equation (13) are the functions of unknown constants $\varpi_{ij}\left(i, j = 1, 2\right)$, therefore, they are minimized with respect to ϖ_{ij} and subsequently the optimum values of $\varpi_{ij}\left(i, j = 1, 2\right)$ and $M\left(\check{\mathcal{T}}_{ij}\right)_{opt.}\left(i, j=1, 2\right)$ are obtained as

$$\varpi_{ij_{opt.}} = \frac{M\left(t_{jm}\right)}{M\left(t_{iu}\right) + M\left(t_{jm}\right)};\left(i, j = 1, 2\right) \tag{18}$$

$$M\left(\check{\mathcal{T}}_{ij}\right)_{opt.} = \frac{M\left(t_{iu}\right) . M\left(t_{jm}\right)}{M\left(t_{iu}\right) + M\left(t_{jm}\right)}; \ \left(i, j = 1, 2\right) \tag{19}$$

Further, substituting the values of the mean squared errors of the estimators defined in equations (14)-(17) in equation (18)-(19), the simplified values of $\varpi_{ij_{opt.}}$ and $M\left(\check{\mathcal{T}}_{ij}\right)_{opt.}$ are obtained as

$$\varpi_{11_{opt.}} = \frac{\mu_{11}\left[\mu_{11} A_4 - \left(A_3 + A_4\right)\right]}{\left[\mu_{11}^2 A_4 - \mu_{11}\left(A_3 + A_4 - A_1\right) - A_1\right]} \tag{20}$$

$$\varpi_{12_{opt.}} = \frac{\mu_{12}\left[\mu_{12} A_6 - \left(A_5 + A_6\right)\right]}{\left[\mu_{12}^2 A_6 - \mu_{12}\left(A_5 + A_6 - A_1\right) - A_1\right]} \tag{21}$$

$$\varpi_{21_{opt.}} = \frac{\mu_{21}\left[\mu_{21} A_4 - \left(A_3 + A_4\right)\right]}{\left[\mu_{21}^2 A_4 - \mu_{21}\left(A_3 + A_4 - A_2\right) - A_2\right]} \tag{22}$$

$$\varpi_{22_{opt.}} = \frac{\mu_{22}\left[\mu_{22} A_6 - \left(A_5 + A_6\right)\right]}{\left[\mu_{22}^2 A_6 - \mu_{22}\left(A_5 + A_6 - A_2\right) - A_2\right]} \tag{23}$$

$$M\left(\check{\mathcal{T}}_{11}\right)_{opt.} = \frac{1}{n}\frac{\left[\mu_{11} B_1 - B_2\right]S_y^2}{\left[\mu_{11}^2 A_4 - \mu_{11} B_3 - A_1\right]} \tag{24}$$

$$M\left(\check{\mathcal{T}}_{12}\right)_{opt.} = \frac{1}{n}\frac{\left[\mu_{12}\ B_4\ -B_5\right]S_y^2}{\left[\mu_{12}^2\ A_6\ -\mu_{12}\ B_6\ -A_1\right]} \tag{25}$$

$$M\left(\check{\mathcal{T}}_{21}\right)_{opt.} = \frac{1}{n}\frac{\left[\mu_{21}\ B_7\ -B_8\right]S_y^2}{\left[\mu_{21}^2\ A_4\ -\mu_{21}\ B_9\ -A_2\right]} \tag{26}$$

$$M\left(\check{\mathcal{T}}_{22}\right)_{opt.} = \frac{1}{n}\frac{\left[\mu_{22}\ B_{10}\ -B_{11}\right]S_y^2}{\left[\mu_{22}^2\ A_6\ -\mu_{22}\ B_{12}\ -A_2\right]} \tag{27}$$

where $B_1 = A_1A_4$, $B_2 = A_1A_3 + A_1A_4$, $B_3 = A_3 + A_4 - A_1$, $B_4 = A_1A_6$, $B_5 = A_1A_5 + A_1A_6$, $B_6 = A_5 + A_6 - A_1$, $B_7 = A_2A_4$, $B_8 = A_2A_3 + A_2A_4$, $B_9 = A_3 + A_4 - A_2$, $B_{10} = A_2A_6$

$B_{11} = A_2A_5 + A_2A_6$, $B_{12} = A_5 + A_6 - A_2$ and $\mu_{ij}\left(i, j = 1, 2\right)$ are the fractions of the sample drawn afresh at the current(second) wave.

2.3.3. Optimum Rotation Rate for the Estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$

Since the mean squared errors of the proposed estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$ are the function of the $\mu_{ij}\left(i, j = 1, 2\right)$, hence to estimate population mean $\bar{Y}$ with maximum precision and minimum cost, an amicable fraction of sample drawn afresh is required at the current wave. For this the mean squared errors of the estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$ in equations (24) - (27) have been minimized with respect to $\mu_{ij}\left(i, j = 1, 2\right)$. Hence an optimum rotation rate has been obtained for each of the estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$ and given as :

$$\mu_{11} = \frac{C_2 \pm \sqrt{C_2^2 - C_1\ C_3}}{C_1} \tag{28}$$

$$\mu_{12} = \frac{C_5 \pm \sqrt{C_5^2 - C_4\ C_6}}{C_4} \tag{29}$$

$$\mu_{21} = \frac{C_8 \pm \sqrt{C_8^2 - C_7\ C_9}}{C_7} \tag{30}$$

$$\mu_{22} = \frac{C_{11} \pm \sqrt{C_{11}^2 - C_{10}\ C_{12}}}{C_{10}} \tag{31}$$

where

$C_1 = A_4B_1,\ C_2 = A_4B_2,\ C_3 = A_1B_1 + B_2B_3,\ C_4 = A_6B_4,\ C_5 = A_6B_5,\ C_6 = A_1B_4 + B_5B_6$

$C_7 = A_4B_7,\ C_8 = A_4B_8,\ C_9 = A_2B_7 + B_8B_9,\ C_{10} = A_6B_{10},\ C_{11} = A_6B_{11}$ and $C_{12} = A_2B_{10} + B_{11}B_{12}$.

The real values of $\mu_{ij}(i, j = 1, 2)$ exist, iff $C_2^2 - C_1\ C_3 \ge 0,\ C_5^2 - C_4\ C_6 \ge 0,\ C_8^2 - C_7\ C_9 \ge 0,$ and $C_{11}^2 - C_{10}\ C_{12} \ge 0$ respectively. For any situation, which satisfies these conditions, two real values of $\mu_{ij}(i, j = 1, 2)$ may be possible , hence to choose a value of $\mu_{ij}(i, j = 1, 2)$, it should be taken care of that $0 \le \mu_{ij} \le 1$, all other values of $\mu_{ij}(i, j = 1, 2)$ are inadmissible. If both the real values of $\mu_{ij}(i, j = 1, 2)$ are admissible, the lowest one will be the best choice as it reduces the total cost of the survey. Substituting the admissible value of μ_{ij} say $\mu_{ij}^{(0)}(i, j = 1, 2)$ from equation (28) - (31) in equation (24) - (27) respectively , we get the optimum values of the mean squared errors of the estimators $\check{\mathcal{T}}_{ij}(i, j = 1, 2)$ with respect to ϖ_{ij} as well as $\mu_{ij}(i, j = 1, 2)$ which are given as

$$M\left(\check{\mathcal{T}}_{11}\right)^*_{opt.} = \frac{\left[\mu_{11}^{(0)}\ B_1 - B_2\right]\ S_y^2}{n\left[\mu_{11}^{(0)\,2}\ A_4 - \mu_{11}^{(0)}\ B_3 - A_1\right]} \tag{32}$$

$$M\left(\check{\mathcal{T}}_{12}\right)^*_{opt.} = \frac{\left[\mu_{12}^{(0)}\ B_4 - B_5\right]\ S_y^2}{n\left[\mu_{12}^{(0)\,2}\ A_6 - \mu_{12}^{(0)}\ B_6 - A_1\right]} \tag{33}$$

$$M\left(\check{\mathcal{T}}_{21}\right)^*_{opt.} = \frac{\left[\mu_{21}^{(0)}\ B_7 - B_8\right]\ S_y^2}{n\left[\mu_{21}^{(0)\,2}\ A_4 - \mu_{21}^{(0)}\ B_9 - A_2\right]} \tag{34}$$

$$M\left(\check{\mathcal{T}}_{22}\right)^*_{opt.} = \frac{\left[\mu_{22}^{(0)}\ B_{10} - B_{11}\right]\ S_y^2}{n\left[\mu_{22}^{(0)\,2}\ A_6 - \mu_{22}^{(0)}\ B_{12} - A_2\right]} \tag{35}$$

3. Cost Analysis

The total cost of survey design and analysis over two successive waves is modelled as:

$$C_T = nc_f + mc_r + uc_s \tag{36}$$

where c_f : The average per unit cost of investigating and processing data at previous (first) wave,

c_r : The average per unit cost of investigating and processing retained data at current wave,

c_s : The average per unit cost of investigating and processing freshly drawn data at current wave.

Remark 3.1: $c_f < c_r < c_s$, when there is a large gap between two successive waves, the cost of investigating a single unit involved in the survey sample should be greater than before (at previous occasion) since as time passes by different commodities (software) and services (human resources, daily wages and conveyance) become expensive so the cost incurring at second occasion increases in a considerable amount. Also the average cost of investigating a retained unit from previous wave should be lesser than investigating a freshly drawn sample unit since survey investigator as well as respondent has some experiences from the previous wave.

Theorem 3.1.1: The optimum total cost for the proposed estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$ is derived as

$$C_T\left(\check{\mathcal{T}}_{ij}\right)_{opt.} = n\left\{c_f + c_s + \left(1 - \mu_{ij}^{(0)}\right)\left(c_r - c_s\right)\right\} \forall\ i, j=1, 2 \tag{37}$$

Remark 3.2: The optimum total costs obtained in equation (37) are dependent on the value of n. Therefore, if a suitable guess of n is available, it can be used for obtaining optimum total cost of the survey by above equation. However, in the absence of suitable guess of n, it may be estimated by following Cochran (1977).

4. Efficiency Comparison

To evaluate the performance of the proposed estimators, the estimators $\check{\mathcal{T}}_{ij}\,(i, j=1, 2)$ at optimum conditions, are compared with the sample mean estimator $\bar{y}_n$, when there is no matching from previous wave and the estimator $\hat{\hat{Y}}$ due to Jessen (1942) given by

Since the sample mean estimator $\bar{y}_n$ is unbiased for population mean, so variance of the estimator $\bar{y}_n$ is given by

$$\hat{\hat{Y}} = \psi\,\bar{y}_u + (1 - \psi)\,\bar{y}_m^{'}, \tag{38}$$

where $\bar{y}_m^{'} = \bar{y}_m + \beta_{yx}\left(\bar{x}_n - \bar{x}_m\right)$, β_{yx} is the population regression coefficient of y on x and ψ is an unknown constant to be determined so as to minimize the mean squared error of the estimator $\hat{\hat{Y}}$. The estimators $\bar{y}_n$ and $\hat{\hat{Y}}$ are unbiased for population mean and variance of the estimators $\bar{y}_n$ and $\hat{\hat{Y}}$ at optimum conditions are given as

$$V\left(\bar{y}_n\right) = \frac{1}{n}S_y^2, \tag{39}$$

$$V\left(\hat{\hat{Y}}\right)^{*}_{opt.} = \left(\frac{1}{2}\left(1 + \sqrt{1 - \rho_{yx}^2}\right)\right)\frac{S_y^2}{n}, \tag{40}$$

and the fraction of sample to be drawn afresh for the estimator $\hat{\hat{Y}}$

$$\mu_J = \frac{1}{1 + \sqrt{1 - \rho_{yx}^2}} \tag{41}$$

The percent relative efficiencies $E_{ij}(M)$ and $E_{ij}(J)$ of the estimator $\check{\mathcal{T}}_{ij}\,(i, j=1, 2)$ (under optimum conditions) with respect to $\bar{y}_n$ and $\hat{\hat{Y}}$ are respectively given by

$$E_{ij}(M) = \frac{V(\bar{y}_n)}{M(\check{\mathcal{T}}_{ij})^*_{opt.}} \times 100 \text{ and } E_{ij}(J) = \frac{V\left(\hat{\hat{Y}}\right)^*_{opt.}}{M(\check{\mathcal{T}}_{ij})^*_{opt.}} \times 100 \, (i, j=1, 2). \quad (42)$$

5. Numerical Illustrations and Simulation

5.1. Generalization of empirical study

A generalized study has been done to show the impact of motleying ancillary information in enhancing the performance of the proposed estimators $\check{\mathcal{T}}_{ij}\,(i, j=1, 2)$. To elaborate the scenario, various choices of correlation coefficients of study and auxiliary variables have been considered. The results obtained have been shown in Table 1.

Table 1: Generalized empirical results while the proposed estimators $\check{\mathcal{T}}_{ij}\,(i, j=1, 2)$ have been compared to the estimators $\bar{y}_n$ and $\hat{\hat{Y}}$ for $\rho_{yz_1} = \rho_{yz_2} = \rho_1$ and $\rho_{xz_1} = \rho_{xz_2} = \rho_2$.

$\rho_{z_1z_2} = \rho_{yx} = 0.5$														
ρ_2	ρ_1	μ_J	$\mu_{11}^{(0)}$	$\mu_{12}^{(0)}$	$\mu_{21}^{(0)}$	$\mu_{22}^{(0)}$	$E_{11}(M)$	$E_{12}(M)$	$E_{21}(M)$	$E_{22}(M)$	$E_{11}(j)$	$E_{12}(j)$	$E_{21}(j)$	$E_{22}(j)$
0.4	**0.6**	0.53	0.66	0.58	0.44	0.41	119.69	114.58	135.48	128.91	111.67	106.90	126.41	120.28
	0.7	0.53	0.51	0.45	0.43	0.39	149.98	138.58	157.12	145.01	139.93	129.30	146.60	135.30
	0.8	0.53	0.33	0.32	0.42	0.37	197.61	176.31	187.18	166.66	184.38	164.50	174.64	155.50
0.5	**0.6**	0.53	0.61	0.58	0.42	0.41	117.08	114.58	132.04	128.91	109.24	106.90	123.20	120.28
	0.7	0.53	0.48	0.45	0.41	0.39	145.81	138.58	152.67	145.01	136.04	129.30	142.45	135.30
	0.8	0.53	0.33	0.32	0.40	0.37	191.44	176.31	181.18	166.66	178.61	164.50	169.04	155.50
0.6	**0.6**	0.53	0.58	0.58	0.41	0.41	114.58	114.58	128.91	128.91	106.90	106.99	120.28	120.28
	0.7	0.53	0.46	0.45	0.40	0.39	142.03	138.58	148.66	145.01	132.52	129.30	138.70	135.30
	0.8	0.53	0.33	0.32	0.39	0.37	185.89	176.31	175.84	166.66	173.44	164.50	164.06	155.50
0.7	**0.6**	0.53	0.55	0.58	0.40	0.41	112.22	114.58	126.04	128.91	104.70	106.90	117.60	120.28
	0.7	0.53	0.45	0.45	0.39	0.39	138.58	138.58	145.01	145.01	129.30	129.30	135.30	135.30
	0.8	0.53	0.32	0.32	0.38	0.37	180.88	176.31	171.03	166.66	168.76	164.50	159.57	155.50
$\rho_{z_1z_2} = \rho_{yx} = 0.6$														
ρ_2	ρ_1	μ_J	$\mu_{11}^{(0)}$	$\mu_{12}^{(0)}$	$\mu_{21}^{(0)}$	$\mu_{22}^{(0)}$	$E_{11}(M)$	$E_{12}(M)$	$E_{21}(M)$	$E_{22}(M)$	$E_{11}(j)$	$E_{12}(j)$	$E_{21}(j)$	$E_{22}(j)$
0.4	**0.6**	0.55	0.87	0.69	0.46	0.44	124.52	121.01	143.54	137.34	112.07	108.91	129.18	123.60
	0.7	0.55	0.60	0.49	0.46	0.42	159.70	147.84	167.74	154.84	143.73	133.06	150.97	139.35
	0.8	0.55	0.29	0.33	0.45	0.40	212.25	188.59	201.85	178.44	191.02	169.73	181.58	160.54
0.5	**0.6**	0.55	0.73	0.69	0.45	0.44	122.31	121.01	139.24	137.34	110.08	108.91	125.36	123.60
	0.7	0.55	0.54	0.49	0.44	0.42	154.60	147.84	162.24	154.84	139.14	133.06	145.89	139.35
	0.8	0.55	0.32	0.33	0.43	0.40	204.53	188.59	193.99	178.44	184.08	169.73	174.59	160.59
0.6	**0.6**	0.55	0.66	0.69	0.44	0.44	119.69	121.01	135.48	137.34	107.72	108.91	121.94	123.60
	0.7	0.55	0.51	0.49	0.43	0.42	149.98	147.84	157.12	154.84	134.98	133.06	141.41	139.35
	0.8	0.55	0.33	0.33	0.42	0.40	197.61	188.59	187.18	178.44	177.85	169.73	168.46	160.54
0.7	**0.6**	0.55	0.61	0.69	0.42	0.44	117.08	121.01	132.04	137.34	105.37	108.91	118.84	123.60
	0.7	0.55	0.48	0.49	0.41	0.42	15.81	147.84	152.67	154.84	131.23	133.06	137.41	139.35
	0.8	0.55	0.33	0.33	0.40	0.40	191.44	188.59	181.18	178.44	172.29	169.73	163.06	160.59

Note: The values for $\mu_j, \mu_{11}^{(0)}, \mu_{12}^{(0)}, \mu_{21}^{(0)}$ and $\mu_{22}^{(0)}$ **have been rounded off up to two places of decimal for presentation.**

5.2. Generalized study based on correlation coefficients and optimum total cost model

To validate the proposed cost model, a hypothetical survey design has been assumed in which various choices of correlation coefficient and different input costs have been considered over two successive waves

Table 2: Optimum total cost of the survey design at the current wave of the proposed estimators $\check{\mathcal{T}}_{ij}$ (i, j=1, 2)

$\rho_{yx}=0.5$, n=30, c_f = ₹ 50.00, c_r = ₹ 75.00 and c_s = ₹ 80.00											
ρ_2	ρ_1	μ_J	$\mu_{11}^{(0)}$	$\mu_{12}^{(0)}$	$\mu_{21}^{(0)}$	$\mu_{22}^{(0)}$	$C_T(J)$	$C_T(11)$	$C_T(12)$	$C_T(21)$	$C_T(22)$
0.5	**0.6**	0.53	0.61	0.58	0.42	0.41	3830.4	3842.7	3837.5	3814.4	3812.8
	0.7	0.53	0.48	0.45	0.41	0.39	3830.4	3823.1	3817.7	3813.0	3809.8
	0.8	0.53	0.33	0.32	0.40	0.37	3830.4	3799.9	3798.2	3811.2	3806.3
0.6	**0.6**	0.53	0.58	0.58	0.41	0.41	3830.4	3837.5	3837.5	3812.8	3812.8
	0.7	0.53	0.46	0.45	0.40	0.39	3830.4	3820.2	3817.7	3811.3	3809.8
	0.8	0.53	0.33	0.32	0.39	0.37	3830.4	3799.5	3798.2	3809.3	3806.3
0.7	**0.6**	0.53	0.55	0.58	0.40	0.41	3830.4	3833.4	3837.5	3811.4	3812.8
	0.7	0.53	0.45	0.45	0.39	0.39	3830.4	3817.7	3817.7	38.9.8	3809.8
	0.8	0.53	0.32	0.32	0.38	0.37	3830.4	3798.9	3798.2	3807.7	3806.3
$\rho_{yx}=0.6$, n=30, c_f = ₹ 50.00, c_r = ₹ 75.00 and c_s = ₹ 80.00											
ρ_2	ρ_1	μ_J	$\mu_{11}^{(0)}$	$\mu_{12}^{(0)}$	$\mu_{21}^{(0)}$	$\mu_{22}^{(0)}$	$C_T(J)$	$C_T(11)$	$C_T(12)$	$C_T(21)$	$C_T(22)$
0.5	**0.6**	0.55	0.73	0.69	0.45	0.44	3833.3	3860.9	3854.8	3817.9	3817.0
	0.7	0.55	0.54	0.49	0.44	0.42	3833.3	3832.1	3824.9	3816.9	3813.9
	0.8	0.55	0.32	0.33	0.43	0.40	3833.3	3798.9	3799.8	3815.5	3810.2
0.6	**0.6**	0.55	0.66	0.69	0.44	0.44	3833.3	3849.9	3854.4	3816.1	3817.0
	0.7	0.55	0.51	0.49	0.43	0.42	3833.3	3826.9	3824.9	3814.8	3813.9
	0.8	0.55	0.33	0.33	0.42	0.40	3833.3	3833.3	3799.8	3813.2	3810.2
0.7	**0.6**	0.55	0.61	0.69	0.42	0.44	3833.3	3842.7	3854.8	3814.4	3817.0
	0.7	0.55	0.48	0.49	0.41	0.42	3833.3	3823.1	3824.9	3813.0	3813.9
	0.8	0.55	0.33	0.33	0.40	0.40	3833.3	3833.3	3799.9	3811.2	3810.2

Table 3: Optimum total cost of the survey design at the current wave of the proposed estimators $\check{\mathcal{T}}_{ij}$ (i, j=1, 2)

ρ_{yx} =0.5, n=40, c_f = ₹ 50.00, c_r = ₹ 75.00 and c_s = ₹ 80.00											
ρ_2	ρ_1	μ_J	$\mu_{11}^{(0)}$	$\mu_{12}^{(0)}$	$\mu_{21}^{(0)}$	$\mu_{22}^{(0)}$	$C_T(J)$	$C_T(11)$	$C_T(12)$	$C_T(21)$	$C_T(22)$
0.5	**0.6**	0.53	0.61	0.58	0.42	0.41	5107.2	5123.7	5116.7	5085.8	5083.8
	0.7	0.53	0.48	0.45	0.41	0.39	5107.2	5097.5	5090.3	5084.0	5079.8
	0.8	0.53	0.33	0.32	0.40	0.37	5107.2	5066.6	5064.3	5081.5	5075.0
0.6	**0.6**	0.53	0.58	0.58	0.41	0.41	5107.2	5116.7	5116.7	5083.8	5083.8
	0.7	0.53	0.46	0.45	0.40	0.39	5107.2	5093.6	5090.3	5081.8	5079.8
	0.8	0.53	0.33	0.32	0.39	0.37	5107.2	5066.0	5064.3	5079.1	5075.0
0.7	**0.6**	0.53	0.55	0.58	0.40	0.41	5107.2	5111.2	5116.7	5081.9	5083.8
	0.7	0.53	0.45	0.45	0.39	0.39	5107.2	5090.3	5090.3	5079.8	5079.8
	0.8	0.53	0.32	0.32	0.38	0.37	5107.2	5062.2	5064.3	5077.0	5075.0
ρ_{yx} =0.6, n=40, c_f = ₹ 50.00, c_r = ₹ 75.00 and c_s = ₹ 80.00											
ρ_2	ρ_1	μ_J	$\mu_{11}^{(0)}$	$\mu_{12}^{(0)}$	$\mu_{21}^{(0)}$	$\mu_{22}^{(0)}$	$C_T(J)$	$C_T(11)$	$C_T(12)$	$C_T(21)$	$C_T(22)$
0.5	**0.6**	0.55	0.73	0.69	0.45	0.44	5111.1	5147.9	5139.7	5090.5	5089.3
	0.7	0.55	0.54	0.49	0.44	0.42	5111.1	5109.4	5099.8	5089.2	5085.2
	0.8	0.55	0.32	0.33	0.43	0.40	5111.1	5065.1	5066.3	5087.3	5080.3
0.6	**0.6**	0.55	0.66	0.69	0.44	0.44	5111.1	5133.3	5139.7	5088.1	5089.3
	0.7	0.55	0.51	0.49	0.43	0.42	5111.1	5102.5	5099.8	5086.4	5085.2
	0.8	0.55	0.33	0.33	0.42	0.40	5111.1	5066.5	5066.3	5084.2	5080.3
0.7	**0.6**	0.55	0.61	0.69	0.42	0.44	5111.1	5123.7	5139.7	5085.8	5089.3
	0.7	0.55	0.48	0.49	0.41	0.42	5111.1	5097.5	5099.8	5084.0	5085.2
	0.8	0.55	0.33	0.33	0.40	0.40	5111.1	5066.6	5066.3	5081.5	5080.3

5.3. Monte Carlo Simulation

Monte Carlo simulation has been performed to get an overview of the proposed estimators in practical scenario through considering different choices of n and μ for better analysis.

Following three sets have been considered for the simulation study

Set I : n = 20, μ = 0.35, (m = 13,u = 7),

Set II : n = 20, μ = 0.20, (m = 16,u = 4),

Set III : n = 20, μ = 0.15, (m = 17,u = 3).

5.3.1. Simulation Algorithm

(i) Choose 5000 samples of size n=20 using simple random sampling without replacement on first wave for both the study and auxiliary variable.

(ii) Calculate sample mean $\bar{x}_{n|k}$ and $\bar{z}_{1|k}(n)$ for k =1, 2, - - -, 5000.

(iii) Retain m=17 units out of each n=20 sample units of the study and auxiliary variables at the first wave.

(iv) Calculate sample mean $\bar{x}_{m|k}$ and $\bar{z}_{1|k}(m)$ for k= 1, 2, - - -, 5000.

(v) Select u=3 units using simple random sampling without replacement from N-n=31 units of the population for study and auxiliary variables at second (current) wave.

(vi) Calculate sample mean $\bar{y}_{u|k}$ and $\bar{z}_{2|k}(m)$ for k = 1, 2, - - -, 5000.

(vii) Iterate the parameter ϖ from 0.1 to 0.9 with a step of 0.2.

(viii) Iterate ψ from 0.1 to 0.9 with a step of 0.1 within (ix).

(ix) Calculate the percent relative efficiencies of the proposed estimator $\check{\mathcal{T}}_{ij}$ (i, j=1, 2) with respect to estimator to $\bar{y}_n$ and $\hat{\bar{Y}}$ as

$$E(\check{\mathcal{T}}_{ij}, M) = \frac{\sum_{k=1}^{5000}\left[\check{\mathcal{T}}_{ij|k} - \bar{y}_{n|k}\right]^2}{\sum_{k=1}^{5000}\left[\check{\mathcal{T}}_{ij|k}\right]^2} \times 100 \text{ and } E(\check{\mathcal{T}}_{ij}, J) = \frac{\sum_{k=1}^{5000}\left[\check{\mathcal{T}}_{ij|k} - \hat{\bar{Y}}_{|k}\right]^2}{\sum_{k=1}^{5000}\left[\check{\mathcal{T}}_{ij|k}\right]^2} \times 100 \;;(i, j=1, 2), k=1, 2, ..., 5000.$$

Table 4: Simulation Results when proposed estimator $\check{T}_{ij}\,(i, j=1, 2)$ have been compared to $\bar{y}_n$

SET	ϖ_{ij}	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
I	$E(\check{T}_{11}, M)$	344.81	368.65	396.28	415.00	424.49	424.50	415.11	399.40	372.37
	$E(\check{T}_{12}, M)$	471.37	503.57	532.28	546.96	513.18	526.00	498.87	466.15	422.95
	$E(\check{T}_{21}, M)$	35.71	343.89	319.56	279.27	236.28	192.75	157.04	127.20	104.46
	$E(\check{T}_{22}, M)$	481.75	453.54	399.70	329.37	265.46	202.08	166.51	132.62	107.68
II	$E(\check{T}_{11}, M)$	295.21	331.51	362.39	388.13	413.31	432.55	443.08	440.17	431.41
	$E(\check{T}_{12}, M)$	468.60	517.25	552.45	579.43	599.73	604.44	597.23	570.13	538.56
	$E(\check{T}_{21}, M)$	304.46	323.91	323.39	303.13	271.90	232.93	199.24	168.94	142.09
	$E(\check{T}_{22}, M)$	487.75	499.92	464.47	405.47	338.91	273.84	224.47	184.37	151.52
III	$E(\check{T}_{11}, M)$	218.44	244.19	271.66	303.33	333.45	364.67	393.36	419.85	439.71
	$E(\check{T}_{12}, M)$	461.69	514.46	566.60	619.47	665.56	703.63	731.26	746.47	743.76
	$E(\check{T}_{21}, M)$	231.69	260.16	283.97	304.00	306.75	299.67	281.22	256.46	228.66
	$E(\check{T}_{22}, M)$	505.81	562.87	585.67	578.89	529.97	467.34	397.98	334.78	280.42

Table 5: Simulation results when the proposed estimator $\check{\mathcal{T}}_{11}$ is compared with the estimator $\hat{\hat{\mathrm{Y}}}$

ϖ_{11}	ψ / **SET**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0.1	**I**	326.12	332.43	428.19	602.76	882.27	1232.9	1637.4	2159.7	2787.3
	II	276.72	256.44	298.55	389.14	563.25	742.19	1039.6	1443.9	1886.6
	III	182.55	153.16	137.82	159.12	205.37	293.91	390.68	499.10	660.27
0.2	**I**	360.01	360.10	782.72	656.20	936.12	1351.0	1821.4	2348.6	310.88
	II	310.81	281.27	322.94	426.16	614.49	820.41	1173.6	1571.9	2015.2
	III	209.72	166.19	153.72	179.11	229.31	301.97	423.52	550.37	732.32
0.3	**I**	380.60	384.95	514.06	708.38	1010.3	1442.1	1891.4	2541.3	3284.2
	II	336.86	306.94	351.72	465.95	669.62	907.41	1286.9	1654.1	2139.2
	III	229.76	184.83	170.77	196.28	255.02	336.89	470.10	618.90	818.38
0.4	**I**	401.05	400.43	532.96	746.77	1060.7	1520.9	2012.7	2651.1	3471.6
	II	361.82	330.30	377.40	505.19	714.43	984.13	1391.5	1775.2	2298.0
	III	252.68	205.47	188.91	216.26	278.49	376.42	523.01	679.76	908.90
0.5	**I**	414.01	407.18	549.29	764.41	1090.0	1552.8	2044.6	2715.7	3539.2
	II	383.07	345.26	399.88	533.56	767.19	1051.3	1463.7	1887.3	2469.8
	III	278.45	227.24	209.50	239.44	304.72	411.04	574.21	748.40	994.88
0.6	**I**	411.71	410.70	545.30	755.91	1095.5	1549.5	2067.6	2694.6	3548.5
	II	398.06	361.34	415.99	556.75	797.60	1097.9	1514.8	1979.0	2589.5
	III	303.72	249.08	229.98	261.05	333.80	449.52	625.62	813.86	1085.2
0.7	**I**	400.98	400.17	532.67	742.17	1068.7	1521.9	2029.6	2638.5	3478.8
	II	404.20	269.57	424.83	566.29	814.96	1120.9	1552.3	2022.4	2642.9
	III	327.71	270.01	249.29	281.45	362.61	482.77	674.92	882.95	1164.0
0.8	**I**	383.48	385.90	502.00	704.79	1028.8	1456.8	1960.8	2521.3	3325.5
	II	398.93	368.51	424.67	569.15	812.87	1115.5	1549.6	2011.0	2651.8
	III	350.68	287.18	267.02	300.12	385.81	515.96	718.68	947.91	1240.0
0.9	**I**	360.51	360.82	469.77	663.00	955.74	1348.7	1826.2	2366.4	3125.1
	II	387.49	358.90	413.19	557.72	791.19	1379.7	1524.7	1970.4	2579.0
	III	366.09	300.07	280.72	315.76	404.82	541.84	752.31	995.79	1298.8

Table 6: Simulation results when the proposed estimator $\check{\mathcal{T}}_{12}$ is compared with the estimator $\hat{\bar{Y}}$

ϖ_{12}	SET \ ψ	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0.1	**I**	438.79	454.45	582.11	831.42	1201.8	1694.0	2233.1	2968.4	3831.7
	II	442.66	398.59	469.84	628.23	878.42	1185.5	1648.1	2361.0	2936.5
	III	427.90	343.13	312.59	375.43	472.41	634.04	887.31	1166.2	1504.7
0.2	**I**	481.01	491.89	651.11	890.31	1271.9	1828.8	2451.3	3226.3	4266.5
	II	484.85	439.04	508.95	677.41	967.42	1308.2	1846.5	2485.0	3124.6
	III	475.65	373.99	347.93	407.95	521.89	684.91	962.09	1278.4	1656.4
0.3	**I**	499.05	517.07	686.35	940.19	1348.0	1916.8	2527.0	3388.8	4461.2
	II	517.55	472.81	543.99	725.41	1040.0	1410.6	2000.8	2582.6	3306.2
	III	516.76	407.22	383.11	444.72	572.07	757.85	1047.3	1398.8	1833.6
0.4	**I**	516.41	527.75	692.80	971.12	1377.0	1970.1	2631.3	3458.3	4579.0
	II	545.34	500.29	570.34	768.94	1083.7	1499.5	2117.7	2706.0	3471.8
	III	556.92	445.77	416.14	480.80	614.70	829.45	1138.9	1507.8	1992.2
0.5	**I**	520.23	521.02	696.36	967.83	1375.2	1959.2	2596.5	3454.0	4525.4
	II	559.62	507.51	587.80	789.21	1129.7	1552.7	2165.4	2782.7	3624.1
	III	596.58	480.22	449.92	516.79	655.86	884.66	1221.4	1616.5	2127.7
0.6	**I**	501.75	508.90	670.69	927.60	1343.8	1900.8	2545.1	3324.2	4397.8
	II	560.74	513.26	589.91	794.99	1133.0	1565.5	2162.5	2814.5	3662.3
	III	627.38	510.46	477.32	544.38	696.06	935.55	1286.6	1703.9	2248.6
0.7	**I**	475.33	480.92	636.01	885.73	1272.4	1815.9	2426.3	3158.2	4173.9
	II	548.69	504.44	578.62	776.17	1112.8	1534.4	2132.8	2762.2	3599.1
	III	650.44	531.63	496.40	562.32	724.39	961.78	1335.1	1767.5	2313.1
0.8	**I**	442.41	450.39	581.65	818.15	1191.6	1688.9	2279.4	2929.7	3873.3
	II	519.16	482.41	556.39	748.91	1065.4	1467.8	2041.6	2637.1	3466.6
	III	663.15	538.25	507.13	569.47	733.17	977.60	1353.2	1808.1	2343.7
0.9	**I**	405.82	409.83	530.42	750.32	1077.9	1521.9	2067.0	2677.9	3542.1
	II	485.43	452.18	519.97	705.34	997.25	1379.7	1932.2	2487.1	3243.2
	III	654.95	532.58	504.31	567.16	727.56	972.56	1343.2	1799.6	2322.1

Table 7: Simulation results when the proposed estimator $\check{\mathcal{T}}_{21}$ is compared with the estimator $\hat{\bar{Y}}$

ϖ_{21}	SET \ ψ	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0.1	**I**	338.97	338.12	438.20	602.27	901.95	1257.7	1682.5	2185.7	2853.7
	II	289.16	265.72	309.35	409.45	583.03	777.05	1085.8	1500.7	1961.1
	III	194.16	162.27	147.18	169.77	218.56	290.84	417.32	531.93	700.09
0.2	**I**	341.08	355.92	451.14	604.12	882.83	1245.9	1694.9	2196.8	2886.7
	II	312.55	277.63	323.08	432.83	610.81	825.45	1162.4	1572.5	2029.8
	III	226.59	179.69	165.54	192.53	246.92	327.07	458.49	592.66	787.86
0.3	**I**	310.27	310.43	408.06	562.85	822.07	1156.5	1551.7	2053.6	2646.2
	II	306.79	275.65	320.34	430.40	607.64	828.57	1159.0	1522.2	1966.6
	III	244.65	197.87	182.29	208.92	270.87	359.27	503.59	657.30	870.32
0.4	**I**	266.62	269.46	348.62	491.28	716.91	1009.3	1355.7	1768.7	2298.6
	II	285.62	259.65	300.40	404.39	566.62	778.82	1088.9	1426.1	1851.3
	III	259.87	210.16	193.93	220.93	284.73	384.16	536.82	700.23	931.24
0.5	**I**	222.99	226.65	292.20	411.81	592.26	838.78	1131.5	1482.9	1913.1
	II	255.38	229.96	268.35	359.57	504.58	699.70	973.42	1268.4	1642.9
	III	266.55	215.90	199.93	226.91	291.38	391.08	548.24	717.8	951.67
0.6	**I**	183.72	186.48	240.18	335.81	487.34	687.36	923.04	121.44	1571.7
	II	220.07	199.74	232.47	310.54	434.76	603.52	836.57	1096.1	1407.9
	III	261.83	212.32	196.50	222.57	286.35	384.76	536.17	701.89	930.06
0.7	**I**	151.98	151.39	193.89	275.64	395.51	560.16	758.68	989.83	1286.5
	II	184.07	170.31	194.83	263.85	370.29	513.19	709.12	931.98	1201.6
	III	244.56	200.80	185.50	209.60	269.14	363.34	504.47	663.99	879.04
0.8	**I**	124.08	122.90	159.60	227.06	324.64	455.54	618.43	812.79	1057.1
	II	154.25	142.83	164.30	221.74	311.35	432.74	590.51	782.89	1005.1
	III	224.81	183.61	170.18	191.41	246.68	331.87	458.85	606.02	800.17
0.9	**I**	101.01	101.22	131.92	187.55	268.81	377.36	512.49	673.75	868.16
	II	127.98	120.18	138.06	186.92	261.26	362.84	494.70	657.75	844.50
	III	200.96	162.94	151.25	171.04	220.55	296.36	409.10	542.09	714.32

Table 8: Simulation results when the proposed estimator $\check{\mathcal{T}}_{22}$ is compared with the estimator $\hat{\bar{Y}}$

ϖ_{22}	ψ / SET	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
0.1	**I**	457.41	464.45	598.53	833.06	1222.4	1727.8	2280.5	3012.7	3960.2
	II	467.70	420.57	493.51	671.77	917.64	1258.4	1742.7	2481.8	3090.8
	III	471.41	378.45	347.07	417.54	523.23	700.91	985.84	1294.1	1660.1
0.2	**I**	446.59	443.03	585.78	800.42	1159.9	16648.2	2207.2	2917.1	3854.2
	II	480.06	428.10	497.99	673.58	938.32	1290.7	1790.5	2452.5	3116.9
	III	528.24	417.18	384.72	451.31	575.64	760.29	1077.0	1413.8	1830.2
0.3	**I**	383.86	388.26	504.92	699.77	1018.4	1438.3	1925.3	2563.7	3330.3
	II	442.47	396.88	464.53	624.20	875.24	1203.7	1678.5	2236.8	2856.7
	III	341.14	430.65	402.61	465.7	597.23	788.46	1112.2	1462.5	1926.8
0.4	**I**	312.67	317.80	407.50	579.53	838.79	1180.7	1596.0	2093.7	2727.8
	II	381.82	346.73	403.41	541.23	757.30	1046.6	1461.5	1930.6	2490.8
	III	527.18	420.54	394.10	451.06	576.66	772.73	1081.0	1434.4	1888.2
0.5	**I**	249.86	254.63	326.43	463.84	663.01	940.42	1273.4	1670.7	2156.6
	II	317.68	285.88	334.87	448.36	627.38	871.77	1219.9	1589.4	2054.2
	III	485.30	388.39	363.79	412.40	530.19	709.26	990.21	1314.4	1732.8
0.6	**I**	199.22	202.29	259.91	365.76	528.67	746.62	1001.8	1322.7	1709.6
	II	258.73	234.80	273.28	365.36	510.39	710.26	985.07	1295.4	1654.2
	III	425.04	341.89	318.71	361.92	465.0	622.42	863.99	1149.0	1510.5
0.7	**I**	159.98	160.52	204.73	292.77	418.76	593.77	803.81	1050.3	1366.1
	II	206.90	191.63	218.51	297.20	416.59	577.01	798.46	1051.3	1353.4
	III	355.08	291.67	267.62	306.67	392.04	532.08	733.74	974.43	1290.4
0.8	**I**	129.24	128.14	165.99	237.26	338.32	475.01	644.65	848.08	1103.6
	II	167.97	155.56	178.67	241.65	338.98	471.19	642.88	855.11	1095.0
	III	299.61	244.69	227.33	255.76	329.85	444.27	610.04	811.62	1072.3
0.9	**I**	104.04	104.34	135.79	193.67	277.07	389.17	528.32	695.07	895.60
	II	136.23	127.98	146.83	199.20	278.15	386.27	526.53	701.69	899.91
	III	250.56	202.66	188.41	213.37	275.45	370.39	508.81	677.67	893.24

7.Rendition of Results

7.1. Results from Generalized Empirical

a) The optimum values $\mu_{11}^{(0)}, \mu_{12}^{(0)}, \mu_{21}^{(0)}$ and $\mu_{22}^{(0)}$ exist for almost each combination of correlation coefficients. For increasing values of correlation of study and ancillary information, the values $\mu_{11}^{(0)}, \mu_{12}^{(0)}, \mu_{21}^{(0)}$ and $\mu_{22}^{(0)}$ decrease, which in accordance with Sukhatme et al (1984.)

b) As the correlation between study and ancillary information is increased, the percent relative efficiencies increase and the proposed estimators perform better than $\bar{y}_n$ and $\hat{\bar{Y}}$.

c) The proposed estimators provide a lesser fraction of fresh sample drawn afresh as compared to the estimator $\hat{\bar{Y}}$ for almost every considered choice of correlation coefficients. The estimator $\check{\mathcal{T}}_{21}$ performs best in terms of percent relative efficiency and the estimator $\check{\mathcal{T}}_{22}$ performs best in terms of sample drawn afresh at current occasion.

d) The optimum total cost of the survey decreases for increasing correlation between study and ancillary character. The estimator $\check{\mathcal{T}}_{22}$ requires the least total cost for the survey at the current occasion.

7.2 Simulation Results

a) From Table 4 to Table 8, it can be seen that the proposed estimators $\check{\mathcal{T}}_{ij}\left(i, j=1, 2\right)$ are efficient over $\bar{y}_n$ and $\hat{\bar{Y}}$ for all the considered sets.

b) Also in simulation study, it is observed that the estimator $\check{\mathcal{T}}_{22}$ is most efficient over the estimator $\hat{\bar{Y}}$ for all considered set but when $\check{\mathcal{T}}_{22}$ is compared to $\bar{y}_n$ it is most efficient among all proposed estimator only for set III.

c) As the fraction of sample drawn afresh is increased, the performance of all four estimators enhances.

8. Ratiocination

The entire detailed generalized and simulation studies attest that accompanying a motleying ancillary character with the study character certainly serves the purpose in long lag of two successive waves. The proposed estimators $\check{\mathcal{T}}_{ij}\left(\text{i, j=1, 2}\right)$ prove to be worthy in terms of precision and cost since all the proposed estimators provide a lesser fraction of freshly drawn sample at current occasion as compared to the estimators due to Jessen (1942). The minute observation suggest that the estimators $\check{\mathcal{T}}_{21}$ and $\check{\mathcal{T}}_{22}$ are providing approximately same fraction of sample to be drawn afresh at the current occasion but the total cost of survey is least for the estimator $\check{\mathcal{T}}_{22}$. Since both the estimators $\check{\mathcal{T}}_{21}$ and $\check{\mathcal{T}}_{22}$ are better than the sample mean estimator and the estimator due to Jessen (1942) and $\check{\mathcal{T}}_{21}$ is best in terms of precision but for little amount of precision, the cost of survey cannot be put on stake. Hence according to the requirement of survey investigator, one is free to choose any of the estimators out of $\check{\mathcal{T}}_{21}$ and $\check{\mathcal{T}}_{22}$. Hence the proposed estimators are recommended to the survey statisticians for their practical use.

UNIT - III

SEARCH OF GOOD ROTATION PATTERNS UNDER NON-RESPONSE:

APPLICATION OF IMPUTATION ECHNIQUES FOR ESTIMATING POPULATION MEAN AT CURRENT OCCASION

CHAPTER – 9*

Intercession of Non-Response through Imputation in Longitudinal Surveys for Population Mean

* Following is the publication based on the work of this chapter:--

1. Priyanka, K. and Mittal, R. (2016): Intercession of Non-Response through Imputation in Longitudinal Surveys for Population Mean. Journal of Modern Applied Statistical Methods, (Accepted For Publication).

Intercession of Non-Response through Imputation in Longitudinal Surveys for Population Mean

1. Introduction

However we assume to be ascertain that we can fetch complete response from the respondents but still we come across the situations where a less or huge portion of sample tends to be non-responding and hence sample provides incomplete information for statistical treatment. We may choose to go with incomplete information but that would entertain the false facts especially when the parameters to be estimated are affected with slight change in observation. The problem of sampling on two successive occasions was first considered by Jessen (1942) and latter this idea was extended by Patterson (1950), Narain (1953), Eckler (1955), Gordon (1983), Arnab & Okafor (1992), Feng & Zou (1997), Biradar & Singh(2001), Singh & Priyanka (2008a), Singh et al.(2013a) , Bandhopadhyay & Singh (2014) and many others.

Longitudinal surveys focus on studying and analyzing the trends and dynamics of those real life scenarios which intend to be monitored multiple times since one time canvassing of characteristics may not supply the very essential attributes of the character under study. In recent times, usage of longitudinal surveys has heighten enough for longitudinal analysis and in many cases; longitudinal surveys are carefully designed to permit the derivation of sophisticated analysis of the long dynamics of social and economic processes. The scenario of incompleteness becomes worst when one is interested in collecting data for more than one occasion because, even though you have a complete sample frame but may fail to obtain response in one or other ways (MCAR, MAR, OAR, PD and DNR).For example, In a Survey of different mines one may be interested in the total or mean yield from the mine. Now it may be possible that total or mean yield cannot be recorded since it had been mined completely, it had been shut down due to governmental issues or a natural calamity ruined the entire mine. So the unavailable sample units (mines) or the missing data for some mines could be imputed with more suitable entities (compensating the unavailability of response), so to get rid over the incompleteness of the data and to negotiate with the negative impact of non-response.

Immense efforts have been put together by Rubin (1976), Sande (1979), Kalton et al. (1981), Kalton & Kasprzyk (1982) and Singh & Singh (1991) by considering complete data set and discarding all those units for which information was not available for at least one time. One may site Lee et al. (1994, 1995), Singh & Horn (2002), Ahmed et al. (2006), Singh & Priyanka (2007b), Singh (2009) and Singh et al. (2013b) for various new estimators for estimation of parameters by method of imputation using additional auxiliary information.

Inspired by above motivating efforts we have implicitly assumed Missing Completely at Random and aspired to develop more worthy estimator for population mean while sampling over successive occasion using an additional auxiliary information, stable in nature over the occasions, by imputing missing data in the presence of non-response. For this an exponential ratio type estimator has been clubbed with a chain type ratio to exponential ratio type estimator over successive occasion to estimate population mean. The properties of the proposed estimator have been elaborated theoretically and empirically considering that (i) non-response may arise on both occasions, (ii) it may occur only at first occasion or (iii) it may occur only at second occasion while comparing the proposed estimator with estimator having complete response for all sample units at each occasion. A Simulation study has also been put through to substantiate the empirical results considering all mentioned three cases for different possibilities of non-response in the sample selected on different occasions.

2. Survey Design and Analysis

2.1. Notations

Let $U = \left(U_1, U_2, ..., U_N\right)$ be the N- element finite population, which has been sampled over two occasions. The characters under study is denoted by x(y) on the first (second) occasion, respectively. It is assumed that information on an almost stable auxiliary variable z, with the known population mean is available on both the occasions. We assume that there is non-response at both the occasions. A simple random sample without replacement s_n of n units has been drawn on the first occasion. Let the number of responding unit out of n sampled units, which are drawn at the first occasion, be denoted by r_1, the set of responding units in s_n by R_1 and that of non-responding by R_1^c. A random sub-sample s_m of $m = n\lambda$ unit is retained (matched) for its use on the current (second) occasion from the units which responded (r_1) at the first occasion and it is assumed these matched units are completely responding at the current (second) occasion as well. A fresh simple random sample (without replacement), s_u of $u = n - m = n\mu$ units, is drawn on the second occasion from the non-sampled units of the population so that the sample size on the second occasion remains the same i.e. n. Let the number of responding units out of u sampled units which are drawn afresh at current occasion, be denoted by r_2, the set of responding unit in s_u by R_2, and that of non-responding units by R_2^c. λ and μ $\left(0 \le \mu, \lambda \le 1, \lambda + \mu = 1\right)$ are the fractions of matched and fresh sample, respectively, at the current(second) occasion. For every unit $i \in R_j \left(j = 1, 2\right)$, the values $x_i \left(y_i\right)$ are observed, but for the units $i \in R_j^c \left(j = 1, 2\right)$ the values $x_i \left(y_i\right)$ are missing and instead imputed values are derived. The following notations have been used hereafter:

$\bar{X}, \bar{Y}, \bar{Z}$: Population means of the variables x, y and z respectively.

$\bar{y}_u, \bar{z}_u, \bar{y}_{r_2}, \bar{z}_{r_2}, \bar{x}_m, \bar{y}_m, \bar{z}_m, \bar{x}_{r_1}, \bar{y}_{r_1}, \bar{z}_{r_1}, \bar{x}_n, \bar{z}_n$: Sample mean of respective variate based on the sample sizes shown in suffice.

$\rho_{yx}, \rho_{xz}, \rho_{yz}$: Correlation coefficient between the variables shown in suffices.

S_x^2, S_y^2, S_z^2 : Population mean square of variables x, y and z respectively.

$f_1 = \left(\frac{r_1}{n}\right), f_2 = \left(\frac{r_2}{u}\right)$: The fraction of respondents at first and second occasions respectively.

$t_1 = (1 - f_1)$, $t_2 = (1 - f_2)$: The fraction of non- respondents at first and second occasions respectively.

2.2. Design

To estimate the population mean $\bar{Y}$ on the current (second) occasion, an estimator T_u has been proposed utilizing the concept of exponential ratio type estimators based on sample of the size u= nμ drawn afresh on the current (second) occasion. Considering that non-response occurs at current occasion, the missing values occurring in the sample of size u are replaced by imputed values. Hence, the following imputation method has been proposed to cope up with the problem of non-response in sample s_u

$$y_{\bullet i} = \begin{cases} y_i & \text{if} \quad i \in R_2 \\ \dfrac{1}{u - r_2}\left\{u\, \bar{y}_{r_2} \exp\left(\dfrac{\bar{\bar{Z}} - \bar{z}_{r_2}}{\bar{\bar{Z}} + \bar{z}_{r_2}}\right) - r_2\, \bar{y}_{r_2}\right\} & \text{if} \quad i \in R_2^c \end{cases} \tag{1}$$

where $\bar{y}_{r_2} = \dfrac{1}{r_2} \sum_{i \in R_2} y_i$ and $\bar{z}_{r_2} = \dfrac{1}{r_2} \sum_{i \in R_2} z_i$.

and hence the estimator for $\bar{Y}$ at current occasion is given by

$$T_u = \frac{1}{u} \sum_{i \in s_u} y_{\bullet i} = \frac{1}{u}\left(\sum_{i \in R_2} y_{\bullet i} + \sum_{i \in R_2^c} y_{\bullet i}\right) = \bar{y}_{r_2} \exp\left(\frac{\bar{\bar{Z}} - \bar{z}_{r_2}}{\bar{\bar{Z}} + \bar{z}_{r_2}}\right) \tag{2}$$

The second estimator T_m is based on sample size $m = n\lambda$ common to the both occasions utilizing information retained from first occasion. Since non- response is assumed to be occurring on first occasion as well so the missing values occurring in the sample of size n are replaced by imputed values. The following imputation technique has been suggested

$$x_{\bullet i} = \begin{cases} x_i & \text{if} \quad i \in R_1 \\ \dfrac{1}{n - r_1}\left\{n\, \bar{x}_{r_1} \exp\left(\dfrac{\bar{\bar{Z}} - \bar{z}_{r_1}}{\bar{\bar{Z}} + \bar{z}_{r_1}}\right) - r_1\, \bar{x}_{r_1}\right\} & \text{if} \quad i \in R_1^c \end{cases} \tag{3}$$

where $\bar{x}_{r_1} = \frac{1}{r_1} \sum_{i \in R_1} x_i$ and $\bar{z}_{r_1} = \frac{1}{r_1} \sum_{i \in R_1} z_i$.

Considering above proposed imputation method the estimator based on sample s_n is altered to

$$\bar{x}_n^* = \frac{1}{n} \sum_{i \in s_n} x_{\bullet i} = \frac{1}{n} \left(\sum_{i \in R_1} x_{\bullet i} + \sum_{i \in R_1^c} x_{\bullet i} \right) = \bar{x}_{r_1} \exp\left(\frac{\bar{Z} - \bar{z}_{r_1}}{\bar{Z} + \bar{z}_{r_1}} \right) \tag{4}$$

Therefore, Estimator based on sample size m common to both occasions which utilizes the missing values by above method of imputation is given by

$$T_m = \bar{y}_m^* \left(\frac{\bar{x}_n^*}{\bar{x}_m^*} \right) \tag{5}$$

where $\bar{y}_m^* = \bar{y}_m \exp\left(\frac{\bar{Z} - \bar{z}_m}{\bar{Z} + \bar{z}_m} \right)$, $\bar{x}_m^* = \bar{x}_m \exp\left(\frac{\bar{Z} - \bar{z}_m}{\bar{Z} + \bar{z}_m} \right)$ and $\bar{x}_n^* = \bar{x}_{r_1} \exp\left(\frac{\bar{Z} - \bar{z}_{r_1}}{\bar{Z} + \bar{z}_{r_1}} \right)$.

Considering the convex combination of the two estimators T_u and T_m, we have the final estimator of population mean $\bar{Y}$ on the current occasion as

$$T = \alpha\, T_u + (1 - \alpha) T_m \tag{6}$$

where $\alpha (0 \le \alpha \le 1)$ is a constant to be determined so as to minimize the mean squared error of the proposed estimator T.

2.3. Analysis

The properties of the proposed estimators T are derived under the following large sample approximations

$\bar{y}_{r_2} = \bar{Y}(1 + e_0)$, $\bar{y}_m = \bar{Y}(1 + e_1)$, $\bar{x}_m = \bar{X}(1 + e_2)$, $\bar{x}_{r_1} = \bar{X}(1 + e_3)$, $\bar{z}_{r_2} = \bar{Z}(1 + e_4)$, $\bar{z}_m = \bar{Z}(1 + e_5)$, $\bar{z}_{r_1} = \bar{Z}(1 + e_6)$, $\bar{y}_u = \bar{Y}(1 + e_7)$, $\bar{x}_n = \bar{X}(1 + e_8)$ and $\bar{z}_n = \bar{Z}(1 + e_9)$ such that $|e_i| < 1 \;\forall\; i = 0, 1, 2, 3, 4, 5, 6, 7, 8$ and 9 .

2.3.1. Bias and Mean Square Error of the Estimators T

The estimators T_u and T_m are exponential ratio and chain type ratio to exponential ratio type in nature respectively. Hence they are biased for population mean $\bar{Y}$. Therefore, the final estimator T defined in equation (6) is also biased estimator of $\bar{Y}$. The bias $B(.)$ and mean square error $M(.)$ of the proposed estimator T are obtained (ignoring finite population corrections) up to first order of approximations and thus we have following theorems:

Theorem 2.3.1.Bias of the estimator T to the first order of approximations is obtained as

$$B(T) = \alpha\, B(T) + (1 - \alpha)\, B(T) \tag{7}$$

where
$$B(T_u) = \frac{1}{r_2}\bar{Y}\left(\frac{3}{8}\frac{C_{002}}{\bar{Z}^2} - \frac{1}{2}\frac{C_{011}}{\bar{Y}\,\bar{Z}}\right) \tag{8}$$

and
$$B(T_m) = \bar{Y}\left(\frac{1}{m}\left(\frac{C_{200}}{\bar{X}^2} - \frac{C_{110}}{\bar{X}\bar{Y}}\right) + \frac{1}{r_1}\left(\frac{3}{8}\frac{C_{002}}{\bar{Z}^2} + \frac{C_{110}}{\bar{X}\bar{Y}} - \frac{1}{2}\frac{C_{011}}{\bar{Y}\bar{Z}} - \frac{C_{200}}{\bar{X}^2}\right)\right) \tag{9}$$

where $C_{rst} = E\left[(x_i - \bar{X})^r (y_i - \bar{Y})^s (z_i - \bar{Z})^t\right];\ (r, s, t) \geq 0$.

Proof: The bias of the estimator T is given by

$$B(T) = E\left[T - \bar{Y}\right] = \alpha\, B(T_u) + (1 - \alpha) B(T_m)$$

where $B(T_u) = E\left[T_u - \bar{Y}\right]$ and $B(T_m) = E\left[T_m - \bar{Y}\right]$

Using large sample approximations and retaining terms up-to the first order of approximations, the expression for $B(T_u)$ and $B(T_m)$ are obtained as in equation (8) and (9) and hence the expression for bias of the estimator T is obtained as in equation (7).

Theorem 2.3.2. Mean square error of the estimator T to the first order of approximations is obtained as

$$M(T) = \alpha^2\ M(T_u) + (1-\alpha)^2 M(T_m) + 2\ \alpha(1-\alpha)Cov(T_u, T_m) \tag{10}$$

$$M(T_u)\ =\ \frac{1}{r_2} A\ S_y^2 \tag{11}$$

$$M(T_m)\ =\ \left(\frac{1}{m}B + \frac{1}{r_1}C\right) S_y^2 \tag{12}$$

where $A = (5/4) - \rho_{yz}$, $B = 2 - 2\ \rho_{yx}$ and $C = 2\ \rho_{yx} - \rho_{yz} - (3/4)$.

Proof: The mean square error of the proposed estimator T is given by

$$M(T) = E\left[T - \bar{Y}\right]^2 = E\left[\alpha^2\left(T_u - \bar{Y}\right) + (1-\alpha)^2\left(T_m - \bar{Y}\right)\right]^2$$

$$= \alpha^2\ M(T_u) + (1-\alpha)^2 M(T_m) + 2\ \alpha(1-\alpha)Cov(T_u, T_m)$$

where $M\ (T_u)\ =\ E\left[T_u\ -\ \bar{Y}\right]^2$ and $M\ (T_m) = E\ \left[T_m\ -\ \bar{Y}\right]^2$.

Since x and y denote the same study character over two occasions and z being auxiliary variate positively correlated to x and y, therefore, looking at the stability nature (see Reddy (1978)) of the coefficient of variation and following Cochran (1977) and Feng & Zou (1997), the coefficient of variation of x, y and z are considered to be approximately same.

The estimators T_u and T_m are based on two independent samples of sizes u and m respectively, hence $Cov(T_u, T_m) = 0$. Considering population size is sufficiently large (i.e. $N \rightarrow \infty$), therefore finite population corrections are ignored and using large sample approximations and retaining terms upto the first order of approximations, the expression for $M(T_u)$ and $M(T_m)$ are obtained as given in equations (11) and (12) and hence the expressions for mean squarde errors of estimators T are obtained as in equation (10).

2.3.2. Minimum Mean Square Error of the Proposed Estimator T

Since the mean squared error of the estimator T given in equation (10) is the function of unknown constant α, therefore, it has been minimized with respect to α and subsequently the optimum value of α is obtained as

$$\alpha_{opt.} = M(T_m)/\left(M(T_u) + M(T_m)\right) \tag{13}$$

Now substituting the values of $\alpha_{opt.}$ in equation (10), we obtain the optimum mean squared error of the estimator T as

$$M(T)_{opt.} = \left(M(T_u)\ .\ M(T_m)\right)\Big/\left(M(T_u) + M(T_m)\right) \tag{14}$$

Further, substituting the value of the mean squared error of the estimators defined in equations (2) and (5) in equation (13) and (14) respectively, the simplified values of $\alpha_{opt.}$ and $M(T)_{opt.}$ are obtained as

$$\alpha_{opt.} = \mu f_2\ \left[\mu\, C - \left(f_1 B + C\right)\right]\Big/\left[\mu^2 f_2\ C - \mu\left(f_1 f_2 B + f_2 C - f_1 A\right) - A\right] \tag{15}$$

$$M(T)_{opt.} = \left[\mu\, C_1 - C_2\right] S_y^2 \big/ n\left[\mu^2\ C_3 - \mu\, C_4 - C_5\right] \tag{16}$$

where $C_1 = AC$, $C_2 = AC + f_1 A\, B$, $C_3 = f_2 C$, $C_4 = f_1 f_2\ B + f_2\ C - f_1 A$, $C_5 = f_1 A$, $A = \frac{5}{4} - \rho_{yz}$, $B = 2 - 2\,\rho_{yx}$, $C = 2\,\rho_{yx} - \rho_{yz} - \frac{3}{4}$ and μ is the fraction of the sample drawn afresh at the current(second) occasion.

Remark 2.3.1: $M(T)_{opt.}$ derived in equation (16) is a function of μ. To estimate the population mean on each occasion the better choice of μ are 1(case of no matching); however, to estimate the change in mean from one occasion to other, μ should be 0(case of complete matching). But intuition suggests that the optimum choices of μ are desired to devise the amicable strategy for both the problems simultaneously.

2.4. Optimum Replacement Strategies for the Estimator T

The key design parameter affecting the estimates of change is the overlap between successive samples. Maintaining high overlap between repeats of a survey is operationally convenient, since many sampled units have been located and have some experience in the survey. Hence to decide about the optimum value of μ (fractions of samples to be drawn afresh on current occasion) so that $\bar{Y}$ may be estimated with maximum precision and minimum cost, we minimize the mean square error $M(T)_{opt.}$ in equation (16) with respect to μ.

The optimum value of μ so obtained is one of the two roots given by

$$\mu = \left(D_2 \pm \sqrt{D_2^2 - D_1\, D_3}\right) \Big/ D_1 \tag{17}$$

where $D_1 = C_1\, C_3$, $D_2 = C_2\, C_3$, $D_3 = C_1\, C_5 + C_2\, C_4$.

The real value of μ exist, iff $D_2^2 - D_1\, D_3 \geq 0$. For any situation, which satisfies these conditions, two real values of μ may be possible , hence to choose a value of μ, it should be taken care of that $0 \leq \mu \leq 1$, all other values of μ are inadmissible. If both the real values of μ are admissible, the lowest one will be the best choice as it reduces the total cost of the survey. Substituting the admissible value of μ say μ_0 from equation (17) in equation (16), we get the optimum value of the mean square error of the estimator T with respect to α as well as μ which is given as

$$M(T)^*_{opt.} = \left[\mu_0\, C_1 - C_2\right] S_y^2 \Big/ n\left[\mu_0^2\, C_3 - \mu_0\, C_4 - C_5\right] \tag{18}$$

3. Special Cases

3.1. Case I: When there is Non-Response only at the First Occasion (Previous Occasion)

When there is a presence of non-response, the proposed estimator T for population mean $\bar{Y}$ changes to

$$T_1 = \varphi\, T_u^{\theta} + (1 - \varphi)\, T_m \tag{19}$$

where $T_u^{\theta} = \bar{y}_u \exp\left(\frac{\bar{Z} - \bar{z}_u}{\bar{Z} + \bar{z}_u}\right)$ and T_m is defined in equation (5) and $\varphi\left(0 \le \varphi \le 1\right)$ is a real constant to be determined so as to minimize the mean square error of the estimator T_1.

In this case, the optimum value of fraction of sample drawn afresh is obtained as

$$\hat{\mu} = \left(H_2 \pm \sqrt{H_2^2 - H_1 H_3}\right) \Big/ H_1 = \mu_1 (\text{say})$$

and the minimum mean square error of the estimator T_1 at the admissible value of $\hat{\mu}$ is derived as

$$M\left(T_1\right)^*_{\text{opt.}} = \left[\mu_1 G_1 - G_2\right] S_y^2 \Big/ n\left[\mu_1^2 C - \mu_1 G_3 - G_4\right] \tag{20}$$

where

$H_1 = CG_1$, $H_2 = CG_2$, $H_3 = G_1G_4 + G_2G_3$, $G_1 = AC$, $G_2 = AC + f_1AB$, $G_3 = f_1B + C - f_1A$, $G_4 = f_1A$ and $f_1 = r_1/n$.

3.2. Case II: When there is Non-Response only at the Second (Current) Occasion

The estimator for population mean $\bar{Y}$ at the current occasion in the presence of non-response at current occasion is given by

$$T_2 = \psi T_u + \left(1 - \psi\right) T_m^{\theta} \tag{21}$$

where $T_m^{\theta} = \bar{x}_n^* \left(\frac{\bar{y}_m^*}{\bar{x}_m^*}\right)$, $\bar{y}_m^* = \bar{y}_m \exp\left(\frac{\bar{Z} - \bar{z}_m}{\bar{Z} + \bar{z}_m}\right)$, $\bar{x}_m^* = \bar{x}_m \exp\left(\frac{\bar{Z} - \bar{z}_m}{\bar{Z} + \bar{z}_m}\right)$ and

$\bar{x}_n^* = \bar{x}_n \exp\left(\frac{\bar{Z} - \bar{z}_n}{\bar{Z} + \bar{z}_n}\right)$ T_u is defined in equation (2) and $\psi\left(0 \le \psi \le 1\right)$ is a real constant to be determined so as to minimize the mean square error of the estimator T_2.

In this case, the optimum value of fraction of sample drawn afresh is obtained as

$$\hat{\mu} = \left(H_5 \pm \sqrt{H_5^2 - H_4 H_6}\right) \Big/ H_4 = \mu_2 (\text{say})$$

and the minimum mean square error of the estimator T_2 at the admissible value of $\hat{\mu}$ is derived as

$$M\left(T_2\right)^*_{opt.} = \left[\mu_2\ G_5\ -G_6\right]S_y^2 \big/ n\left[\mu_2^2\ G_7\ -\mu_2\ G_3\ -A\right] \tag{22}$$

where

$H_4 = G_5G_7,\ H_5 = G_6G_7,\ H_6 = AG_5 + G_6G_8,\ G_5 = AC,\ G_6 = AB + AC,\ G_7 = f_2C,$

$G_8 = f_2B + f_2C - A$ and $f_2 = r_2/u.$

4. Efficiency Comparison

The percent relative loss in the efficiency of the proposed estimator T has been recorded to infer about the effect of incompleteness in the data over the occasions with respect to the estimator T_{CR} under the same circumstances but for complete response over the occasions.

Considering following estimator of population mean for complete response:

$$T_{CR} = \xi\ T_u^{\theta} + \left(1-\xi\right)T_m^{\theta} \tag{23}$$

where $\xi\left(0 \le \xi \le 1\right)$ is a real constant to be determined so as to minimize the mean square error of the estimator T_{CR}.

The optimum mean squared error for the estimator T_{CR} with respect to ξ as well as μ is obtained as

$$M\left(T_{CR}\right)^*_{opt.} = \left[\mu^* B_1 - B_2\right]S_y^2 \big/ n\left[\mu^{*2}\ C - \mu^* B_3 - A\right] \tag{24}$$

where $\mu^* = \left(B_5 \pm \sqrt{B_5^2\ - B_4\ B_6}\right) \Big/ B_4$, $B_4 = B_1C,\ B_5 = B_2C,\ B_6 = AB_1 + B_2B_3,\ B_1 = AC,$

$B_2 = AB + AC,\ B_3 = B + C - A,\ A = \left(5/4\right) - \rho_{yz},\ B = 2 - 2\ \rho_{yx}$ and $C = 2\ \rho_{yx} - \rho_{yz} - \left(3/4\right).$

The percent relative loss in precision of the estimators T, T_1 and T_2 with respect to the estimator T_{CR} under their respective optimality conditions are given by

$$\left.\begin{aligned} L_0 &= \frac{M(T)^*_{opt.} - M(T_{CR})^*_{opt.}}{M(T)^*_{opt.}} \times 100 \\ L_1 &= \frac{M(T_1)^*_{opt.} - M(T_{CR})^*_{opt.}}{M(T_1)^*_{opt.}} \times 100 \\ L_2 &= \frac{M(T_2)^*_{opt.} - M(T_{CR})^*_{opt.}}{M(T_2)^*_{opt.}} \times 100 \end{aligned}\right\} \tag{25}$$

5. Numerical Illustrations and Simulation

Empirical validation has been carried out by Monte Carlo Simulation. Real life situation of completely known finite population has been considered.

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

The population comprise of N = 51 states of United States. Let y_i be the total energy consumption during 2008 in the i^{th} state of U. S., x_i be the total energy consumption during 2003 in the i^{th} state of U. S. and z_i denote the total energy consumption during 2001 in the i^{th} state of U. S.

For the considered population, the value of μ defined in equation (17) and the percent relative loss in precision L_0, L_1 and L_2 defined in equation (25) of the estimator T, T_1 and T_2 respectively with respect to estimator T_{CR} have been computed and are presented in Table 1. To judge about the performance of the estimator in the presence of different percentages of non-response, a more general illustration has been worked out by considering choices of correlation coefficients of study and auxiliary variables on different occasions. These results have been shown in Table 2 to Table 4.

To validate the above empirical results, Monte Carlo simulation has also been performed for the considered population. For better analysis, the above simulation experiments were repeated for different choices of t_1 and t_2.

5.1. Simulation Algorithm

(i) Choose 5000 samples of size n=25 using simple random sampling without replacement on first occasion for both the study and auxiliary variable.

(ii) For $f_1 = 0.88$, choose $r_1=22$ responding units out of n=25 samples units.

(iii) Calculate sample mean $\bar{x}_{r_1 \mid k}$ and $\bar{z}_{r_1 \mid k}$ for k =1, 2, - - -, 5000.

(iv) Retain m=15 units out of each $r_1=22$ sample units of the study and auxiliary variables at the first occasion.

(v) Calculate sample mean $\bar{x}_{m \mid k}$ and $\bar{z}_{m \mid k}$ for k= 1, 2, - - -, 5000.

(vi) Select u=10 units using simple random sampling without replacement from N-n=26 units of the population for study and auxiliary variables at second (current) occasion.

(vii) For $f_2 = 0.90$, choose $r_2=9$ responding units out of u=10 samples units.

(viii) Calculate sample mean $\bar{y}_{r_2 \mid k}$, $\bar{y}_{m \mid k}$ and $\bar{z}_{r_2 \mid k}$ for k = 1, 2, - - -, 5000.

(ix) Iterate the parameter α from 0.1 to 0.9 with a step of 0.2.

(x) Iterate ξ from 0.1 to 0.9 with a step of 0.1 within (ix).

(xi) Calculate the percent relative loss in efficiencies of the proposed estimator T, T_1 and T_2 with respect to estimator to T_{CR} as

$$L(T) = \frac{\sum_{k=1}^{5000}\left[T_{\mid k} - T_{CR \mid k}\right]^2}{\sum_{k=1}^{5000}\left[T_{\mid k}\right]^2} \times 100 \ , \qquad L(T_1) = \frac{\sum_{k=1}^{5000}\left[T_{\mid k} - T_{CR \mid k}\right]^2}{\sum_{k=1}^{5000}\left[T_{1\mid k}\right]^2} \times 100$$

and $$L(T_2) = \frac{\sum_{k=1}^{5000}\left[T_{2\mid k} - T_{CR \mid k}\right]^2}{\sum_{k=1}^{5000}\left[T_{2\mid k}\right]^2} \times 100 , \quad k=1, 2, ..., 5000.$$

Table 1: Empirical Comparison of the proposed estimators T, T_1 and T_2 with respect to the estimator T_{CR}.

	Percent Relative Loss in Efficiency for estimator T		**Percent Relative Loss in Efficiency for estimator** T_1		**Percent Relative Loss in Efficiency for estimator** T_2	
μ^*	μ_0	L_0	μ_1	L_1	μ_2	L_2
0.3554	**0.5562**	3.3442	**0.4521**	-29.4932	**0.4779**	9.5745

Table 2: Percent relative loss (L_0) when estimator T is compared to the estimator T_{CR} in the presence of non-response on both the occasions.

ρ_{yz}			0.1			0.3			0.5			0.7		
t_1	t_2	ρ_{yx}	μ_0	μ^*	L_0	μ_0	μ^*	L_0	μ_0	μ^*	L_0	μ_0	μ^*	L_0
0.05	0.05	0.6	0.50	0.54	11.43	0.40	0.52	-4.11	0.89	0.49	-17.49	0.54	0.45	-35.58
		0.7	0.56	0.58	5.51	0.52	0.55	-9.12	0.44	0.52	-24.43	0.79	0.48	-38.66
		0.8	0.62	0.62	-1.66	0.59	0.60	-15.51	0.56	0.57	-29.79	0.48	0.53	-44.78
		0.9	0.70	0.70	-11.14	0.68	0.68	-24.02	0.65	0.65	-37.18	0.61	0.62	-50.78
	0.10	0.6	0.42	0.54	15.99	0.24	0.52	-1.03	*	0.49	*	0.61	0.45	-30.27
		0.7	0.52	0.58	10.40	0.46	0.55	-4.82	0.32	0.52	-21.23	*	0.48	*
		0.8	0.60	0.62	3.23	0.56	0.60	-11.03	0.51	0.57	-25.80	0.40	0.53	-41.59
		0.9	0.68	0.70	-12.99	0.65	0.68	-26.35	0.61	0.65	-40.09	0.55	0.62	-54.46
	0.15	0.6	0.34	0.54	19.89	0.07	0.52	1.53	*	0.49	*	0.69	0.45	-24.10
		0.7	0.48	0.58	15.02	0.40	0.55	-0.95	0.19	0.52	-9.34	0.76	0.48	-36.60
		0.8	0.57	0.62	8.04	0.53	0.60	-6.70	0.47	0.57	-22.06	0.31	0.53	-38.98
		0.9	0.67	0.70	-1.64	0.6	0.68	-15.25	0.61	0.65	-29.19	0.56	0.62	-43.65
0.15	0.05	0.6	0.55	0.54	13.68	0.46	0.52	-2.78	0.91	0.49	-17.50	0.59	0.45	-36.80
		0.7	0.61	0.58	8.79	0.57	0.55	-6.75	0.49	0.52	-2306	0.81	0.48	-38.72
		0.8	0.66	0.62	2.75	0.64	0.60	-11.94	0.60	0.57	-27.17	0.53	0.53	-43.28
		0.9	0.73	0.70	-5.37	0.71	0.68	-19.01	0.69	0.65	-33.02	0.65	0.62	-47.60
	0.10	0.6	0.48	0.54	18.86	0.32	0.52	1.00	*	0.49	*	0.65	0.45	-31.11
		0.7	0.57	0.58	14.20	0.52	0.55	-1.92	0.39	0.52	-19.29	*	0.48	*
		0.8	0.64	0.62	8.10	0.61	0.60	-7.03	0.56	0.57	-22.76	0.46	0.53	-39.64
		0.9	0.72	0.70	-0.25	0.70	0.68	-14.28	0.67	0.65	-28.68	0.63	0.62	-43.68
	0.15	0.6	0.41	0.54	23.48	0.17	0.52	3.44	*	0.49	*	0.72	0.45	-24.63
		0.7	0.53	0.58	19.41	0.46	0.55	2.54	0.27	0.52	-16.42	*	0.48	*
		0.8	0.62	0.62	13.42	0.58	0.60	-2.21	0.52	0.57	-18.54	0.38	0.53	-36.51
		0.9	0.71	0.70	4.91	0.69	0.68	-9.52	0.65	0.65	-24.36	0.61	0.62	-39.83
0.20	0.05	0.6	0.57	0.54	14.81	0.49	0.52	-2.11	0.91	0.49	-17.51	0.61	0.45	-37.40
		0.7	0.63	0.58	10.44	0.60	0.55	-5.55	0.52	0.52	-22.37	0.83	0.48	-38.75
		0.8	0.68	0.62	5.00	0.66	0.60	-10.13	0.62	0.57	-25.85	0.56	0.53	-42.53
		0.9	0.75	0.70	-2.41	0.73	0.68	-16.46	0.71	0.65	-30.91	0.67	0.62	-46.00
	0.10	0.6	0.51	0.54	20.31	0.36	0.52	2.02	*	0.49	*	0.67	0.45	-31.53
		0.7	0.59	0.58	16.13	0.55	0.55	-0.45	0.42	0.52	-18.32	*	0.48	*
		0.8	0.66	0.62	10.59	0.63	0.60	-4.99	0.59	0.57	-21.22	0.49	0.53	-38.66
		0.9	0.74	0.70	2.89	0.73	0.68	-11.55	0.69	0.65	-26.40	0.65	0.62	-41.92
	0.15	0.6	0.45	0.54	25.31	0.22	0.52	4.91	*	0.49	20.08	0.74	0.45	-24.89
		0.7	0.56	0.58	21.65	0.49	0.55	4.31	0.31	0.52	-15.08	*	0.48	*
		0.8	0.64	0.62	16.17	0.60	0.60	0.06	0.55	0.57	-16.76	0.42	0.53	-35.26
		0.9	0.73	0.70	8.28	0.70	0.68	-6.59	0.67	0.65	-21.89	0.63	0.62	-37.89

Note: (*) denotes that percent relative loss does not exist since value of optimum μ_0 does not exist.

Table 3: Percent relative loss (L_1) when estimator T is compared to the estimator T_{CR} in the presence of non-response at first occasion.

ρ_{yz}		0.1			0.3			0.5			0.7		
t_1	ρ_{yx}	μ_1	μ^*	L_1	μ_1	μ^*	L_1	μ_1	μ^*	L_1	μ_1	μ^*	L_1
0.05	0.6	0.56	0.54	6.31	0.54	0.52	-8.51	0.51	0.49	-23.91	0.48	0.45	-40.15
	0.7	0.60	0.58	0.41	0.57	0.55	13.79	0.55	0.52	-28.50	0.51	0.48	-43.95
	0.8	0.64	0.62	-6.61	0.62	0.60	-20.10	0.59	0.57	-33.99	0.56	0.53	-48.49
	0.9	0.72	0.70	-15.85	0.70	0.68	-28.40	0.67	0.65	-41.23	0.64	0.62	-54.48
0.10	0.6	0.59	0.54	7.17	0.56	0.52	-8.12	0.54	0.49	-24.05	0.50	0.45	-40.98
	0.7	0.62	0.58	1.18	0.60	0.55	-12.84	0.57	0.52	-28.05	0.54	0.48	-44.12
	0.8	0.66	0.62	-4.62	0.64	0.60	-18.52	0.62	0.57	-32.87	0.58	0.53	-47.93
	0.9	0.73	0.70	-13.15	0.71	0.68	-26.07	0.69	0.65	-39.30	0.66	0.62	-53.03
0.15	0.6	0.61	0.54	8.03	0.59	0.52	-7.72	0.56	0.49	-24.20	0.53	0.45	-41.80
	0.7	0.64	0.58	3.22	0.62	0.55	-11.87	0.59	0.52	-27.60	0.56	0.48	-44.29
	0.8	0.68	0.62	-2.60	0.66	0.60	-16.92	0.64	0.57	-31.75	0.60	0.53	-47.37
	0.9	0.74	0.70	-10.42	0.73	0.68	-23.71	0.71	0.65	-37.36	0.68	0.62	-51.58

Table 4: Percent relative loss (L_2) when estimator T is compared to the estimator T_{CR} in the presence of non-response at second occasion.

ρ_{yz}		0.1			0.3			0.5			0.7		
t_2	ρ_{yx}	μ_2	μ^*	L_2	μ_2	μ^*	L_2	μ_2	μ^*	L_2	μ_2	μ^*	L_2
0.05	0.6	0.47	0.54	10.32	0.37	0.52	-4.77	0.89	0.49	-17.48	0.52	0.45	-34.97
	0.7	0.54	0.58	3.90	0.50	0.55	-10.30	0.41	0.52	-25.11	0.78	0.48	-38.63
	0.8	0.60	0.62	-3.83	0.57	0.60	-17.27	0.53	0.57	-31.09	0.45	0.53	-45.52
	0.9	0.69	0.70	-13.97	0.67	0.68	-26.47	0.64	0.65	-39.23	0.59	0.62	-52.35
0.10	0.6	0.39	0.54	14.57	0.20	0.52	-2.04	*	0.49	*	0.59	0.45	-29.84
	0.7	0.49	0.58	8.52	0.44	0.55	-6.25	0.28	0.52	-22.20	*	0.48	-2.15
	0.8	0.58	0.62	0.84	0.54	0.60	-13.00	0.49	0.57	-27.31	0.37	0.53	-42.55
	0.9	0.67	0.70	-9.41	0.65	0.68	-22.25	0.62	0.65	-35.37	0.57	0.62	-48.89
0.15	0.6	0.31	0.54	18.12	0.02	0.52	-0.90	*	0.49	*	0.67	0.45	-23.84
	0.7	0.45	0.58	12.86	0.37	0.55	-2.67	0.14	0.52	-20.36	*	0.48	*
	0.8	0.55	0.62	5.42	0.51	0.60	-8.90	0.44	0.57	-23.79	0.27	0.53	-40.21
	0.9	0.66	0.70	-4.83	0.63	0.68	-18.05	0.59	0.65	-31.56	0.54	0.62	-45.54

Note: (*) denotes that percent relative loss does not exist since value of optimum μ_2 does not exist.

Table 5: Simulation result when the proposed estimator T is compared with the estimator T_{CR} when non-response occurs on both the occasion

α \ ξ	SET	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.1	I	-23.85	6.10	33.28	43.81	57.06	62.36	66.77
	II	-24.09	5.21	22.58	47.06	53.96	60.85	66.88
	III	-13.18	16.70	33.33	43.50	59.94	66.85	68.68
0.3	I	-46.87	-19.46	10.19	29.28	43.39	54.65	60.04
	II	-54.92	-25.77	2.14	30.79	41.07	50.64	60.43
	III	-42.99	-9.49	12.51	33.32	51.20	58.26	59.94
0.5	I	-80.12	-47.12	-9.73	12.33	31.92	45.49	51.44
	II	-90.55	-56.39	-16.28	14.25	28.86	41.93	51.74
	III	-85.44	-39.60	-9.91	16.85	37.16	46.19	50.30
0.7	I	-111.24	-70.35	-27.92	-2.13	21.43	36.56	43.82
	II	-118.92	-75.86	-31.04	2.00	19.47	34.44	44.77
	III	-123.06	-68.43	-33.23	0.2.37	24.28	34.80	40.76
0.9	I	-125.68	-82.26	-37.31	-8.64	16.15	32.29	40.13
	II	-127.09	-81.47	-35.76	-3.02	17.91	32.01	42.22
	III	-148.24	-88.23	-49.20	-12.81	14.80	27.73	34.46

I: n=25, $\mu = 0.40$, t_1=0.12, t_2=0.10 , II: n=25, $\mu = 0.40$, t_1=0.16, t_2=0.20

III: n=25, $\mu = 0.40$, t_1=0.28, t_2=0.30

Table 6: Simulation result when the proposed estimator T_1 is compared with the estimator T_{CR} when non-response occurs only on first occasion

φ \ ξ	SET	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.1	I	-20.98	-9.49	30.43	39.38	53.20	65.04	67.24
	II	-20.86	18.17	34.12	49.80	64.98	67.15	71.37
	III	-88.05	-50.30	-15.35	8.46	32.66	39.43	51.38
0.3	I	-56.20	-28.72	7.70	24.69	45.06	55.00	60.02
	II	-42.21	-4.87	17.89	36.77	52.01	58.26	64.38
	III	-126.69	-82.31	-44.79	-133.22	15.23	31.43	38.67
0.5	I	-88.42	-53.78	-15.51	9.19	31.80	44.73	50.29
	II	-75.35	-35.25	-4.74	19.44	40.41	49.53	55.66
	III	-178.35	-117.28	-71.18	-34.22	-2.78	17.97	25.30
0.7	I	-119.74	-80.62	-36.35	-3.79	19.56	35.45	41.96
	II	-109.31	-60.23	-24.21	2.17	28.08	40.13	47.12
	III	-211.19	-144.73	-92.32	-52.77	-14.78	7.84	16.30
0.9	I	-138.83	-92.46	-47.51	-12.53	12.47	29.92	37.16
	II	-131.52	-76.53	-37.68	-7.70	20.50	34.10	41.37
	III	-218.85	-149.08	-95.39	-53.67	-15.79	5.98	14.28

I: n=25, $\mu = 0.40$, t_1=0.12 , II: n=25, $\mu = 0.40$, t_1=0.16

III: n=25, $\mu = 0.40$, t_1=0.28

Table 7: Simulation result when the proposed estimator T_2 is compared with the estimator T_{CR} when non-response occurs only on first occasion

ψ \ ξ	SET	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.1	I	0.20	21.90	39.82	53.53	65.07	71.45	74.45
	II	0.49	11.24	39.58	54.34	64.62	70.92	72.31
	III	0.97	22.26	40.05	53.15	65.12	70.63	74.91
0.3	I	-24.01	2.56	25.24	41.91	55.55	63.79	67.82
	II	-24.03	2.40	24.51	42.73	55.57	63.87	66.25
	III	-24.0	2.45	25.41	42.42	55.88	63.72	67.49
0.5	I	-51.48	-19.09	7.28	29.29	45.59	55.71	60.37
	II	-53.87	-22.26	6.66	29.17	44.95	55.14	58.55
	III	-38.49	-20.32	7.25	28.09	44.54	54.64	59.88
0.7	I	-76.61	-40.99	-8.27	17.96	36.49	48.50	53.69
	II	-84.57	-48.72	-12.11	15.02	33.79	46.31	50.53
	III	-69.30	-47.13	-13.11	12.94	32.66	44.53	51.04
0.9	I	-91.20	-51.71	-16.71	11.37	31.41	44.66	50.21
	II	-110.70	-66.95	-27.96	4.02	25.74	39.31	44.18
	III	102.42	-73.68	-36.88	-1.03	21.95	35.42	42.77

I: n=25, $\mu = 0.40$, t_2=0.20, II: n=25, $\mu = 0.40$, t_2=0.30

III: n=25, $\mu = 0.40$, t_2=0.40

9. Rendition of Results

The performance of an estimator in successive sampling in the presence of non-response is generally judged on the basis of percent relative loss in efficiency (lesser is loss better is the estimator) and in terms of optimum value of fraction of fresh sample to be drawn on current (second) occasion which in turns is directly associated to the cost of survey. Following interpretation can be drawn from Tables 1- 7,

(1)From Table-1, it is observed that

(a) Optimum values μ_0, μ_1 and μ_2 for the estimators T, T_1 and T_2 respectively exist for the considered Population and $\mu^* < \mu_1 < \mu_2 < \mu_0$, which justifies the applicability of the proposed estimators T, T_1 and T_2 at optimum conditions. This also signifies that portion consisting more non response requires a more number of units in the sample to be drawn afresh on current occasion.

(b) Lesser percent relative loss in efficiency is observed in terms of precision indicating the proposed estimator T (at optimal conditions) to be considerable if non response appears in the survey design. This result justifies the use of additional auxiliary information which is stable over time embedding with exponential type structure at both occasions in two occasion successive sampling.

(2) In Table-2, we see that

(a) For a fixed value of t_1, as percentage of non-response on current occasion increases the amount of loss increases which is natural and for a fixed t_2 as the value of ρ_{yx} increases the loss in efficiency decreases, in fact a little gain is observed when compared to the estimator T_{CR}.

(b) For fixed amount of t_1 and t_2 as the value of ρ_{yz} increases, the loss in efficiency decrease.

(c) As the percentage of non-response at first occasion increases for fixed value of t_2, loss in efficiency increases.

(3) From Table -3 and Table-4 we observe that

(a) For fixed values of t_1 and t_2, increasing values of ρ_{yx} and ρ_{yz} the percent relative loss in efficiency decreases.

(b) As we keep on increasing the value of t_1 and t_2, percent relative loss increases.

(4) From the simulation results presented in Table-5, 6 and 7, where the estimators T, T_1 and T_2 are respectively compared to estimator T_{CR}, following results can be drawn

(a) The values for $L(T), L(T_1)$ and $L(T_2)$ increase as the value of ξ increase for fixed choice of α. (b) As we increase the value of α, φ and ψ respectively for fixed choices of ξ, the value of $L(T), L(T_1)$ and $L(T_2)$ decrease and gain is also observed over the estimator T_{CR}.

10. Conclusion

The thorough analysis of proposed estimators utilizing information on an additional auxiliary variable in the presence of non-response with variety of cases depending upon the occurrence of non-response, seems to be interesting enough as an amalgamation of exponential structure with ratio type estimator because even in the midst of non- response, the proposed method of imputation not just provides lesser percent relative loss in efficiency of the estimator but it also helps in reducing the cost of survey. So the proposed estimator T can be considered for its practical use in the presence of non-response by survey practitioners.

CHAPTER – 10*

A fresh approach for Intercession of Non-Response in Multivariate Longitudinal Designs

---

* Following is the publication based on the work of this chapter:--

1. Priyanka, K. and Mittal, R. (2016): A fresh approach for Intercession of Non-Response in Multivariate Longitudinal Designs. Communication in Statistics (Theory and Methods), (Accepted For publication).

A fresh approach for Intercession of Non-Response in Multivariate Longitudinal Designs

1. Introduction

Probably a most smartly designed sample survey may bout with incompleteness of data due to completely stochastic nature of non-response and cannot ascertain the complete response from the respondents due to many uncertainties lying in the conduction of survey. However we always have a data to conclude with but in an incomplete form of the true fact of the study. The severity of spoiling inference gets worse when the samples are drawn on multiple waves. The research related to sampling on successive waves has started in the international arena since early 1942. Some of the work in the last five decades are summarized below: Jessen (1942) was pioneer to start the work and latter this idea was extended by Patterson (1950), Narain (1953), Eckler (1955), Gordon (1983), Arnab and Okafor (1992), Feng and Zou (1997), Biradar and Singh (2001), Singh and Priyanka (2008a), Singh et al. (2013a), Bandyopadhyay and Singh(2014), Priyanka and Mittal(2014), Priyanka and Mittal(2015a, 2015b)and many others. Longitudinal surveys focus on studying and analyzing the trends and dynamics of those real life scenarios which intend to be monitored multiple times since one time canvassing of characteristics may not supply the very essential attributes of the character under study.

The scenario of incompleteness becomes awful when one is interested in collecting data for more than one wave because, even though we have a complete sample frame but may fail to obtain response in one or other ways (MCAR, MAR, OAR, PD and DNR). For example, in a survey of different retail outlets of consumable commodities, one may be keen to estimate average daily sales of a specific product in a particular outlet. Now it may not be possible to record the total or average daily sale of specific product from any of the retail outlet since the outlet might be shut down, outlet might be closed that day, the sale of product might be discontinued from that outlet; the product might be out of stock in the outlet or the producing company might have discontinued the production of the product due to any issue. So rather than discarding the unavailable sample units, suitable values can be imputed in place of unavailable sample units so as to overcome incompleteness of the data due to non-response.

Immense efforts have been put together by Rubin (1976), Sande (1979), Kalton et al.(1981), Kalton and Kasprzyk (1982), Singh and Singh (1991) by considering complete data set and discarding all those units for which information was not available for at least one time. One may site Lee et al. (1994, 1995), Singh and Horn (2002), Ahmed et al. (2006), Singh and Priyanka (2007b), Singh (2009),Singh et al. (2010), Diana and Perri (2010) and Singh et al. (2013b) for various new estimators for estimation of parameters in presence of non-response.

So many authors have effort fully commanded the literature by an exceptional use of multi-auxiliary information while sampling over two or more waves for the estimation of population mean when complete response of sample units is available but as far as our knowledge is concerned, it is the very initial effort to use imputation technique in order to handle non-response using multi-auxiliary information for the estimation of population mean on successive waves. Motivated by above argument, MCAR have been assumed implicitly and completely known multi-auxiliary information have been utilized to estimate the population mean in the presence of non-response while sampling over two successive waves. For this an exponential ratio type estimator has been clubbed with chain type ratio to exponential ratio type estimator over successive waves to estimate population mean. The problem of non-response (incompleteness in data) has been handled by imputation technique. Fresh imputation technique has been devised. The possible cases in which non-response may creep in two successive waves have been elaborated in details. The properties have been discussed theoretically as well as empirically. The proposed estimators under devised imputation techniques have been compared with a multivariate weighted estimator due to Priyanka et al. (2015). A detailed simulation algorithm has been designed and applied to substantiate the empirical and theoretical results.

2. Survey Design and Analysis

2.1 Notations

Let $U = (U_1, U_2, ..., U_N)$ be the N- element finite population, which has been sampled over two successive waves. The characters under study is denoted by x(y) on the first (second) wave, respectively. It is assumed that information on almost stable auxiliary variables $z_1, z_2, ..., z_p$, with the known population mean, are available on both successive waves. It has been assumed that there is non-response on both successive waves (occasions). A simple random sample without replacement s_n of n units has been drawn on the first wave. Let the number of responding unit out of n sampled units, which are drawn at the first wave, be denoted by r_1, the set of responding units in s_n by R_1 and that of non-responding by R_1^c. A random sub-sample s_m of $m = n\lambda$ unit is retained (matched) for its use on the current (second) wave from the units which responded (r_1) at the first wave and it is assumed that these matched units are completely responding at the current (second) wave as well. A fresh simple random sample (without replacement), s_u of $u = n - m = n\mu$ units, is drawn on the second wave from the non-sampled units of the population so that the sample size on the second wave remains the same. Let the number of responding units out of u sampled units which are drawn afresh at current wave, be denoted by r_2, the set of responding unit in s_u by R_2, and that of non-responding units by R_2^c. λ and μ $(\lambda + \mu = 1)$ are the fractions of matched and fresh sample, respectively, at the current (second) wave. For every unit $j \in R_k$ $(k = 1, 2)$, the values $x_j (y_j)$ are observed, but for the units $j \in R_k^c$ $(k = 1, 2)$ the values $x_j (y_j)$ are missing and instead values are imputed. The following notations have been used hereafter:

$\bar{X}, \bar{Y}, \bar{Z}_i$: Population means of the variables x, y and $z_1, z_2, ..., z_p$ respectively.

$\bar{y}_u, \bar{z}_u, \bar{y}_{r_2}, \bar{z}_i(r_2), \bar{x}_m, \bar{y}_m, \bar{z}_i(m), \bar{x}_{r_1}, \bar{y}_{r_1}, \bar{z}_i(r_1), \bar{x}_n, \bar{z}_i(n)$: Sample mean of respective variates based on the sample sizes shown in suffice.

$\rho_{yx}, \rho_{xz}, \rho_{yz_i}, \rho_{z_iz_j}$: Correlation coefficient between the variables shown in suffices for

$i \neq j=1, 2, ..., p$.

$S_x^2, S_y^2, S_{z_i}^2$: Population mean square of variables x, y and $z_1, z_2, ..., z_p$ respectively.

$f_1=\left(\frac{r_1}{n}\right), f_2=\left(\frac{r_2}{u}\right)$: The fraction of respondents at first and second waves respectively.

$t_1=(1-f_1), t_2=(1-f_2)$: The fraction of non- respondents at first and second waves respectively.

2.2 Survey Design under Proposed Imputation Technique

To estimate the population mean $\bar{Y}$ on the current (second) wave, utilizing p-additional auxiliary information which are stable over time and are readily available on both successive waves. Considering the case of non-response at current wave, the missing values in the sample of size u, are replaced by imputed values. A fresh imputation technique have been proposed to manage non-response as under

$$y_{\bullet j}=\begin{cases} y_j & \text{if} \quad j \in R_2 \\ \dfrac{1}{u-r_2}\left\{u\,\bar{y}_{r_2}\,\exp\left(\dfrac{\bar{Z}_i-\bar{z}_i(r_2)}{\bar{Z}_i+\bar{z}_i(r_2)}\right)-r_2\,\bar{y}_{r_2}\right\} & \text{if} \quad j \in R_2^c \end{cases} \tag{1}$$

where $\bar{y}_{r_2}=\frac{1}{r_2}\sum_{j\in R_2} y_j$ and $\bar{z}_i(r_2)=\frac{1}{r_2}\sum_{j\in R_2} z_{i_j}, i=1, 2, ..., p$.

and hence the estimator $T(i, u)\{i=1, 2, ..., p\}$ under above proposed imputation technique becomes

$$T(i,u)=\bar{y}_{r_2}\,\exp\left(\frac{\bar{Z}_i-\bar{z}_i(r_2)}{\bar{Z}_i+\bar{z}_i(r_2)}\right), i=1, 2, ..., p. \tag{2}$$

A multivariate weighted estimator T_u based on sample of the size u= nμ drawn afresh on the current (second) wave is proposed as

$$T_u = \mathbf{W}_u' \mathbf{T}_{\exp}(u) \tag{3}$$

where $\mathbf{W}_u$ is a column vector of p-weights given by $\mathbf{W}_u = [w_{u_1} \quad w_{u_2} . \quad . \quad . \quad w_{u_p}]'$

and $\mathbf{T}_{\exp}(u) = \begin{bmatrix} T(1,u) \\ T(2,u) \\ \vdots \\ T(p,u) \end{bmatrix}$, such that $\mathbf{1}'\mathbf{W_u} = 1$, where $\mathbf{1}$ is a column vector of order p.

The second estimator T_m, based on sample of size m, is also proposed as weighted multivariate chain type ratio to exponential ratio estimator and hence is given by

$$T_m = \mathbf{W}'_{\mathbf{m}}\mathbf{T}_{\exp}(m,n) \tag{4}$$

where $\mathbf{W_m}$ is a column vector of p-weights as $\mathbf{W_m} = [w_{m_1} \quad w_{m_2} \cdot \quad \cdot \quad \cdot \quad w_{m_p}]'$

and $\mathbf{T}_{\exp}(m,n) = \begin{bmatrix} T(1,m,n) \\ T(2,m,n) \\ \vdots \\ T(p,m,n) \end{bmatrix}$, Such that $\mathbf{1}'\mathbf{W_m} = 1$, where $\mathbf{1}$ is a column vector of order p.

Since non- response is assumed to be occurring on first wave as well, so the missing values occurring in the sample of size n are replaced by imputed values. For finding relevant imputed values following imputation technique has been proposed

$$x_{\cdot j} = \begin{cases} x_j & \text{if} \quad j \in R_1 \\ \dfrac{1}{n - r_1}\left\{n\,\bar{x}_{r_1} \exp\left(\dfrac{\bar{Z}_i - \bar{z}_i(r_1)}{\bar{Z}_i + \bar{z}_i(r_1)}\right) - r_1\,\bar{x}_{r_1}\right\} & \text{if} \quad j \in R_1^c \end{cases} \tag{5}$$

where $\bar{x}_{r_1} = \dfrac{1}{r_1}\sum_{j\in R_1} x_j$ and $\bar{z}_i(r_1) = \dfrac{1}{r_1}\sum_{j\in R_1} z_{i_j}(r_1)$, i=1, 2, ..., p.

In the light of above proposed imputation technique the estimator based on sample s_n becomes

$$\bar{x}_n^* = \bar{x}_{r_1} \exp\left(\frac{\bar{Z}_i - \bar{z}_i(r_1)}{\bar{Z}_i + \bar{z}_i(r_1)}\right), \text{ i=1, 2, ..., p} \tag{6}$$

Now the estimator based on sample size m common to both successive waves is proposed as

$$T(i, m, n) = \left(\frac{\bar{y}^*(i, m)}{\bar{x}^*(i, m)}\right)\bar{x}^*(i, n) \tag{7}$$

where $\bar{y}^*(i, m) = \bar{y}_m \exp\left(\frac{\bar{Z}_i - \bar{z}_i(m)}{\bar{Z}_i + \bar{z}_i(m)}\right)$, $\bar{x}^*(i, m) = \bar{x}_m \exp\left(\frac{\bar{Z}_i - \bar{z}_i(m)}{\bar{Z}_i + \bar{z}_i(m)}\right)$

and $\bar{x}^*(i, n) = \bar{x}_{r_1} \exp\left(\frac{\bar{Z}_i - \bar{z}_i(r_1)}{\bar{Z}_i + \bar{z}_i(r_1)}\right)$ for i=1, 2, 3, ..., p.

The optimum weights $\mathbf{W_u}$ and $\mathbf{W_m}$ in T_u and T_m are chosen by minimizing their mean squared errors respectively.

Now a convex linear combination of the two estimators T_u and T_m has been considered to define the final estimator of population mean $\bar{Y}$ on the current wave in the presence of non-response on both successive waves and is given as

$$T_{|p}(NR) = \alpha\, T_u + (1 - \alpha)\, T_m \tag{8}$$

where $\alpha(0 \le \alpha \le 1)$ is a constant to be determined so as to minimize the mean squared error of the proposed estimator $T_{|p}(NR)$.

2.3. Analysis

The properties of the proposed estimator $T_{|p}(NR)$ are derived under the following large sample approximations

$\bar{y}_{r_2} = \bar{Y}(1 + e_0)$, $\bar{y}_m = \bar{Y}(1 + e_1)$, $\bar{x}_m = \bar{X}(1 + e_2)$, $\bar{x}_{r_1} = \bar{X}(1 + e_3)$, $\bar{z}_i(r_2) = \bar{Z}_i(1 + e_{4i})$,

$\bar{z}_i(m) = \bar{Z}_i(1 + e_{5i})$, $\bar{z}_i(r_1) = \bar{Z}_i(1 + e_{6i})$, $\bar{x}_n = \bar{X}(1 + e_7)$, $\bar{z}_i(n) = \bar{Z}_i(1 + e_{8i})$ such that

$|e_k| < 1\ \forall$ i= 1, 2, ..., p and k = 0, 1, 2, 3,4, 5, 6, 7 and 8.

Under the above transformations, the estimators T_u and T_m take the following forms:

$$T(i, u) = \frac{\bar{Y}}{8}\left(8 + 8e_0 - 4e_{4i} - 4e_0e_{4i} + 3e_{4i}^2\right) \text{ for } i=1, 2, ..., p \tag{9}$$

$$\begin{aligned} T(i, m, n) = \frac{\bar{Y}}{8}\big(&8 + 8e_1 - 8e_2 + 8e_3 - 4e_{6i} - 8e_1e_2 + 8e_1e_3 - 4e_1e_{6i} - 8e_2e_3 \\ &+ 4e_2e_{6i} - 4e_3e_{6i} + 8e_2^2 + 3e_{6i}^2\big) \quad \text{for } i=1, 2, ..., p \end{aligned} \tag{10}$$

2.3.1. Bias and Mean Squared Error of the Estimator $T_{|p}(NR)$

The estimators T_u and T_m are exponential ratio and chain type ratio to exponential ratio type in nature respectively. Hence they are biased for population mean $\bar{Y}$. Therefore, the final estimator $T_{|p}(NR)$ defined in equation (8) is also biased estimator of $\bar{Y}$. The bias $B(.)$ and mean square error $M(.)$ of the proposed estimator $T_{|p}(NR)$ are obtained (ignoring finite population corrections) up to first order of approximations and thus we have following theorems:

Theorem 2.3.1: The bias of the proposed estimator $T_{|p}(NR)$ to the first order of approximation has been derived as

$$B\left(T_{|p}(NR)\right) = \alpha\, B(T_u) + (1-\alpha)\, B(T_m) \tag{11}$$

$$B(T_u) = \frac{1}{r_2}\mathbf{W}_u'\mathbf{B}_u \tag{12}$$

$$B(T_m) = \mathbf{W}_m'\left(\frac{1}{m}B_{m1} + \frac{1}{r_1}\mathbf{B}_{m2}\right) \tag{13}$$

where $\mathbf{B}_u = \left(B_1(u), B_2(u), \ldots, B_p(u)\right)'$, $\quad B_i(u) = \bar{Y}\left(\frac{3}{8}\frac{C_{002}}{\bar{Z}_i^2} - \frac{1}{2}\frac{C_{011}}{\bar{Y}\,\bar{Z}_i}\right)$,

for i =1, 2, 3, ..., p

$B_{m1} = \bar{Y}\left(\frac{C_{200}}{\bar{X}^2} - \frac{C_{110}}{\bar{X}\bar{Y}}\right)$, $\mathbf{B}_{m2} = \left(Bm_{21}, Bm_{22}, \ldots, Bm_{2p}\right)'$

where $Bm_{2i} = \bar{Y}\left(\frac{3}{8}\frac{C_{002}}{\bar{Z}_i^2} + \frac{C_{110}}{\bar{X}\bar{Y}} - \frac{1}{2}\frac{C_{011}}{\bar{Y}\bar{Z}_i} - \frac{C_{200}}{\bar{X}^2}\right)$, $C_{rst} = E\left[\left(x_i - \bar{X}\right)^r\left(y_i - \bar{Y}\right)^s\left(z_i - \bar{Z}\right)^t\right]$;

$(r, s, t) \geq 0$ for $i = 1, 2, 3, \ldots, p$.

Theorem 2.3.2: The mean squared error of the estimator $T_{|p}(NR)$ is obtained as

$$M\left(T_{|p}(NR)\right) = \alpha^2\, M(T_u) + (1-\alpha)^2\, M(T_m) + 2\alpha(1-\alpha)\mathrm{Cov}(T_u, T_m) \tag{14}$$

$$M(T_u) = \mathbf{W}_u'\,\Omega_u\mathbf{W}_u \tag{15}$$

$$M(T_m) = (B)\mathbf{W}_m'\,\mathbf{E}\,\mathbf{W}_m + \mathbf{W}_m'\,\Omega_m\mathbf{W}_m \tag{16}$$

where $\mathbf{W_u} = \left[w_{u_1} \ w_{u_2} \ ... \ w_{u_p} \right]'$, $\mathbf{W_m} = \left[w_{m_1} \ w_{m_2} \ ... \ w_{m_p} \right]'$, **E** is a unit matrix of order $p \times p$,

$\Omega_u = \left(\frac{1}{r_2} - \frac{1}{N}\right) \Omega_{u*}$, $\Omega_m = \left(\frac{1}{r_1} - \frac{1}{N}\right) \Omega_{m*}$ where

$$\Omega_{u*} = \begin{bmatrix} \Omega u_{11} & \Omega u_{12} & . & . & . & \Omega u_{1p} \\ \Omega u_{21} & \Omega u_{22} & . & . & . & \Omega u_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \Omega u_{p1} & \Omega u_{p2} & . & . & . & \Omega u_{pp} \end{bmatrix}_{p \times p} \quad \text{and} \quad \Omega_{m*} = \begin{bmatrix} \Omega m_{11} & \Omega m_{12} & . & . & . & \Omega m_{1p} \\ \Omega m_{21} & \Omega m_{22} & . & . & . & \Omega m_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \Omega m_{p1} & \Omega m_{p2} & . & . & . & \Omega m_{pp} \end{bmatrix}$$

where $B = \left(\frac{1}{m} - \frac{1}{N}\right) B_1, B_1 = 2\,\bar{Y}^2 \left(1 - \rho_{yx}\right) C_0^2$, $\Omega u_{ii} = \bar{Y}^2 \left(C_0^2 + \frac{1}{4} C_{z_i}^2 - \rho_{yz_i} C_0 C_{z_i} \right)$,

$\Omega u_{ij} = \bar{Y}^2 \left(C_0^2 - \frac{1}{2} \rho_{yz_i} C_0 C_{z_i} - \frac{1}{2} \rho_{yz_j} C_0 C_{z_j} + \frac{1}{4} \rho_{z_i z_j} C_{z_i} C_{z_j} \right)$,

$\Omega m_{ii} = \bar{Y}^2 \left(C_0^2 \left(2\rho_{yx} - 1\right) - \rho_{yz_i} C_0 C_{z_i} + \frac{1}{4} C_{z_i}^2 \right)$ and

$\Omega m_{ij} = \bar{Y}^2 \left(C_0^2 \left(2\rho_{yx} - 1\right) - \frac{1}{2} \rho_{yz_i} C_0 C_{z_i} - \frac{1}{2} \rho_{yz_j} C_0 C_{z_j} + \frac{1}{4} \rho_{z_i z_j} C_{z_i} C_{z_j} \right)$ $\forall\ i \neq j = 1, 2, 3, ..., p$.

and $\text{Cov}\left(T_u, T_m\right) = 0$ as they are based on two independent samples.

2.4. Choice of Optimal Weights

To find the optimization of the weight vector $\mathbf{W_u} = \left[w_{u_1} \ w_{u_2} \ ... \ w_{u_p} \right]'$, the mean squared error $M(T_u)$ given in equation (15) is minimized subject to the condition $\mathbf{1}'\mathbf{W}_u = 1$ using the method of Lagrange's Multiplier explained as:

To find the extrema using Lagrange's Multiplier Technique, we define L_u as

$$L_u = \mathbf{W}_u' \, \Omega_u \mathbf{W}_u - \lambda_u(\mathbf{1}'\mathbf{W}_u - 1), \tag{17}$$

where **1** is a unit column vector of order p and λ_u is the Lagrangian multiplier.

Now, by differentiating equation (17) partially with respect to $\mathbf{W}_u$ and equating it to zero we have

$$\frac{\partial L_u}{\partial \mathbf{W}_u} = \frac{\partial}{\partial \mathbf{W}_u} \left[\mathbf{W}_u' \, \Omega_u \, \mathbf{W}_u - \lambda_u(\mathbf{1}'\mathbf{W}_u - 1) \right] = 0$$

This implies that, $2\,\Omega_u \mathbf{W}_u - \lambda_u \mathbf{1} = \mathbf{0}$, which yields

$$\mathbf{W}_u = \frac{\lambda_u}{2}\,\Omega_u^{-1}\mathbf{1} \tag{18}$$

Now pre- multiplying equation (18) by $\mathbf{1}'$, we get

$$\frac{\lambda_u}{2} = \frac{1}{\mathbf{1}'\Omega_u^{-1}\mathbf{1}} \tag{19}$$

Thus, using equation (19) in equation (18), we obtain the optimal weight vector as

$$\mathbf{W}_{u_{opt.}} = \frac{\Omega_u^{-1}}{\mathbf{1}'\Omega_u^{-1}\mathbf{1}} \tag{20}$$

In similar manners, the optimal of the weight $\mathbf{W}_m = \left[w_{m_1}\ w_{m_2}\ \ldots w_{m_p}\right]'$, is obtained by minimizing $M(T_m)$ subject to the constraint $\mathbf{1}'\mathbf{W}_m = 1$ using the method of Lagrange's multiplier, for this we define

$$L_m = (B)\mathbf{W}_m'\,\mathbf{E}\,\mathbf{W}_m + \mathbf{W}_m'\,\Omega_m \mathbf{W}_m - \lambda_m(\mathbf{1}'\mathbf{W}_m - 1),$$

where λ_m is the Lagrangian multiplier.

Now, differentiating L_m with respect to $\mathbf{W}_m$ and equating to 0, we get

$$\mathbf{W}_{m_{opt.}} = \frac{\Omega_m^{-1}}{\mathbf{1}'\Omega_m^{-1}\mathbf{1}} \tag{21}$$

Then substituting the optimum values of $\mathbf{W}_u$ and $\mathbf{W}_m$ in equations (15) and (16) respectively, the optimum mean square errors of the estimators are obtained as:

$$M(T_u)_{opt.} = \left(\frac{1}{r_2} - \frac{1}{N}\right)\frac{1}{\mathbf{1}'\Omega_{u*}^{-1}\mathbf{1}} \tag{22}$$

$$M(T_m)_{opt.} = \left(\frac{1}{m} - \frac{1}{N}\right)B_1 + \left(\frac{1}{r_1} - \frac{1}{N}\right)\frac{1}{\mathbf{1}'\Omega_{m*}^{-1}\mathbf{1}} \tag{23}$$

2.5. Minimum Mean Squared Error of the Proposed Estimator $T_{|p}(NR)$

Since the mean squared error of the estimator $T_{|p}(NR)$ given in equation (14) is a function of unknown constant α, therefore, it has been minimized with respect to α and subsequently the optimum value of α is obtained as

$$\alpha_{opt.} = \frac{M(T_m)_{opt.}}{M(T_u)_{opt.} + M(T_m)_{opt.}} \tag{24}$$

Now substituting the values of $\alpha_{opt.}$ in equation (14), we obtain the optimum mean squared error of the estimator $T_{|p}(NR)$ as

$$M(T_{|p}(NR))^*_{opt.} = \frac{M(T_u)_{opt.} \cdot M(T_m)_{opt.}}{M(T_u)_{opt.} + M(T_m)_{opt.}} \tag{25}$$

Further, substituting the optimum values of the mean squared error of the estimator given in equations (22) and (23) in equation (24) and (25) respectively, the simplified values $\alpha_{opt.}$ and $M(T_{|p}(NR))^*_{opt.}$ are obtained as

$$\alpha_{opt.} = \frac{\mu\left[\mu f_2 C - (f_1 f_2 B_1 + f_2 C)\right]}{\left[\mu^2 f_2 C - \mu (f_1 f_2 B_1 + f_2 C - f_1 A) - f_1 A\right]} \tag{26}$$

$$M(T_{|p}(NR))^*_{opt.} = \frac{1}{n}\frac{\left[\mu C_1 - C_2\right]}{\left[\mu^2 C_3 - \mu C_4 - C_5\right]} \tag{27}$$

where $A = \frac{1}{\mathbf{1}'\Omega_{u*}^{-1}\mathbf{1}}, B_1 = 2\bar{Y}^2(1-\rho_{yx})C_0^2, C = \frac{1}{\mathbf{1}'\Omega_{m*}^{-1}\mathbf{1}}, C_1 = A\,C,\ C_2 = f_1 A\, B_1 + A\,C,$

$C_3 = f_2 C,\ C_4 = f_1 f_2 B_1 + f_2 C - f_1 A,\ C_5 = f_1 A$ and μ is the fraction of the sample drawn afresh at the current (second) wave.

2.6. Optimum Replacement Strategy for the Estimator $T_{|p}(NR)$

The idea of longitudinal surveys is mainly concerned with obtaining efficient estimates with minimal cost in carrying out the survey. So it is technically convenient to maintain a high overlap between repeats of the survey which provides the advantage due to many sampled units being located and have some experience in the survey. Hence the decision of the optimum value of μ should be made (fractions of samples to be drawn afresh on the current occasion) so that $\bar{Y}$ may be estimated with maximum precision and minimum cost, we minimize the mean squared error $M(T_{|p}(NR))^*_{opt.}$ in equation (27) with respect to μ and thus the optimum value of μ so obtained is one of the two roots given by

$$\hat{\mu} = \frac{D_2 \pm \sqrt{D_2^2 - D_1 D_3}}{D_1} \tag{28}$$

where $D_1 = C_1C_3$, $D_2 = C_2C_3$ and $D_3 = C_1C_5 + C_2C_4$.

The real value of μ exist, iff $D_2^2 - D_1D_3 \geq 0$. For any situation, which satisfies this condition, two real values of μ may be possible, hence choose a value of μ such that $0 \leq \mu \leq 1$. All other values of μ are inadmissible. If both the real values of μ are admissible, the lowest one will be the best choice as it reduces the total cost of the survey. Substituting the admissible value of μ say $\mu_{T_{|p}}^{NR}$ from (28) in to the equation (27), we get the optimum value of the mean squared error of the estimator $T_{|p}(NR)$ with respect to α as well as μ which, is given as

$$M\left(T_{|p}(NR)\right)_{opt.}^{**} = \frac{1}{n}\frac{\left[\mu_{T_{|p}}^{NR} C_1 - C_2\right]}{\left[\mu_{T_{|p}}^{NR^2} C_3 - \mu_{T_{|p}}^{NR} C_4 - C_5\right]} . \tag{29}$$

3. Special Cases

3.1. Case I: When Non-Response occurs only at the First (Previous) wave

When there is a presence of non-response only at first wave, the proposed estimator $T_{|p}(NR)$ for population mean $\bar{Y}$ reduces to

$$T_{|p}(P) = \varphi\, T_u^{\theta} + (1-\varphi) T_m \tag{30}$$

where $T_u^{\theta} = \mathbf{W}_{\mathbf{u}}'\mathbf{T}_{\exp}(u)$, where $\mathbf{W}_{\mathbf{u}}$ is a column vector of p-weights given by

$\mathbf{W}_{\mathbf{u}} = \left[w_{u_1}\ w_{u_2} \ldots w_{u_p}\right]'$

and $\mathbf{T}_{\exp}(u) = \begin{bmatrix} T(1,u) \\ T(2,u) \\ \vdots \\ T(p,u) \end{bmatrix}$, where $T(i, u) = \bar{y}_u \exp\left(\frac{\bar{Z}_i - \bar{z}_i(u)}{\bar{Z}_i + \bar{z}_i(u)}\right)$ for $i = 1, 2, 3, \ldots, p$

such that $\mathbf{1}'\mathbf{W_u} = 1$, where $\mathbf{1}$ is a column vector of order p and T_m is defined in equation (4) and $\varphi(0 \le \varphi \le 1)$ is a real constant to be determined so as to minimize the mean squared error of the estimator $T_{|p}(P)$.

In this case, the optimum value of fraction of sample drawn afresh is obtained as

$$\hat{\mu} = \left(H_2 \pm \sqrt{H_2^2 - H_1 H_3}\right) \Big/ H_1 = \mu_{T_{|p}}^{P} \text{ (say)}$$

and the minimum mean squared error of the estimator $T_{|p}(P)$ at the admissible value of $\hat{\mu}$ is derived as

$$M\left(T_{|p}(P)\right)^{**}_{opt.} = \left[\mu_{T_{|p}}^{P} G_1 - G_2\right] \Big/ n\left[\mu_{T_{|p}}^{P\,2} C - \mu_{T_{|p}}^{P} G_3 - G_4\right] \tag{31}$$

where

$H_1 = CG_1$, $H_2 = CG_2$, $H_3 = G_1G_4 + G_2G_3$, $G_1 = AC$, $G_2 = AC + f_1AB_1$, $G_3 = f_1B_1 + C - f_1A$,

$G_4 = f_1A$, $A = \dfrac{1}{\mathbf{1}'\Omega_{u*}^{-1}\mathbf{1}}$, $B_1 = 2\,\bar{Y}^2\left(1 - \rho_{yx}\right)C_0^2$, $C = \dfrac{1}{\mathbf{1}'\Omega_{m*}^{-1}\mathbf{1}}$ and $f_1 = r_1/n$.

3.2. Case II: When Non-Response occurs only at the Second (Current) wave

The estimator for population mean $\bar{Y}$ in the presence of non-response at current wave only, is given by

$$T_{|p}(C) = \psi\, T_u + (1-\psi)T_m^{\theta} \tag{32}$$

where $T_m^{\theta} = \mathbf{W_m'}\mathbf{T}_{\exp}(m, n)$, where $\mathbf{W_m}$ is a column vector of p-weights

$\mathbf{W_m} = \left[w_{m_1}\ w_{m_2} \ldots w_{m_p}\right]'$

and $\mathbf{T}_{\exp}(m,n) = \begin{bmatrix} T(1,m,n) \\ T(2,m,n) \\ \vdots \\ T(p,m,n) \end{bmatrix}$, where $T(i, m, n) = \left(\dfrac{\bar{y}^*(i, m)}{\bar{x}^*(i, m)}\right)\bar{x}^*(i, n)$

where $\bar{y}^*(i, m) = \bar{y}_m \exp\left(\dfrac{\bar{Z}_i - \bar{z}_i(m)}{\bar{Z}_i + \bar{z}_i(m)}\right)$, $\bar{x}^*(i, m) = \bar{x}_m \exp\left(\dfrac{\bar{Z}_i - \bar{z}_i(m)}{\bar{Z}_i + \bar{z}_i(m)}\right)$

and $\bar{x}^*(i, n) = \bar{x}_n \exp\left(\dfrac{\bar{Z}_i - \bar{z}_i(n)}{\bar{Z}_i + \bar{z}_i(n)}\right)$ for i=1, 2, 3, ..., p.

Such that $\mathbf{1}'\mathbf{W_m} = 1$, where $\mathbf{1}$ is a column vector of order p.

T_u has been defined in equation (3) and $\psi\left(0 \le \psi \le 1\right)$ is a real constant to be determined so as to minimize the mean squared error of the estimator $T_{|p}\left(C\right)$.

In this case, the optimum value of fraction of sample drawn afresh is obtained as

$$\hat{\mu} = \left(H_5 \pm \sqrt{H_5^2 - H_4\ H_6}\right) \Big/ H_4 = \mu^{C}_{T_{|p}}\left(\text{say}\right) \tag{33}$$

and the minimum mean squared error of the estimator $T_{|p}\left(C\right)$ at the admissible value of $\hat{\mu}$ is derived as

$$M\left(T_{|p}\left(C\right)\right)^{**}_{\text{opt.}} = \left[\mu^{C}_{T_{|p}}\ G_5 - G_6\right] \Big/ n\left[\mu^{C^2}_{T_{|p}}\ G_7 - \mu^{C}_{T_{|p}}\ G_8 - A\right] \tag{34}$$

where $H_4 = G_5G_7$, $H_5 = G_6G_7$, $H_6 = AG_5 + G_6G_8$, $G_5 = AC$, $G_6 = AB_1 + AC$, $G_7 = f_2C$, $G_8 = f_2B_1 + f_2C - A$, $A = \dfrac{1}{\mathbf{1}'\,\Omega_{u*}^{-1}\,\mathbf{1}}$, $B_1 = 2\ \bar{Y}^2\left(1 - \rho_{yx}\right)C_0^2$, $C = \dfrac{1}{\mathbf{1}'\,\Omega_{m*}^{-1}\,\mathbf{1}}$ and $f_2 = r_2/u$.

4. Efficiency with Increased Number of Auxiliary Variables

As we know that increasing the number of auxiliary variables typically increases the precision of the estimates. In this section here, we verify this property for the proposed estimator as under: Let $T_{|p}\left(NR\right)$ and $T_{|q}\left(NR\right)$ be two proposed estimators based on p and q auxiliary variables respectively such that $p < q$, then $M\left(T_{|p}\left(NR\right)\right) \ge M\left(T_{|q}\left(NR\right)\right)$, i.e.

$$M\left(T_{|p}\left(NR\right)\right) - M\left(T_{|q}\left(NR\right)\right) \ \ge\ 0 \tag{35}$$

$$\frac{1}{n}\frac{\left[\mu\, A_pC_p\ - A_p\ (B + C_p)\right]}{\left[\mu^2\ C_p\ - \mu\,(B + C_p\ + A_p) - A_p\right]} - \frac{1}{n}\frac{\left[\mu\, A_qC_q\ - A_q\ (B + C_q)\right]}{\left[\mu^2\ C_q\ - \mu\,(B + C_q\ + A_q) - A_q\right]} \ \ge\ 0$$

On simplification, we get

$$\left(A_p - A_q\right)\left[\left(\mu - 1\right)^2\left(\mu\, C_pC_q + \frac{A_pA_q\left(C_p - C_q\right)}{\left(A_p - A_q\right)}\right) - \mu\, B\left(\left(C_p - C_q\right)\left(\mu - 1\right) - B\right)\right] \ge 0$$

This reduces to the condition

$$\left(A_p - A_q\right) \ \ge\ 0 \tag{36}$$

So from Section 2.5 above, we get

$$\frac{1}{\mathbf{1}'\,\Omega_{p}^{-1}\,\mathbf{1}}-\frac{1}{\mathbf{1}'\,\Omega_{q}^{-1}\,\mathbf{1}}\ \geq 0$$

$$\mathbf{1}'\,\Omega_{q}^{-1}\,\mathbf{1} \geq \mathbf{1}'\,\Omega_{p}^{-1}\,\mathbf{1}$$

Following Rao (2006), the matrix Ω_q can be partitioned and can be written as

$$\Omega_q = \begin{pmatrix} \Omega_p & \mathbf{F} \\ \mathbf{F}' & \mathbf{G} \end{pmatrix}$$

where $\mathbf{F}$, $\mathbf{F}'$ and $\mathbf{G}$ are matrices deduced from Ω_q such that their order never exceeds q-p and always greater than or equal to 1.Then,

$$\Omega_q^{-1} = \begin{pmatrix} \Omega_p^{-1} + \mathbf{HJH}' & -\mathbf{HJ} \\ -\mathbf{JH}' & \mathbf{J} \end{pmatrix} \tag{37}$$

where $\mathbf{J} = \left(\mathbf{G} - \mathbf{F}'\Omega_p^{-1}\mathbf{F}\right)^{-1}$ and $\mathbf{H} = \Omega_p^{-1}\mathbf{F}$. (See Rao (2006) and Olkin (1958))

Now rewriting $\mathbf{1}'\Omega_q^{-1}\mathbf{1}$ by putting the value of Ω_q^{-1} from equation (37), we get

$$\mathbf{1}'\Omega_q^{-1}\mathbf{1} = \begin{pmatrix} \mathbf{1}_p & \mathbf{1}_{q-p} \end{pmatrix}' \begin{pmatrix} \Omega_p^{-1} + \mathbf{HJH}' & -\mathbf{HJ} \\ -\mathbf{JH}' & \mathbf{J} \end{pmatrix} \begin{pmatrix} \mathbf{1}_p \\ \mathbf{1}_{q-p} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{1}_p'\left(\Omega_p^{-1} + \mathbf{HJH}'\right) - \mathbf{1}_{q-p}'\mathbf{JH}' & -\mathbf{1}_p'\ \mathbf{HJ} + \mathbf{1}_{q-p}'\mathbf{J} \end{pmatrix} \begin{pmatrix} \mathbf{1}_p \\ \mathbf{1}_{q-p} \end{pmatrix}$$

$$= \mathbf{1}_p'\left(\Omega_p^{-1} + \mathbf{HJH}'\right)\mathbf{1}_p - \mathbf{1}_{q-p}'\mathbf{JH}'\,\mathbf{1}_p - \mathbf{1}_p'\ \mathbf{HJ}\ \mathbf{1}_{q-p} + \mathbf{1}_{q-p}'\mathbf{J}\ \mathbf{1}_{q-p}$$

$$\Rightarrow \mathbf{1}'\ \Omega_q^{-1}\ \mathbf{1} - \mathbf{1}_p'\ \left(\Omega_p^{-1}\right)\ \mathbf{1}_p = \mathbf{1}_p'\left(\mathbf{HJH}'\right)\mathbf{1}_p - \mathbf{1}_{q-p}'\mathbf{JH}'\,\mathbf{1}_p - \mathbf{1}_p'\ \mathbf{HJ}\ \mathbf{1}_{q-p} + \mathbf{1}_{q-p}'\mathbf{J}\ \mathbf{1}_{q-p}$$

$$\mathbf{1}'\ \Omega_q^{-1}\ \mathbf{1} - \mathbf{1}_p'\ \left(\Omega_p^{-1}\right)\ \mathbf{1}_p = \begin{pmatrix} \mathbf{1}_p & \mathbf{1}_{q-p} \end{pmatrix}' \begin{pmatrix} \mathbf{HJH}' & -\mathbf{HJ} \\ -\mathbf{JH}' & \mathbf{J} \end{pmatrix} \begin{pmatrix} \mathbf{1}_p \\ \mathbf{1}_{q-p} \end{pmatrix}$$

$$\mathbf{1}'\ \Omega_q^{-1}\ \mathbf{1} - \mathbf{1}_p'\ \left(\Omega_p^{-1}\right)\ \mathbf{1}_p = \mathbf{1}' \begin{pmatrix} \mathbf{H} \\ -\mathbf{I} \end{pmatrix} \mathbf{J}\ \begin{pmatrix} \mathbf{H} & -\mathbf{I} \end{pmatrix}\ \mathbf{1}\ \geq\ 0$$

The latter follows since **J** is positive definite so that $\mathbf{R}'\mathbf{J}\ \mathbf{R}\ \geq\ 0$ for all **R,**

where $\mathbf{R} = \begin{pmatrix} \mathbf{H} & -\mathbf{I} \end{pmatrix}\ \mathbf{1}$.

Hence from equation (35), we have

$$\mathrm{M}\left(\mathrm{T}_{|p}\left(\mathrm{NR}\right)\right) - \mathrm{M}\left(\mathrm{T}_{|q}\left(\mathrm{NR}\right)\right)\ \geq\ 0$$

This leads to the result that utilizing more auxiliary variables provides more efficient estimates in terms of mean squared error for the proposed estimator.

5. Complexity in compliance with multi-collinearity

Case 1: when the p-auxiliary variables are mutually uncorrelated i.e. $\rho_{z_i z_j} = 0$ $\forall\ i \neq j = 1, 2, ..., p$, the proposed multivariate estimators are applicable and in this case optimum value for μ say $\mu_{\oplus}^{NR}$, $\mu_{\oplus}^{P}$ and $\mu_{\oplus}^{C}$ and the mean squared error of the estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ with respect to φ and μ are obtained as

$$\mu_{\oplus}^{NR} = \left(D_2^{\oplus} \pm \sqrt{D_2^{\oplus^2} - D_1^{\oplus} D_3^{\oplus}}\right) \Big/ D_1^{\oplus}$$

$$\mu_{\oplus}^{P} = \left(H_2^{\oplus} \pm \sqrt{H_2^{\oplus^2} - H_1^{\oplus} H_3^{\oplus}}\right) \Big/ H_1^{\oplus}$$

$$\mu_{\oplus}^{C} = \left(H_5^{\oplus} \pm \sqrt{H_5^{\oplus^2} - H_4^{\oplus} H_6^{\oplus}}\right) \Big/ H_4^{\oplus}$$

$$M\left(T_{|p}(NR)\right)_{opt.}^{**} = \left[\mu_{\oplus}^{NR} C_1^{\oplus} - C_2^{\oplus}\right] \Big/ n\left[\mu_{\oplus}^{NR^2} C_3^{\oplus} - \mu_{T_{|p}}^{NR} C_4^{\oplus} - C_5^{\oplus}\right]$$

$$M\left(T_{|p}(P)\right)_{opt.}^{**} = \left[\mu_{\oplus}^{P} G_1^{\oplus} - G_2^{\oplus}\right] \Big/ n\left[\mu_{\oplus}^{P^2} C^{\oplus} - \mu_{\oplus}^{P} G_3^{\oplus} - G_4^{\oplus}\right]$$

$$M\left(T_{|p}(C)\right)_{opt.}^{**} = \left[\mu_{\oplus}^{C} G_5^{\oplus} - G_6^{\oplus}\right] \Big/ n\left[\mu_{\oplus}^{C^2} G_7^{\oplus} - \mu_{\oplus}^{C} G_8^{\oplus} - A^{\oplus}\right]$$

$D_1^{\oplus} = C_1^{\oplus} C_3^{\oplus}$, $D_2^{\oplus} = C_2^{\oplus} C_3^{\oplus}$, $D_3^{\oplus} = C_1^{\oplus} C_5^{\oplus} + C_2^{\oplus} C_4^{\oplus}$, $H_1^{\oplus} = C^{\oplus} G_1^{\oplus}$, $H_2^{\oplus} = C^{\oplus} G_2^{\oplus}$, $H_3^{\oplus} = G_1^{\oplus} G_4^{\oplus} + G_2^{\oplus} G_3^{\oplus}$,

$G_1^{\oplus} = A^{\oplus} C^{\oplus}$, $G_2^{\oplus} = A^{\oplus} C^{\oplus} + f_1 A^{\oplus} B_1^{\oplus}$, $G_3^{\oplus} = f_1 B_1^{\oplus} + C^{\oplus} - f_1 A^{\oplus}$, $G_1^{\oplus} = f_1 A^{\oplus}$, $H_4^{\oplus} = G_5^{\oplus} G_7^{\oplus}$, $H_5^{\oplus} = G_6^{\oplus} G_7^{\oplus}$,

$H_6^{\oplus} = A^{\oplus} G_5^{\oplus} + G_6^{\oplus} G_8^{\oplus}$, $G_5^{\oplus} = A^{\oplus} C^{\oplus}$, $G_6^{\oplus} = A^{\oplus} B_1^{\oplus} + A^{\oplus} C^{\oplus}$, $G_7^{\oplus} = f_2 C^{\oplus}$, $G_8^{\oplus} = f_2 B_1^{\oplus} + f_2 C^{\oplus} - A^{\oplus}$,

$A^{\oplus} = \frac{1}{\mathbf{1}'\Delta_{u*}^{-1}\mathbf{1}}$, $B_1^{\oplus} = 2\,\bar{Y}^2\left(1 - \rho_{yx}\right) C_0^2$, $C^{\oplus} = \frac{1}{\mathbf{1}'\Delta_{m*}^{-1}\mathbf{1}}$, $f_1 = r_1/n$ and $f_2 = r_2/u$.

where

$$\Delta_{u*} = \begin{bmatrix} \Delta u_{11} & \Delta u_{12} & . & . & . & \Delta u_{1p} \\ \Delta u_{21} & \Delta u_{22} & . & . & . & \Delta u_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \Delta u_{p1} & \Delta u_{p2} & . & . & . & \Delta u_{pp} \end{bmatrix}_{p \times p} \quad \text{and} \quad \Delta_{m*} = \begin{bmatrix} \Delta m_{11} & \Delta m_{12} & . & . & . & \Delta m_{1p} \\ \Delta m_{21} & \Delta m_{22} & . & . & . & \Delta m_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \Delta m_{p1} & \Delta m_{p2} & . & . & . & \Delta m_{pp} \end{bmatrix}$$

$$\Delta u_{ii} = \bar{Y}^2\left(C_0^2 + \frac{1}{4}C_{z_i}^2 - \rho_{yz_i}C_0C_{z_i}\right),\ \Delta u_{ij} = \bar{Y}^2\left(C_0^2 - \frac{1}{2}\rho_{yz_i}C_0C_{z_i} - \frac{1}{2}\rho_{yz_j}C_0C_{z_j}\right),$$

$$\Delta m_{ii} = \bar{Y}^2\left(C_0^2\left(2\rho_{yx} - 1\right) - \rho_{yz_i}C_0C_{z_i} + \frac{1}{4}C_{z_i}^2\right) \text{ and}$$

$$\Delta m_{ij} = \bar{Y}^2\left(C_0^2\left(2\rho_{yx} - 1\right) - \frac{1}{2}\rho_{yz_i}C_0C_{z_i} - \frac{1}{2}\rho_{yz_j}C_0C_{z_j}\right)\ \forall\ i \neq j = 1, 2, 3, ..., p.$$

Case 2: When the p-auxiliary variables are mutually correlated i.e. $\rho_{z_iz_j} \neq 0$ $\forall\ i \neq j = 1, 2, ..., p$. In this case if there is high correlation between p-auxiliary variates, then such a problem can be addressed as a problem of multi-collinearity in survey sampling.

6. Efficiency Comparison

In order to examine the performance of the proposed estimators under non-response, the estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ have been compared to the estimator $T_{|p}$ due to Priyanka et al. (2015).

Hence, following Olkin (1958), Raj (1965), Artes and Garcia (2005) and Singh et al. (2011) we consider $C_0 = C_{z_i}; \forall\ i = 1, 2, 3, ..., p$ approximately and hence, the optimum value of μ for the case (i) when non-response occurs on both the occasion, (ii) non-response occurs only at first occasion (iii) non-response occurs only at second occasion, say $\hat{\mu}^{NR}_{T_{|p}}$, $\hat{\mu}^{P}_{T_{|p}}$ and $\hat{\mu}^{C}_{T_{|p}}$ and optimum value of mean squared errors $M\left(T_{|p}(NR)\right)^{**}_{opt.}$, $M\left(T_{|p}(P)\right)^{**}_{opt.}$ and $M\left(T_{|p}(C)\right)^{**}_{opt.}$ of the proposed estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ reduce to

The optimum value of μ is given by

$$\mu^{NR^*}_{T_{|p}} = \left(D_2^* \pm \sqrt{D_2^{*2} - D_1^*\ D_3^*}\right) \Big/ D_1^*$$

(38)

$$\mu^{P^*}_{T_{|p}} = \left(H_2^* \pm \sqrt{H_2^{*2} - H_1^*\ H_3^*}\right) \Big/ H_1^* \qquad (40)$$

$$\mu^{C^*}_{T_{|p}} = \left(H_5^* \pm \sqrt{H_5^{*2} - H_4^*\ H_6^*}\right) \Big/ H_4^* \qquad (42)$$

and the optimum mean squared error of the estimators given as

$$M\left(T_{|p}(NR)\right)^{***}_{opt.} = \left[\mu^{NR^*}_{T_{|p}} C^*_1 - C^*_2\right] \Big/ n\left[\mu^{NR^{*2}}_{T_{|p}} C^*_3 - \mu^{NR^*}_{T_{|p}} C^*_4 - C^*_5\right] \quad (39)$$

$$M\left(T_{|p}(P)\right)^{***}_{opt.} = \left[\mu^{P^*}_{T_{|p}} G^*_1 - G^*_2\right] \Big/ n\left[\mu^{P^{*2}}_{T_{|p}} C^* - \mu^{P^*}_{T_{|p}} G^*_3 - G^*_4\right] \quad (41)$$

$$M\left(T_{|p}(C)\right)^{***}_{opt.} = \left[\mu^{C^*}_{T_{|p}} G^*_5 - G^*_6\right] \Big/ n\left[\mu^{C^{*2}}_{T_{|p}} G^*_7 - \mu^{C^*}_{T_{|p}} G^*_8 - A^*\right] \quad (43)$$

where $D^*_1 = C^*_1 C^*_3,\ D^*_2 = C^*_2 C^*_3,\ D^*_3 = C^*_1 C^*_5 + C^*_2 C^*_4,\ C^*_1 = A^* C^*,\ C^*_2 = f_1 A^* B^*_1 + A^* C^*,\ C^*_3 = f_2 C^*,$

$C^*_4 = f_1 f_2 B^*_1 + f_2 C^* - f_1 A^*,\ C^*_5 = f_1 A^*,\ H^*_1 = C^* G^*_1,\ H^*_2 = C^* G^*_2,\ H^*_3 = G^*_1 G^*_4 + G^*_2 G^*_3,\ G^*_1 = A^* C^*,$

$G^*_2 = A^* C^* + f_1 A^* B^*_1,\ G^*_3 = f_1 B^*_1 + C^* - f_1 A^*,\ G^*_4 = f_1 A^*,\ H^*_4 = G^*_5 G^*_7,\ H^*_5 = G^*_6 G^*_7,\ H^*_6 = A^* G^*_5 + G^*_6 G^*_8,$

$G^*_5 = A^* C^*,\ G^*_6 = A^* B^*_1 + A^* C^*,\ G^*_7 = f_2 C^*,\ G^*_8 = f_2 B^*_1 + f_2 C^* - A^*,\ f_1 = r_1/n,\ A^* = \dfrac{1}{1' Ł^{-1}_{u*} 1},$

$B^*_1 = 2\left(1 - \rho_{yx}\right) S^2_y \quad , C^* = \dfrac{1}{1' Ł^{-1}_{m*} 1}$

$$Ł_{u*} = \begin{bmatrix} Łu_{11} & Łu_{12} & . & . & . & Łu_{1p} \\ Łu_{21} & Łu_{22} & . & . & . & Łu_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ Łu_{p1} & Łu_{p2} & . & . & . & Łu_{pp} \end{bmatrix}_{p \times p} \quad \text{and} \quad Ł_{m*} = \begin{bmatrix} Łm_{11} & Łm_{12} & . & . & . & Łm_{1p} \\ Łm_{21} & Łm_{22} & . & . & . & Łm_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ Łm_{p1} & Łm_{p2} & . & . & . & Łm_{pp} \end{bmatrix}_{p \times p}$$

$B^*_1 = 2\left(1 - \rho_{yx}\right) S^2_y,\ Łu_{ii} = \left(\dfrac{5}{4} - \rho_{yz_i}\right) S^2_y,\ Łu_{ij} = \left(1 - \dfrac{1}{2}\rho_{yz_i} - \dfrac{1}{2}\rho_{yz_j} + \dfrac{1}{4}\rho_{z_i z_j}\right) S^2_y,$

$Łm_{ii} = \left(2\rho_{yx} - \rho_{yz_i} - \dfrac{3}{4}\right) S^2_y$ and $Łm_{ij} = \left(2\rho_{yx} - \dfrac{1}{2}\rho_{yz_i} - \dfrac{1}{2}\rho_{yz_j} + \dfrac{1}{4}\rho_{z_i z_j} - 1\right) S^2_y \ \forall\ i \neq j = 1, 2, 3, ..., p.$

6.1 Comparison of the proposed estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ with respect to estimator $T_{|p}(PR)$ due to Priyanka et al. (2015)

The estimator $T_{|p}(PR)$ proposed by Priyanka et al. (2015) is given as

$$T_{|p}(PR) = \xi\, T^{\theta}_u + (1 - \xi)\, T^{\theta}_m \quad (44)$$

where T_u^{θ} and T_m^{θ} are discussed in equation (30) and (32) respectively and $\xi\left(0 \le \xi \le 1\right)$ is a real constant to be determined so as to minimize the mean squared error of the estimator $T_{|p}\left(PR\right)$.

The optimum value of fraction of sample to be drawn afresh say $\mu_{T_{|p}}^{PR}$ and optimum value of mean squared error $M\left(T_{|p}\left(PR\right)\right)_{opt.}^{**}$ of the estimator $T_{|p}\left(PR\right)$ is given by

$$\mu_{T_{|p}}^{PR} = \frac{J_2^* \pm \sqrt{J_2^{*2} - J_1^* J_3^*}}{J_1^*} \tag{45}$$

$$M\left(T_{|p}\left(PR\right)\right)_{opt.}^{**} = \frac{1}{n}\frac{\left[\mu_{T_{|p}}^{PR} I_1^* - I_2^*\right]}{\left[\mu_{T_{|p}}^{PR^2} C^0 - \mu^*_{T_{|p}} I_3^* - A^0\right]} \tag{46}$$

where $J_1^* = C^0 I_1^*$, $J_2^* = C^0 I_2^*$, $J_3^* = A^0 I_1^* + I_2^* I_3^*$, $A^0 = \frac{1}{\mathbf{1}'\mathbf{H}_{u*}^{-1}\mathbf{1}}$, $C^0 = \frac{1}{\mathbf{1}'\mathbf{H}_{m*}^{-1}\mathbf{1}}$, $I_1^* = A^0 C^0$,
$I_2^* = A^0 B_1^* + A^0 C^0$, $I_3^* = B_1^* + C^0 - A^0$, $B_1^* = 2\left(1 - \rho_{yx}\right)S_y^2$,

$$\mathbf{H}_{u*} = \begin{bmatrix} hu_{11} & hu_{12} & . & . & . & hu_{1p} \\ hu_{21} & hu_{22} & . & . & . & hu_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ hu_{p1} & hu_{p2} & . & . & . & hu_{pp} \end{bmatrix}_{p \times p} \quad \text{and} \quad \mathbf{H}_{m*} = \begin{bmatrix} hm_{11} & hm_{12} & . & . & . & hm_{1p} \\ hm_{21} & hm_{22} & . & . & . & hm_{2p} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ hm_{p1} & hm_{p2} & . & . & . & hm_{pp} \end{bmatrix}_{p \times p}$$

$B_1^* = 2\left(1 - \rho_{yx}\right)S_y^2$, $hu_{ii} = \left(\frac{5}{4} - \rho_{yz_i}\right) S_y^2$, $hu_{ij} = \left(1 - \frac{1}{2}\rho_{yz_i} - \frac{1}{2}\rho_{yz_j} + \frac{1}{4}\rho_{z_i z_j}\right) S_y^2$,

$hm_{ii} = \left(2\rho_{yx} - \rho_{yz_i} - \frac{3}{4}\right) S_y^2$ and $hm_{ij} = \left(2\rho_{yx} - \frac{1}{2}\rho_{yz_i} - \frac{1}{2}\rho_{yz_j} + \frac{1}{4}\rho_{z_i z_j} - 1\right) S_y^2 \ \forall\ i \ne j = 1, 2, 3, ..., p.$

The percent relative loss in precision of the proposed estimators $T_{|p}\left(NR\right)$, $T_{|p}\left(P\right)$ and $T_{|p}\left(C\right)$ have been recorded to infer about the effect of incompleteness in the data over the successive waves with respect to the estimator $T_{|p}\left(PR\right)$ and are given under their respective optimal conditions as

$$\left.\begin{aligned} L_{|p}^{0} &= \frac{M\left(T_{|p}(NR)\right)_{opt.}^{***} - M\left(T_{|p}(PR)\right)_{opt.}^{**}}{M\left(T_{|p}(NR)\right)_{opt.}^{***}} \times 100, \\ L_{|p}^{1} &= \frac{M\left(T_{|p}(P)\right)_{opt.}^{***} - M\left(T_{|p}(PR)\right)_{opt.}^{**}}{M\left(T_{|p}(P)\right)_{opt.}^{***}} \times 100, \\ L_{|p}^{2} &= \frac{M\left(T_{|p}(C)\right)_{opt.}^{***} - M\left(T_{|p}(PR)\right)_{opt.}^{**}}{M\left(T_{|p}(C)\right)_{opt.}^{***}} \times 100 . \end{aligned}\right\} \quad (47)$$

7. Numerical Illustrations and Monte Carlo Simulation

Empirical validation has been carried out by Monte Carlo Simulation. Real life situation of completely known finite population has been considered.

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

For carrying out numerical illustration we have considered the case of three auxiliary information (i.e. p=3) which are stable over time and are available at both the occasions. The population comprise of N = 51 states of the United States. Let

y_j: The total energy consumption during 2007 in the j^{th} state of U. S.

x_j: The total energy consumption during 2002 in the j^{th} state of U. S.

z_{1j}: The total energy consumption during 2001 in the j^{th} state of U. S.

z_{2j}: The total energy consumption during 2000 in the j^{th} state of U. S.

z_{3j}: The total energy consumption during 1999 in the j^{th} state of U. S.

For the considered population, the values of μ defined in equations (38), (40) and (42) and the percent relative loss in efficiencies $L_1^0, L_2^0, L_3^0, L_1^1, L_2^1, L_3^1, L_1^2, L_2^2$ and L_3^2 defined in equation (47) of the estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ for p=1, 2 and 3, respectively with respect to estimator $T_{|p}(PR)$ have been computed and are presented in Table 1.

To judge about the performance of the estimator in the presence of different percentages of non-response, a more general illustration has been worked out by considering choices of correlation coefficients of study and auxiliary variables on different waves. These results have been shown in Table-2 to Table-6.

To validate the above empirical results, Monte Carlo simulation has also been performed for the considered population. For convenience the choices of t_1 and t_2 are considered 0.20 and 0.30 respectively. The simulation results are shown in Table-7 to Table-9.

7.1 Simulation Algorithm

(i) Choose 5000 samples of size n=25 using simple random sampling without replacement on first wave for both the study and auxiliary variables.

(ii) For $f_1 = 0.80$, choose $r_1 = 20$ responding units out of n=25 samples units.

(iii) Calculate sample mean $\bar{x}_{r_1\,|\,k}$, $\bar{z}_{1_{|k}}(r_1)$, $\bar{z}_{2_{|k}}(r_1)$ and $\bar{z}_{3_{|k}}(r_1)$ for k =1, 2, - - -, 5000.

(iv) Retain m=15 units out of each $r_1 = 20$ sample units of the study and auxiliary variables at the first wave.

(v) Calculate sample mean $\bar{x}_{m\,|\,k}$, $\bar{z}_{1_{m|k}}$, $\bar{z}_{2_{m|k}}$ and $\bar{z}_{3_{m|k}}$ for k= 1, 2, - - -, 5000.

(vi) Select u=10 units using simple random sampling without replacement from N-n=26 units of the population for study and auxiliary variables at second (current) wave.

(vii) For $f_2 = 0.70$, choose $r_2 = 7$ responding units out of u=10 samples units.

(viii) Calculate sample mean $\bar{y}_{r_2\,|\,k}$, $\bar{y}_{m\,|\,k}$, $\bar{z}_{1_{|k}}(r_2)$, $\bar{z}_{2_{|k}}(r_2)$ and $\bar{z}_{3_{|k}}(r_2)$ for k = 1, 2, - - -, 5000.

(ix) Iterate the parameter α from 0.1 to 0.9 with a step of 0.1.

(x) Iterate ξ from 0.1 to 0.9 with a step of 0.1 within (ix).

(xi) Calculate the percent relative loss in efficiencies of the proposed estimator $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ with respect to estimator respect to $T_{|p}(PR)$ for p=1, 2 and 3 as

$$L_0(p) = \frac{\sum_{k=1}^{5000}\left[T_{|p|k}(NR) - T_{|p|k}(PR)\right]^2}{\sum_{k=1}^{5000}\left[T_{|p|k}(NR)\right]^2} \times 100\ ,\quad L_1(p) = \frac{\sum_{k=1}^{5000}\left[T_{|p|k}(P) - T_{|p|k}(PR)\right]^2}{\sum_{k=1}^{5000}\left[T_{|p|k}(P)\right]^2} \times 100\ ,$$

and $$L_2(p) = \frac{\sum_{k=1}^{5000} \left[T_{|p|k}(C) - T_{|p|k}(PR) \right]^2}{\sum_{k=1}^{5000} \left[T_{|p|k}(C) \right]^2} \times 100, \quad k=1, 2, ..., 5000.$$

Similarly, the algorithm has been modified for the case when non-response occurs only at previous wave or only at current wave.

Table 1: **Empirical Comparison of the proposed estimators** $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ **with respect to the estimator** $T_{|p}(PR)$.

	$\mu^{PR}_{T_{|p}}$	$\mu^{NR}_{T_{|p}}$	$\mu^{P}_{T_{|p}}$	$\mu^{C}_{T_{|p}}$	$L^0_{|p}$	$L^1_{|p}$	$L^2_{|p}$
p=1	0.5355	0.5555	0.6284	0.4444	4.0912	1.3253	2.2902
p=2	0.5196	0.4831	0.6157	0.3539	3.4008	0.7533	2.1065
p=3	0.5137	0.4212	0.6109	0.2765	3.0501	0.5326	1.9302

Table 2: Percent relative loss when estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ **are compared to the estimator** $T_{|p}(PR)$ **for p=1.**

t_1	t_2	ρ_{yx}	0.6							0.8						
		ρ_{yz_1}	μ_1^{NR}	μ_1^{P}	μ_1^{C}	μ_1^{PR}	L_1^0	L_1^1	L_1^2	μ_1^{NR}	μ_1^{P}	μ_1^{C}	μ_1^{PR}	L_1^0	L_1^1	L_1^2
0.05	0.05	0.6	0.61	0.50	0.59	0.47	2.67	-0.27	2.82	0.53	0.58	0.50	0.56	3.02	0.53	2.38
		0.7	0.54	0.48	0.52	0.45	2.31	-0.51	2.68	0.48	0.56	0.45	0.53	2.79	0.36	2.31
	0.10	0.6	0.74	0.50	0.73	0.47	6.27	-0.27	6.33	0.47	0.58	0.45	0.56	5.27	0.53	4.52
		0.7	0.61	0.48	0.59	0.45	5.51	-0.51	5.77	0.40	0.56	0.37	0.53	4.85	0.36	4.23
0.10	0.05	0.6	0.63	0.52	0.59	0.47	2.52	-0.54	2.82	0.55	0.60	0.50	0.56	3.66	0.77	2.38
		0.7	0.57	0.50	0.52	0.45	1.94	-1.03	2.68	0.51	0.58	0.45	0.53	3.28	0.73	2.31
	0.10	0.6	0.76	0.52	0.73	0.47	6.20	-0.54	6.33	0.50	0.60	0.45	0.56	6.03	0.77	4.52
		0.7	0.63	0.50	0.59	0.45	5.26	-1.03	5.77	0.43	0.58	0.37	0.53	5.47	0.73	4.23
0.15	0.05	0.6	0.65	0.55	0.59	0.47	2.37	-0.82	2.82	0.58	0.62	0.50	0.56	4.31	1.61	2.38
		0.7	0.59	0.53	0.52	0.45	1.56	-1.54	2.68	0.53	0.60	0.45	0.53	3.76	1.10	2.31
	0.10	0.6	0.77	0.55	0.73	0.47	6.14	-0.82	6.33	0.53	0.62	0.45	0.56	6.79	1.61	4.52
		0.7	0.65	0.53	0.59	0.45	5.01	-1.54	5.77	0.46	0.60	0.37	0.53	6.09	1.10	4.23

Table 3: Percent relative loss when estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ are compared to the estimator $T_{|p}(PR)$ for p=2 and $\rho_{z_1z_2} = 0$

		ρ_{yx}		**0.6**							**0.8**						
t_1	t_2	ρ_{yz_1}	ρ_{yz_2}	μ_2^{NR}	μ_2^{P}	μ_2^{C}	μ_2^{PR}	L_2^0	L_2^1	L_2^2	μ_2^{NR}	μ_2^{P}	μ_2^{C}	μ_2^{PR}	L_2^0	L_2^1	L_2^2
0.10	0.05	0.4	0.3	0.48	0.57	0.42	0.53	3.08	0.58	2.25	0.58	0.62	0.54	0.58	4.00	1.39	2.41
			0.4	0.44	0.57	0.38	0.52	2.88	0.45	2.17	0.57	0.61	0.53	0.57	3.88	1.28	2.40
			0.5	0.35	0.56	0.27	0.51	2.49	0.26	1.93	0.56	0.60	0.51	0.56	3.73	1.13	2.39
		0.5	0.3	0.42	0.56	0.35	0.52	2.77	0.38	2.11	0.57	0.61	0.52	0.57	3.83	1.23	2.40
			0.4	0.35	0.56	0.27	0.51	2.49	0.26	1.93	0.56	0.60	0.51	0.56	3.73	1.13	2.39
			0.5	*	0.55	*	0.50	*	0.10	*	0.54	0.60	0.49	0.55	3.58	1.00	2.37
	0.10	0.4	0.3	0.38	0.57	0.31	0.53	5.13	0.58	4.00	0.55	0.62	0.50	0.58	6.45	1.39	4.66
			0.4	0.31	0.57	0.23	0.52	4.71	0.45	3.67	0.53	0.61	0.48	0.57	6.31	1.28	4.61
			0.5	0.12	0.56	*	0.51	3.65	0.26	*	0.51	0.60	0.46	0.56	6.12	1.13	4.55
		0.5	0.3	0.26	0.56	*	0.52	4.48	0.38	*	0.52	0.61	0.47	0.57	6.24	1.23	4.59
			0.4	0.12	0.56	*	0.51	3.65	0.26	*	0.51	0.60	0.46	0.56	6.12	1.13	4.55
			0.5	*	0.55	*	0.55	*	0.10	*	0.49	0.60	0.43	0.55	5.92	1.00	4.47
0.15	0.05	0.4	0.3	0.51	0.60	0.42	0.53	3.50	0.87	2.25	0.61	0.64	0.54	0.58	4.79	2.09	2.41
			0.4	0.47	0.59	0.38	0.52	3.24	0.67	2.17	0.60	0.63	0.53	0.57	4.62	1.92	2.40
			0.5	0.38	0.58	0.27	0.51	2.77	0.40	1.93	0.58	0.62	0.51	0.56	4.40	1.70	2.39
		0.5	0.3	0.45	0.59	0.35	0.52	3.10	0.58	2.11	0.59	0.63	0.52	0.57	4.55	1.85	2.40
			0.4	0.38	0.58	0.27	0.51	2.77	0.40	1.93	0.58	0.62	0.51	0.56	4.40	1.70	2.39
			0.5	*	0.57	*	0.50	*	0.15	*	0.57	0.62	0.49	0.55	4.19	1.50	2.37
	0.10	0.4	0.3	0.41	0.60	0.31	0.53	5.69	0.87	4.00	0.57	0.64	0.50	0.58	7.34	2.09	4.66
			0.4	0.35	0.59	0.23	0.52	5.23	0.67	3.67	0.56	0.63	0.48	0.57	7.16	1.92	4.61
			0.5	*	0.58	*	0.51	*	0.40	*	0.54	0.62	0.46	0.56	6.90	1.70	4.55
		0.5	0.3	0.30	0.59	*	0.52	4.94	0.58	*	0.55	0.63	0.47	0.57	7.07	1.85	4.59
			0.4	0.17	0.58	*	0.51	4.13	0.40	*	0.54	0.62	0.46	0.56	6.90	1.70	4.55
			0.5	*	0.57	*	0.50	*	0.15	*	0.52	0.62	0.43	0.55	6.65	1.50	4.47

Note:" *" denotes that percent relative loss cannot be obtained since μ_2^{NR} , μ_2^{P} and μ_2^{C} do not exist.

Table 4: Percent relative loss when estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ are compared to the estimator $T_{|p}(PR)$ for p=2 and $\rho_{z_1z_2} > 0$

		ρ_{yx}		**0.7**							0.8						
t_1	t_2	ρ_{yz_1}	ρ_{yz_2}	μ_2^{NR}	μ_2^{P}	μ_2^{C}	μ_2^{PR}	L_2^0	L_2^1	L_2^2	μ_2^{NR}	μ_2^{P}	μ_2^{C}	μ_2^{PR}	L_2^0	L_2^1	L_2^2
0.10	0.05	0.4	0.3	0.51	0.58	0.45	0.54	3.28	0.74	2.31	0.59	0.63	0.55	0.58	4.13	1.52	2.42
			0.4	0.49	0.58	0.43	0.53	3.15	0.63	2.27	0.59	0.62	0.54	0.58	4.04	1.43	2.42
			0.5	0.45	0.57	0.38	0.52	2.90	0.45	2.18	0.57	0.61	0.53	0.57	3.89	1.29	2.40
		0.5	0.3	0.46	0.57	0.40	0.52	2.99	0.52	2.22	0.58	0.61	0.53	0.57	3.94	1.34	2.41
			0.4	0.45	0.57	0.38	0.52	2.90	0.45	2.18	0.57	0.61	0.53	0.57	3.89	1.29	2.40
			0.5	0.39	0.56	0.32	0.51	2.65	0.32	2.04	0.56	0.61	0.51	0.56	3.78	1.18	2.39
	0.10	0.4	0.3	0.43	0.58	0.37	0.54	5.48	0.74	4.24	0.56	0.63	0.51	0.58	6.60	1.52	4.69
			0.4	0.40	0.58	0.33	0.53	5.25	0.63	4.09	0.55	0.62	0.50	0.58	6.49	1.43	4.67
			0.5	0.31	0.57	0.24	0.52	4.74	0.45	3.69	0.53	0.61	0.48	0.57	6.32	1.29	4.62
		0.5	0.3	0.35	0.57	0.28	0.52	4.94	0.52	3.86	0.54	0.61	0.49	0.57	6.38	1.34	4.64
			0.4	0.31	0.57	0.24	0.52	4.74	0.45	3.69	0.53	0.61	0.48	0.57	6.32	1.29	4.62
			0.5	0.20	0.56	0.12	0.51	4.11	0.32	3.13	0.52	0.61	0.46	0.56	6.18	1.18	4.57
0.15	0.05	0.4	0.3	0.54	0.60	0.45	0.54	3.77	1.11	2.31	0.62	0.65	0.55	0.58	4.98	2.28	2.42
			0.4	0.52	0.60	0.43	0.53	3.59	0.95	2.27	0.61	0.64	0.54	0.58	4.85	2.15	2.42
			0.5	0.48	0.59	0.38	0.52	3.26	0.68	2.18	0.60	0.63	0.53	0.57	4.63	1.93	2.40
		0.5	0.3	0.49	0.59	0.40	0.52	3.38	0.78	2.22	0.60	0.64	0.53	0.57	4.71	2.01	2.41
			0.4	0.48	0.59	0.38	0.52	3.26	0.68	2.18	0.60	0.63	0.53	0.57	4.63	1.93	2.40
			0.5	0.42	0.58	0.32	0.51	2.95	0.49	2.04	0.59	0.63	0.51	0.56	4.47	1.77	2.39
	0.10	0.4	0.3	0.46	0.60	0.37	0.54	6.11	1.11	4.24	0.58	0.65	0.51	0.58	7.55	2.28	4.69
			0.4	0.43	0.60	0.33	0.53	5.83	0.95	4.09	0.57	0.64	0.50	0.58	7.41	2.15	4.67
			0.5	0.35	0.59	0.24	0.52	5.26	0.68	3.69	0.56	0.63	0.48	0.57	7.17	1.93	4.62
		0.5	0.3	0.38	0.59	0.28	0.52	5.48	0.78	3.86	0.56	0.64	0.49	0.57	7.25	2.01	4.64
			0.4	0.35	0.59	0.24	0.52	5.26	0.68	3.69	0.56	0.63	0.48	0.57	7.17	1.93	4.62
			0.5	0.25	0.58	0.12	0.51	4.60	0.49	3.13	0.54	0.63	0.46	0.56	6.99	1.77	4.57

Table 5: Percent relative loss when estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ are compared to the estimator $T_{|p}(PR)$ for p=3 and $\rho_{z_i z_j} = 0,\ i \neq j=1,\ 2$ and 3

t_1	t_2	ρ_{yx}	ρ_{yz_1}	ρ_{yz_2}	ρ_{yz_3}	μ_3^{NR}	μ_3^{P}	μ_3^{C}	μ_3^{PR}	L_3^0	L_3^1	L_3^2
0.10	0.05	0.6	0.5	0.6	0.7	0.53	0.48	0.47	0.43	1.37	-1.57	2.63
		0.7	0.5	0.6	0.7	0.61	0.52	0.56	0.46	2.33	-0.68	2.76
		0.7	0.4	0.6	0.5	0.91	0.54	0.90	0.49	3.52	-0.16	3.53
	0.10	0.6	0.5	0.6	0.7	0.57	0.48	0.53	0.43	4.58	-1.57	5.55
		0.7	0.5	0.6	0.7	0.70	0.52	0.67	0.46	5.85	-0.68	6.08
		0.7	0.4	0.6	0.5	*	0.54	*	0.49	*	-0.16	*
0.15	0.05	0.6	0.5	0.6	0.7	0.55	0.51	0.47	0.43	0.75	-2.35	2.63
		0.7	0.5	0.6	0.7	0.63	0.54	0.56	0.46	2.11	-1.03	2.76
		0.7	0.4	0.6	0.5	0.91	0.56	0.90	0.49	3.52	-0.24	3.53
	0.10	0.6	0.5	0.6	0.7	0.60	0.51	0.53	0.43	4.09	-2.35	5.55
		0.7	0.5	0.6	0.7	0.72	0.54	0.67	0.46	5.74	-1.03	6.08
		0.7	0.4	0.6	0.5	*	0.56	*	0.49	*	-0.24	*

Note:" *" denotes that percent relative loss cannot be obtained since μ_3^{NR} and μ_3^{C} do not exist.

Table 6: Percent relative loss when estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ are compared to the estimator $T_{|p}(PR)$ for p=3 and $\rho_{z_i z_j} > 0,\ i \neq j=1,\ 2$ and 3

t_1	t_2	ρ_{yx}	ρ_{yz_1}	ρ_{yz_2}	ρ_{yz_3}	$\rho_{z_1z_2}$	$\rho_{z_1z_3}$	$\rho_{z_2z_3}$	μ_3^{NR}	μ_3^{P}	μ_3^{C}	μ_3^{PR}	L_3^0	L_3^1	L_3^2
0.10	0.05	0.6	0.2	0.5	0.3	0.5	0.5	0.4	0.78	0.53	0.76	0.48	3.13	-0.24	3.18
		0.7	0.4	0.7	0.6	0.3	0.3	0.4	0.68	0.53	0.64	0.48	2.75	-0.40	2.93
		0.7	0.3	0.7	0.5	0.3	0.4	0.2	0.65	0.52	0.61	0.47	2.58	-0.50	2.85
	0.10	0.6	0.2	0.5	0.3	0.5	0.5	0.4	*	0.53	*	0.48	*	-0.24	*
		0.7	0.4	0.7	0.6	0.3	0.3	0.4	0.84	0.53	0.83	0.48	6.73	-0.40	6.77
		0.7	0.3	0.7	0.5	0.3	0.4	0.2	0.78	0.52	0.75	0.47	6.34	-0.50	6.44
0.15	0.05	0.6	0.2	0.5	0.3	0.5	0.5	0.4	0.79	0.56	0.76	0.48	3.11	-0.36	3.18
		0.7	0.4	0.7	0.6	0.3	0.3	0.4	0.70	0.55	0.64	0.48	2.67	-0.60	2.93
		0.7	0.3	0.7	0.5	0.3	0.4	0.2	0.66	0.55	0.61	0.47	2.45	-0.75	2.85
	0.10	0.6	0.2	0.5	0.3	0.5	0.5	0.4	*	0.56	*	0.48	*	-0.36	*
		0.7	0.4	0.7	0.6	0.3	0.3	0.4	0.85	0.55	0.83	0.48	6.72	-0.60	6.77
		0.7	0.3	0.7	0.5	0.3	0.4	0.2	0.79	0.55	0.75	0.47	6.29	-0.75	6.44

Note:" *" denotes that percent relative loss cannot be obtained since μ_3^{NR} and μ_3^{C} do not exist.

Table 7: Simulation result when the proposed estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ are compared with the estimator $T_{|p}(PR)$ for p=1.

α \ ξ		**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
0.1	$L_0(1)$	4.69	25.85	39.22	42.75	35.11	21.05	-3.15	-38.59
	$L_1(1)$	0.90	24.37	38.66	43.07	37.62	18.89	-0.75	-35.43
	$L_2(1)$	1.37	23.89	37.14	40.14	33.76	20.21	-4.67	-44.45
0.2	$L_0(1)$	-5.40	18.02	32.87	36.19	28.29	11.87	-15.82	-53.04
	$L_1(1)$	-11.75	14.60	31.12	35.49	27.73	10.08	-14.42	-54.92
	$L_2(1)$	-9.06	15.91	30.94	34.16	27.34	11.71	-16.95	-55.56
0.3	$L_0(1)$	-12.92	13.30	28.21	32.15	23.77	6.90	-22.95	-62.45
	$L_1(1)$	-24.09	5.72	23.99	28.33	19.73	1.35	-26.10	-71.40
	$L_2(1)$	-15.88	10.62	26.39	30.32	23.21	6.20	-23.77	-66.00
0.4	$L_0(1)$	-15.69	11.10	26.75	3.46	22.51	5.16	-24.35	-64.80
	$L_1(1)$	-33.76	-0.23	18.17	22.29	14.24	-5.97	-35.43	-84.45
	$L_2(1)$	-17.57	9.26	25.01	29.10	22.02	4.51	-25.51	-68.47
0.5	$L_0(1)$	-12.78	13.65	28.57	32.70	24.59	7.90	-20.67	-60.31
	$L_1(1)$	-36.59	-2.44	15.71	19.71	12.28	-8.83	-39.94	-90.22
	$L_2(1)$	-14.00	12.23	27.18	31.49	24.51	7.40	-22.14	-63.46
0.6	$L_0(1)$	-5.17	19.70	33.68	37.35	30.23	14.65	-12.06	-49.41
	$L_1(1)$	-34.57	-0.71	16.64	21.06	13.75	-7.25	-38.24	-87.08
	$L_2(1)$	-5.76	18.22	32.40	36.48	30.07	14.18	-13.05	-51.25
0.7	$L_0(1)$	5.83	28.04	40.16	43.61	37.56	23.33	-0.43	-34.37
	$L_1(1)$	-27.38	4.60	20.89	25.80	18.71	-1.02	-31.38	-76.74
	$L_2(1)$	5.42	26.94	39.52	43.82	37.28	22.85	-1.43	-35.29
0.8	$L_0(1)$	17.30	37.15	47.39	50.65	45.16	32.86	11.74	-17.69
	$L_1(1)$	-16.36	13.69	27.39	31.64	25.31	7.78	-20.21	-61.48
	$L_2(1)$	17.50	36.13	47.02	50.10	45.13	32.54	11.38	-18.25
0.9	$L_0(1)$	28.57	45.51	54.08	57.30	52.75	41.73	23.51	-1.34
	$L_1(1)$	-3.40	23.42	35.85	39.00	33.86	18.01	-7.06	-43.40
	$L_2(1)$	26.87	45.02	54.26	56.88	52.69	41.83	23.11	-1.29

Table 8: **Simulation result when the proposed estimators** $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ **are compared with the** estimator $T_{|p}(PR)$ for p=2

α \ ξ		**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
0.1	$L_0(2)$	3.24	24.82	38.69	42.08	35.69	21.83	-1.95	-36.46
	$L_1(2)$	-1.59	22.49	37.08	43.83	39.29	20.32	-0.75	-34.47
	$L_2(2)$	1.38	23.81	37.82	41.41	35.36	22.41	-0.19	-38.59
0.2	$L_0(2)$	-7.60	17.11	31.87	35.82	28.84	12.59	-13.83	-49.94
	$L_1(2)$	-13.68	12.97	30.11	35.17	28.76	10.84	-14.24	-53.05
	$L_2(2)$	-9.37	15.88	31.40	35.26	28.87	14.26	-12.41	-49.97
0.3	$L_0(2)$	-15.10	11.60	27.23	31.73	24.14	7.60	-21.35	-59.74
	$L_1(2)$	-26.02	3.52	22.67	27.85	20.82	1.92	-25.45	-69.23
	$L_2(2)$	-16.43	10.28	26.64	31.020	24.79	8.49	-19.25	-60.42
0.4	$L_0(2)$	-17.88	9.23	25.61	29.89	22.75	5.74	-23.06	-62.48
	$L_1(2)$	-36.76	-2.83	16.62	21.80	15.04	-4.95	-34.53	-82.91
	$L_2(2)$	-18.66	8.56	24.92	29.37	23.34	6.60	-21.84	-63.60
0.5	$L_0(2)$	-15.12	11.83	27.05	32.06	24.67	8.19	-19.67	-58.42
	$L_1(2)$	-39.90	-5.36	13.95	19.09	12.82	-8.08	-39.02	-88.85
	$L_2(2)$	-15.49	11.05	26.74	31.71	25.38	9.06	-19.14	-59.58
0.6	$L_0(2)$	-7.84	17.69	32.61	36.60	30.14	14.78	-11.44	-48.09
	$L_1(2)$	-38.71	-3.76	14.68	20.25	14.09	-6.85	-37.53	-85.90
	$L_2(2)$	-7.93	16.69	31.76	36.45	30.63	15.33	-10.88	-48.29
0.7	$L_0(2)$	3.20	26.17	38.64	42.80	37.39	23.38	0.60	-33.38
	$L_1(2)$	-31.54	1.65	19.08	24.97	18.34	-0.74	-30.85	-75.83
	$L_2(2)$	3.15	25.11	38.55	42.82	37.59	23.57	0.13	-33.45
0.8	$L_0(2)$	14.97	35.38	46.39	49.91	44.92	32.73	12.04	-17.13
	$L_1(2)$	-20.15	10.42	25.66	30.94	25.35	7.99	-19.99	-61.94
	$L_2(2)$	15.34	34.31	45.93	49.66	45.32	32.88	12.32	-16.99
0.9	$L_0(2)$	26.39	43.96	53.49	56.63	52.54	41.62	23.70	-1.10
	$L_1(2)$	-6.90	20.34	34.31	38.47	33.12	18.02	-7.06	-42.84
	$L_2(2)$	26.88	43.21	53.19	56.38	52.74	41.89	24.11	-2.25

Table 9: Simulation result when the proposed estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ are compared with the estimator $T_{|p}(PR)$ for p=3.

α \ ξ		**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
0.1	$L_0(3)$	2.51	22.50	36.19	39.97	33.42	21.13	-0.29	-34.02
	$L_1(3)$	-3.89	19.77	33.48	41.28	36.68	19.54	-0.66	-30.84
	$L_2(3)$	1.29	22.31	35.72	39.85	34.12	22.20	2.02	-33.06
0.2	$L_0(3)$	-7.60	15.29	29.35	33.48	26.93	12.23	-11.64	-45.36
	$L_1(3)$	-14.96	10.96	26.74	32.38	26.77	10.64	-13.30	-49.21
	$L_2(3)$	-8.82	14.76	29.46	33.64	27.96	14.62	-9.17	-43.45
0.3	$L_0(3)$	-14.95	10.05	25.14	29.50	22.44	7.32	-18.53	-54.15
	$L_1(3)$	-26.38	1.93	19.67	25.42	19.20	2.58	-23.37	-63.34
	$L_2(3)$	-15.36	9.48	24.92	29.77	24.11	9.18	-15.49	-53.02
0.4	$L_0(3)$	-17.73	7.84	23.46	27.72	21.21	5.72	-20.17	-56.53
	$L_1(3)$	-35.88	-3.83	14.06	19.73	13.92	-3.76	-31.18	-75.58
	$L_2(3)$	17.66	7.84	23.24	28.27	22.74	7.30	-17.95	-55.93
0.5	$L_0(3)$	-14.93	10.50	24.84	29.87	23.13	8.13	-17.18	-52.58
	$L_1(3)$	-39.43	-6.53	11.55	17.07	11.92	-6.71	-35.23	-80.90
	$L_2(3)$	-14.63	10.03	24.82	30.09	24.54	9.46	-15.76	-52.64
0.6	$L_0(3)$	-8.17	16.06	29.66	34.30	28.42	14.39	-9.57	-43.11
	$L_1(3)$	-38.44	-4.96	12.25	17.93	12.99	-5.75	-34.12	-78.27
	$L_2(3)$	-7.84	15.26	29.66	34.56	29.42	15.31	-8.25	-42.64
0.7	$L_0(3)$	2.48	24.18	36.07	40.36	35.54	22.60	1.15	-29.81
	$L_1(3)$	-31.86	-0.07	16.25	22.35	16.98	-0.31	-28.18	-69.46
	$L_2(3)$	2.57	23.25	36.19	40.61	36.02	22.97	1.70	-29.37
0.8	$L_0(3)$	13.67	33.16	43.42	47.33	42.91	31.63	12.41	-14.81
	$L_1(3)$	-21.40	8.02	22.58	28.10	23.38	7.59	-18.46	-56.30
	$L_2(3)$	14.12	32.13	43.36	47.34	43.49	31.89	12.95	-14.44
0.9	$L_0(3)$	24.71	41.54	50.92	54.06	50.49	40.24	23.47	0.12
	$L_1(3)$	-8.81	17.42	31.13	35.43	30.86	16.97	-6.53	-39.93
	$L_2(3)$	25.31	40.84	50.66	54.03	50.82	40.58	24.07	0.20

8. Rendition of Results

The performance of an estimator in successive sampling in the presence of non-response is generally judged on the basis of percent relative loss in efficiency (lesser is loss better is the estimator) and in terms of optimum value of fraction of fresh sample to be drawn on current (second) wave which in turns is directly associated to the cost of survey. Following interpretation can be drawn from Tables 1- 9,

8.1 Results based on empirical study for the considered population

1) From Table-1, it is observed that the Optimum values $\mu_{T_{|p}}^{NR}$, $\mu_{T_{|p}}^{P}$ and $\mu_{T_{|p}}^{C}$ for the estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ respectively exist for the considered Population also $\mu_3^{NR} < \mu_2^{NR} < \mu_1^{NR}$, $\mu_3^{P} < \mu_2^{P} < \mu_1^{P}$ and $\mu_3^{C} < \mu_2^{C} < \mu_1^{C}$, which justifies the applicability of the proposed estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ at optimum conditions. This indicates that a smaller fresh sample is required when more number of auxiliary variables is used.

2) We also observe that $\mu_{T_{|p}}^{PR} < \mu_{T_{|p}}^{P}$ for p=1, 2 and 3. This is probably because in successive sampling we try to reduce the fraction of sample to be drawn afresh and make most use of the information available from previous occasion but in the case when non-response occurs only at previous occasion then for compensating the absence of response at first occasion, more fraction of fresh sample is required.

3) We also see that $L_3^0 < L_2^0 < L_1^0$, $L_3^1 < L_2^1 < L_1^1$ and $L_3^2 < L_2^2 < L_1^2$, which supports the fact that utilization of more number of auxiliary variables decreases the percent relative loss in precision when compared to the estimator due to Priyanka et al. (2015).

8.2 Results extracted from general scenario i.e. by considering different choices of correlation coefficients

1) From Table 2 to Table 6, we see that, even for low correlation coefficients of study and auxiliary variables, the proposed estimators work efficiently and provide lesser loss although as the correlation between the study and auxiliary increases (whether the auxiliary variables are mutually correlated or uncorrelated), the amount of percent relative loss decreases for the estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ when compared with the estimator $T_{|p}(PR)$.

2) From Table 2 to Table 6, we observe that for fixed fraction t_1 if we increase the fraction t_2, the percent relative loss $L^0_{|p}$, p=1, 2 3, decreases and if t_2 is kept fixed and we keep on increasing the fraction t_1, the percent relative loss $L^0_{|p}$, p=1, 2, 3, increases.

3) From Table 2 to Table 6, we see that for increasing correlation between study and auxiliary variables if t_1 is kept fixed the percent relative loss $L^1_{|p}$, p=1, 2, 3, decreases but if we even increases t_1, the percent relative loss $L^1_{|p}$, p=1, 2, 3, also increases.

4)From Table 2 to Table 6, we can infer that if t_2 is kept fixed then increasing the amount of correlation between study and auxiliary variable results in lower percent relative loss $L^2_{|p}$ for p=1, 2, 3 but if we increases t_2 then the percent relative loss also increases.

8.3 Results based on simulation study

1) From simulation results in Table 7 to Table 9 we observe that for fixed choices of α, the percent relative loss $L_0(p)$, $L_1(p)$ and $L_2(p)$ (p=1, 2 and 3) increases initially and start to decrease as ξ is increased when the proposed estimators are compared to the estimator due to Priyanka et al. (2015).

2) Also it is observed that for fixed choices of ξ, the value of $L_0(p)$, $L_1(p)$ and $L_2(p)$ (p=1, 2 and 3) decreases initially and start to increase as α is increased when the proposed estimators are compared to the estimator due to Priyanka et al. (2015).

3) It is also observed that when the proposed estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ utilize more number of auxiliary variable, the percent relative loss $L_0(p)$, $L_1(p)$ and $L_2(p)$ p=1, 2, 3 are observed to have decreasing trend which signifies the use of more number of auxiliary variables.

9. Conclusion

The thorough analysis of proposed multivariate weighted estimators utilizing information on multi-auxiliary variables in the presence of non-response with variety of cases depending upon the occurrence of non-response, seems to be interesting enough as an amalgamation of exponential structure with ratio type estimator because even in the midst of non- response, the proposed method of imputation not just provides lesser percent relative loss in efficiency of the estimator but it also helps in reducing the cost of survey as far as possible when a comparative study is carried out with respect to estimator $T_{|p}(PR)$. Therefore, the proposed estimators $T_{|p}(NR)$, $T_{|p}(P)$ and $T_{|p}(C)$ can be considered for their practical use in the presence of non-response, if any, on successive waves by survey practitioners.

CHAPTER – 11*

Cogitation of Incompleteness in the midst of Imputation in Longitudinal Surveys for Population Mean

---

* Following is the publication based on the work of this chapter:--

1. Priyanka, K. and Mittal, R. (2016): Cogitation of Incompleteness in the midst of Imputation in Longitudinal Surveys for Population Mean. Proceedings of National Conference on ***RSCTA***, held in Ramanujan College, University of Delhi on 11-12th Mach, 2016. (Accepted)

Cogitation of Incompleteness in the midst of Imputation in Longitudinal Surveys for Population Mean

1. Introduction

Sophisticated sample surveys are being designed to prevent the non-response of sample units but it is hard to prevent completely due to the pure stochastic nature of incompleteness. Missing data makes the analysis more miserable when the data has to be collected and analysed on more than one occasion. The problem of sampling on two successive occasions was initiated by Jessen (1942), and latter this idea was explored by Patterson (1950), Narain (1953), Eckler (1955), Gordon (1983), Arnab and Okafor (1992), Feng and Zou (1997), Singh and Singh (2001), Singh and Priyanka (2008a), Singh et al. (2013a), Bandyopadhyay and Singh (2014), Priyanka and Mittal (2014, 2015a, 2015b), Priyanka et al. (2015) and many others.

Longitudinal surveys are mainly about observing characteristics on more than one chance (occasion) so that the dynamics of the characteristic could be understood over a period so as to infer about the behaviors and patterns. In this process a variety of literature has been put on using many explanative twists, definitely enriching the field of study and a vast literature is available for dealing with non-response while sampling over successive occasion.One may cite Rubin (1976), Sande (1979), Kalton et al. (1981), Kalton and Kasprzyk (1982), Singh and Singh (1991) by considering complete data set and discarding all those units for which information was not available for at least one time. Also Lee et al. (1994, 1995), Singh and Horn (2002), Ahmed et al. (2006), Singh and Priyanka (2007b), Singh (2009) and Singh et al. (2013b) can be seen for various new estimators for estimation of parameters by method of imputation using additional auxiliary information in successive sampling.

Various ideas have been dug into conceiving that auxiliary information utilized remains stable in nature while sampling over successive occasions but when difference (gap) between two occasions is sufficiently large, the nature of auxiliary variable may not sustain to be stable. In such a case the nature of auxiliary characteristic turns to be dynamic over the period of observation. Soa completely fresh approach has been made using imputation technique while sampling over two successive occasions to negotiate with the ill effects of non-response. MCAR has been assumed implicitly and a more worthy estimator for population mean while sampling over successive occasion using additional auxiliary information which is changing (dynamic) over the period of observation, by imputing missing data in the presence of non-response. The properties of the proposed estimator have been elaborated theoretically considering that (i) non-response may arise on both occasions, (ii) it may occur only at first occasion or (iii) it may occur only at second occasion while comparing the proposed estimator with estimator having complete response, proposed by Priyanka and Mittal (2016). A Simulation study has also been put through to substantiate the practicability of the proposed estimator.

2. Survey Design and Analysis

2.1 Notations

Let $U = (U_1, U_2, ..., U_N)$ be the N- element finite population, which has been sampled over two occasions. The characters under study is denoted by x(y) on the first (second) occasion, respectively. It is assumed that information on a dynamic (varying) auxiliary variable $z_1(z_2)$, with the known population mean, is available on first (second) occasion. We assume that there is non-response at both the occasions. A simple random sample without replacement s_n of n units has been drawn on the first occasion. Let the number of responding unit out of n sampled units, which are drawn at the first occasion, be denoted by r_1, the set of responding units in s_n by R_1 and that of non-responding by R_1^c. A random sub-sample s_m of $m = n\lambda$ unit is retained (matched) for its use on the current (second) occasion from the units which responded (r_1) at the first occasion and it is intuitive that these matched units will be completely responding at the current (second) occasion as well. A fresh simple random sample (without replacement), s_u of $u = n - m = n\mu$ units, is drawn on the second occasion from the non-sampled units of the population so that the sample size on the second occasion remains the same i.e. n. Let the number of responding units out of u sampled units which are drawn afresh at current occasion, be denoted by r_2, the set of responding unit in s_u by R_2, and that of non-responding units by R_2^c. λ and μ $(\lambda + \mu = 1)$ are the fractions of matched and fresh sample, respectively, at the current(second) occasion. For every unit $i \in R_j\,(j = 1, 2)$, the values $x_i\,(y_i)$ are observed, but for the units $i \in R_j^c\,(j = 1, 2)$ the values $x_i\,(y_i)$ are missing and instead imputed values are derived. The following notations have been used hereafter:

$\bar{X}, \bar{Y}, \bar{Z}_1, \bar{Z}_2$: Population means of the variables x, y, z_1 and z_2 respectively.

$\bar{y}_u, \bar{z}_u, \bar{y}_{r_2}, \bar{z}_2(r_2), \bar{x}_m, \bar{y}_m, \bar{z}_1(m), \bar{z}_2(m), \bar{x}_{r_1}, \bar{z}_1(r_1)$: Sample mean of respective variate based on the sample sizes shown in suffice.

$\rho_{yx}, \rho_{xz_1}, \rho_{xz_2}, \rho_{yz_1}, \rho_{yz_2}, \rho_{z_1z_2}$: Correlation coefficient between the variables shown in suffices.

$S_x^2, S_y^2, S_{z_1}^2, S_{z_2}^2$: Population mean squared of variables x, y, z_1 and z_2 respectively.

$f_1 = \left(\frac{r_1}{n}\right), f_2 = \left(\frac{r_2}{u}\right)$: The fraction of respondents at first and second occasions respectively.

$t_1 = (1 - f_1), t_2 = (1 - f_2)$: The fraction of non- respondents at first and second occasions respectively.

2.2. Formulation of the Proposed Estimator T

To estimate the population mean $\bar{Y}$ on the current (second) occasion, an estimator T_u has been proposed considering that non-response occurs at current occasion andthe missing values occurring in the sample of size u are replaced by imputed values. Hence, the following imputation method has been proposed to cope up with the problem of non-response in sample s_u :

$$y_{\bullet i} = \begin{cases} y_i & \text{if} \quad i \in R_2 \\ \frac{1}{u - r_2}\left\{u\,\bar{y}_{r_2}\,\exp\left(\frac{\bar{Z}_2 - \bar{z}_2(r_2)}{\bar{Z}_2 + \bar{z}_2(r_2)}\right) - r_2\,\bar{y}_{r_2}\right\} & \text{if} \quad i \in R_2^c \end{cases} \tag{1}$$

where $\bar{y}_{r_2} = \frac{1}{r_2}\sum_{i \in R_2} y_i$ and $\bar{z}_2(r_2) = \frac{1}{r_2}\sum_{i \in R_2} z_{2_i}$.

and hence the estimator for $\bar{Y}$ is given by

$$T_u = \bar{y}_{r_2}\,\exp\left(\frac{\bar{Z}_2 - \bar{z}_2(r_2)}{\bar{Z}_2 + \bar{z}_2(r_2)}\right) \tag{2}$$

The second estimator T_m is based on sample size $m = n\lambda$ common to the both occasions utilizing information retained from first occasion. Since non- response is assumed to be occurring on first occasion as well so the missing values occurring in the sample of size n are replaced by imputed values. The following imputation technique has been suggested

$$x_{\bullet i} = \begin{cases} x_i & \text{if} \quad i \in R_1 \\ \frac{1}{n - r_1}\left\{n\,\bar{x}_{r_1}\,\exp\left(\frac{\bar{Z}_1 - \bar{z}_1(r_1)}{\bar{Z}_1 + \bar{z}_1(r_1)}\right) - r_1\,\bar{x}_{r_1}\right\} & \text{if} \quad i \in R_1^c \end{cases} \tag{3}$$

where $\bar{x}_{r_1} = \frac{1}{r_1}\sum_{i \in R_1} x_i$ and $\bar{z}_1(r_1) = \frac{1}{r_1}\sum_{i \in R_1} z_{1_i}$.

Considering above proposed imputation method the estimator based on sample s_n is altered to

$$\bar{x}_n^* = \frac{1}{n}\sum_{i \in s_n} x_{\bullet i} = \bar{x}_{r_1}\,\exp\left(\frac{\bar{Z}_1 - \bar{z}_1(r_1)}{\bar{Z}_1 + \bar{z}_1(r_1)}\right) \tag{4}$$

Therefore, the estimator based on sample size m common to both occasions which utilizes the missing values by above method of imputation is given by

$$T_m = \bar{y}_m^* \left(\frac{\bar{x}_n^*}{\bar{x}_m^*} \right) \tag{5}$$

where $\bar{y}_m^* = \bar{y}_m \ \exp\left(\frac{\bar{Z}_2 - \bar{z}_2(m)}{\bar{Z}_2 + \bar{z}_2(m)} \right)$, $\bar{x}_m^* = \bar{x}_m \ \exp\left(\frac{\bar{Z}_1 - \bar{z}_1(m)}{\bar{Z}_1 + \bar{z}_1(m)} \right)$ and

$\bar{x}_n^* = \bar{x}_{r_1} \ \exp\left(\frac{\bar{Z}_1 - \bar{z}_1(r_1)}{\bar{Z}_1 + \bar{z}_1(r_1)} \right)$.

Considering the convex combination of the two estimators T_u and T_m, we have the final estimator of population mean $\bar{Y}$ on the current occasion as

$$T = \alpha \ T_u + (1 - \alpha) T_m \tag{6}$$

where $\alpha(0 \le \alpha \le 1)$ is a constant to be determined so as to minimize the mean squared error of the proposed estimators T.

2.3. Properties of the Proposed Estimators T

The properties of the proposed estimators T are derived under the following large sample approximations

$\bar{y}_{r_2} = \bar{Y}(1 + e_0)$, $\bar{y}_m = \bar{Y}(1 + e_1)$, $\bar{x}_m = \bar{X}(1 + e_2)$, $\bar{x}_{r_1} = \bar{X}(1 + e_3)$, $\bar{z}_2(r_2) = \bar{Z}_2(1 + e_4)$,
$\bar{z}_2(m) = \bar{Z}_2(1 + e_5)$, $\bar{z}_1(m) = \bar{Z}_1(1 + e_6)$ and $\bar{z}_1(r_1) = \bar{Z}_1(1 + e_7)$ such that $|e_i| < 1 \ \forall \ i = 0, ..., 7$.

2.4. Bias and Mean Squared Error of the Estimators T

The estimators T_u and T_m are exponential ratio and chain type ratio to exponential ratio type in nature respectively. Hence they are biased for population mean $\bar{Y}$. Therefore, the final estimator T defined in equation (6) is also biased estimator of $\bar{Y}$. The bias $B(.)$ and mean squared error $M(.)$ of the proposed estimator T are obtained (ignoring finite population corrections) and thus we have following theorems:

Theorem 2.4.1.Bias of the estimator T to the first order of approximations is obtained as

$$B(T) = \alpha\, B(T_u) + (1-\alpha)\, B(T_m) \tag{7}$$

where
$$B(T_u) = \frac{1}{r_2}\bar{Y}\left(\frac{3}{8}\frac{C_{0002}}{\bar{Z}_2^2} - \frac{1}{2}\frac{C_{0101}}{\bar{Y}\,\bar{Z}_2}\right) \tag{8}$$

and

$$\begin{aligned}B(T_m) = \bar{Y}\Bigg(&\frac{1}{m}\left(\frac{C_{2000}}{\bar{X}^2} - \frac{1}{8}\frac{C_{0020}}{\bar{Z}_1^2} + \frac{3}{8}\frac{C_{0002}}{\bar{Z}_2^2} - \frac{C_{1100}}{\bar{X}\bar{Y}} - \frac{1}{2}\frac{C_{1010}}{\bar{X}\bar{Z}_1} + \frac{1}{2}\frac{C_{1001}}{\bar{X}\bar{Z}_2} + \frac{1}{2}\frac{C_{0110}}{\bar{Y}\bar{Z}_1} - \frac{1}{2}\frac{C_{0101}}{\bar{Y}\bar{Z}_2} - \frac{1}{4}\frac{C_{0011}}{\bar{Z}_1\bar{Z}_2}\right)\\ &+ \frac{1}{r_1}\left(\frac{1}{8}\frac{C_{0020}}{\bar{Z}_1^2} - \frac{C_{2000}}{\bar{X}^2} + \frac{C_{1100}}{\bar{X}\bar{Y}} + \frac{1}{2}\frac{C_{1010}}{\bar{X}\bar{Z}_1} - \frac{1}{2}\frac{C_{1001}}{\bar{X}\bar{Z}_2} - \frac{1}{2}\frac{C_{0110}}{\bar{Y}\bar{Z}_1} + \frac{1}{4}\frac{C_{0011}}{\bar{Z}_1\bar{Z}_2}\right)\Bigg)\end{aligned} \tag{9}$$

where $C_{rstq} = E\left[(x_i - \bar{X})^r (y_i - \bar{Y})^s (z_{1_i} - \bar{Z}_1)^t (z_{2_i} - \bar{Z}_2)^q\right];\ (r, s, t, q) \geq 0$.

Theorem 2.4.2. Mean squared error of the estimator T to the first order of approximations is obtained as

$$M(T) = \alpha^2\, M(T_u) + (1-\alpha)^2\, M(T_m) + 2\,\alpha(1-\alpha)\,Cov(T_u, T_m) \tag{10}$$

$$M(T_u) = \frac{1}{r_2}A_1\, S_y^2 \tag{11}$$

$$M(T_m) = \left(\frac{1}{m}A_2 + \frac{1}{r_1}A_3\right) S_y^2 \tag{12}$$

where $A_1 = (5/4) - \rho_{yz_2}$, $A_2 = (5/2) - 2\,\rho_{yx} + \rho_{yz_1} - \rho_{yz_2} - \rho_{xz_1} + \rho_{xz_2} - (1/2)\rho_{z_1z_2}$

$A_3 = 2\,\rho_{yx} - \rho_{yz_1} + \rho_{xz_1} - \rho_{xz_2} + (1/2)\rho_{z_1z_2} - (5/4)$ and $Cov(T_u, T_m) = 0$.

2.5. Minimum Mean Squared Error of the Proposed Estimator T

Since the mean squared error of the estimator T given in equation (10) is a function of unknown constant α, therefore, it has been minimized with respect to α and subsequently the optimum value of α and hence optimum mean squared error of the estimator T are given respectively as

$$\alpha_{opt.} = M(T_m) \Big/ \big(M(T_u) + M(T_m)\big) \tag{13}$$

$$M(T)_{opt.} = \big(M(T_u) \, . \, M(T_m)\big) \Big/ \big(M(T_u) + M(T_m)\big) \tag{14}$$

Further, substituting the value of the mean squared error of the estimators defined in equations (2) and (5) in equation (13) and (14) respectively, the simplified values of $\alpha_{opt.}$ and $M(T)_{opt.}$ are obtained as

$$\alpha_{opt.} = \mu f_2 \left[\mu A_3 - (f_1 A_2 + A_3)\right] \Big/ \left[\mu^2 f_2 A_3 - \mu\left(f_1 f_2 A_2 + f_2 A_3 - f_1 A_1\right) - f_1 A_1\right] \tag{15}$$

$$M(T)_{opt.} = \left[\mu C_1 - C_2\right] S_y^2 \Big/ n\left[\mu^2 C_3 - \mu C_4 - C_5\right] \tag{16}$$

where $C_1 = A_1 A_3$, $C_2 = f_1 A_1 A_2 + A_1 A_3$, $C_3 = f_2 A_3$, $C_4 = f_2 A_3 + f_1 f_2 A_2 - f_1 A_1$, $C_5 = f_1 A_1$ and μ is the fraction of the sample drawn afresh at the current (second) occasion.

Remark 2.5.1: $M(T)_{opt.}$ derived in equation (16) is a function of μ. To estimate the population mean on each occasion the better choice of μ are 1(case of no matching); however, to estimate the change in mean from one occasion to other, μ should be 0(case of complete matching). But intuition suggests that the optimum choices of μ are desired to devise the amicable strategy for both the problems simultaneously.

2.6. Optimum Replacement Strategies for the Estimator T

The key design parameter affecting the estimates of change is the overlap between successive samples. Maintaining high overlap between repeats of a survey is operationally convenient, since many sampled units have been located and have some experience in the survey. Hence to decide about the optimum value of μ (fractions of samples to be drawn afresh on current occasion) so that $\bar{Y}$ may be estimated with maximum precision and minimum cost, we minimize the mean squared error $M(T)_{opt.}$ in equation (16) with respect to μ.

The optimum value of μ so obtained is one of the two roots given by

$$\mu = \left(D_2 \pm \sqrt{D_2^2 - D_1 D_3}\right) \Big/ D_1 \tag{17}$$

where $D_1 = C_1 C_3$, $D_2 = C_2 C_3$, $D_3 = C_1 C_5 + C_2 C_4$.

The real value of μ exist, iff $D_2^2 - D_1 D_3 \geq 0$. For any situation, which satisfies these conditions, two real values of μ may be possible , hence to choose a value of μ, it should be taken care of that $0 \leq \mu \leq 1$, all other values of μ are inadmissible. If both the real values of μ are admissible, the lowest one will be the best choice as it reduces the total cost of the survey. Substituting the admissible value of μ say μ_0 from equation (17) in equation (16), we get the optimum value of the mean squared error of the estimator T with respect to α as well as μ which is given as

$$M(T)^{*}_{opt.} = \left[\mu_0 C_1 - C_2\right] S_y^2 \Big/ n\left[\mu_0^2 C_3 - \mu_0 C_4 - C_5\right] \tag{18}$$

3. Special Cases

3.1. Case I: When there is Non-Response only at the First Occasion (Previous Occasion)

When there is a presence of non-response, the proposed estimator T for population mean $\bar{Y}$ changes to

$$T_1 = \varphi T_u^{\theta} + (1 - \varphi) T_m \tag{19}$$

where $T_u^{\theta} = \bar{y}_u \exp\left(\dfrac{\bar{\bar{Z}}_2 - \bar{\bar{z}}_2(u)}{\bar{\bar{Z}}_2 + \bar{\bar{z}}_2(u)}\right)$ and T_m is defined in equation (5) and $\varphi(0 \leq \varphi \leq 1)$ is a real constant to be determined so as to minimize the mean squared error of the estimator T_1.

In this case, the optimum value of fraction of sample drawn afreshand the minimum mean squared error of the estimator T_1 at the admissible value of $\hat{\mu}$ are derived respectively as

$$\hat{\mu} = \left(D_5 \pm \sqrt{D_5^2 - D_4\ D_6}\right) \Big/ D_4 = \mu_1 (\text{say})$$

$$M(T_1)^*_{opt.} = \left[\mu_1\ C_6 - C_7\right] S_y^2 \Big/ n\left[\mu_1^2\ A_3 - \mu_1\ C_8 - C_9\right] \qquad (20)$$

where $D_4 = A_3C_6,\ D_5 = A_3C_7,\ D_6 = C_6C_9 + C_7C_8,\ C_6 = A_1A_3,\ C_7 = f_1A_1A_2 + A_1A_3$
$C_8 = A_3 + f_1A_2 - f_1A_1\ ,\ \ C_9 = f_1A_1$ and $f_1 = r_1/n.$

3.2. Case II: When there is Non-Response only at the Second (Current) Occasion

The estimator for population mean $\bar{Y}$ at the current occasion in the presence of non-response at current occasion is given by

$$T_2 = \psi\ T_u + (1 - \psi) T_m^{\theta} \qquad (21)$$

where $T_m^{\theta} = \bar{x}_n^* \left(\dfrac{\bar{y}_m^*}{\bar{x}_m^*}\right), \bar{y}_m^* = \bar{y}_m\ \exp\left(\dfrac{\bar{Z}_2 - \bar{z}_2(m)}{\bar{Z}_2 + \bar{z}_2(m)}\right), \bar{x}_m^* = \bar{x}_m\ \exp\left(\dfrac{\bar{Z}_1 - \bar{z}_1(m)}{\bar{Z}_1 + \bar{z}_1(m)}\right)$

$\bar{x}_n^* = \bar{x}_n\ \exp\left(\dfrac{\bar{Z}_1 - \bar{z}_1(n)}{\bar{Z}_1 + \bar{z}_1(n)}\right)$ and T_u is defined in equation (2) and $\psi(0 \le \psi \le 1)$ is a real

constant to be determined so as to minimize the mean squared error of the estimator T_2.

In this case, the optimum value of fraction of sample drawn afresh and the minimum mean squared error of the estimator T_2 at the admissible value of $\hat{\mu}$ are derived respectively as

$$\hat{\mu} = \left(D_8 \pm \sqrt{D_8^2 - D_7\ D_9}\right) \Big/ D_7 = \mu_2 (\text{say})$$

$$M(T_2)^*_{opt.} = \left[\mu_2\ C_{10} - C_{11}\right] S_y^2 \Big/ n\left[\mu_2^2\ C_{12} - \mu_2\ C_{13} - A_1\right] \qquad (22)$$

where $D_7 = C_{10}C_{12},\ D_8 = C_{11}C_{12},\ D_9 = A_1C_{10} + C_{11}C_{13},\ C_{10} = A_1A_3,\ C_{11} = A_1A_2 + A_1A_3$
$C_{12} = f_2A_3,\ \ C_{13} = f_2A_3 + f_2A_2 - A_1$ and $f_2 = r_2/u.$

4. Efficiency Comparison

The percent relative loss in the efficiency of the proposed estimators T has been recorded to infer about the affect of incompleteness in the data over the occasions with respect to the estimator T_{CR} (Priyanka and Mittal (2016)) under the same circumstances but for complete response over the occasions which is described as

$$T_{CR} = \xi\, T_u^{\theta} + (1-\xi) T_m^{\theta} \tag{23}$$

where T_u^{θ} and T_m^{θ} have been defined in equation (19) and (21) and $\xi(0 \le \xi \le 1)$ is a real constant to be determined so as to minimize the mean squared error of the estimator T_{CR}.

The optimum mean squared error for the estimator T_{CR} with respect to ξ as well as μ is obtained as

$$M(T_{CR})^{*}_{opt.} = \left[\mu^{*} G_1 - G_2\right] S_y^2 \Big/ n\left[\mu^{*2}\, B_3 - \mu^{*} G_3 - B_1\right] \tag{24}$$

where $\mu^{*} = \left(H_2 \pm \sqrt{H_2^2 - H_1\, H_3}\right) \Big/ H_1$,

$H_1 = B_3 G_1,\; H_2 = B_3 G_2,\; H_3 = B_1 G_1 + G_2 G_3,\; G_1 = B_1 B_3,$

$G_2 = B_1 B_2 + B_1 B_3,\;\; G_3 = B_3 + B_2 - B_1,\;\;\; B_1 = (5/4) - \rho_{yz_2},$

$B_2 = (5/2) - 2\,\rho_{yx} + \rho_{yz_1} - \rho_{yz_2} - \rho_{xz_1} + \rho_{xz_2} - (1/2)\rho_{z_1 z_2}$

and $B_3 = 2\,\rho_{yx} - \rho_{yz_1} + \rho_{xz_1} - \rho_{xz_2} + (1/2)\rho_{z_1 z_2} - (5/4)$.

The percent relative loss in precision of the estimators T, T_1 and T_2 with respect to the estimator T_{CR} under their respective optimality conditions are given by

$$\left.\begin{aligned} L_0 &= \frac{M(T)^*_{opt.} - M(T_{CR})^*_{opt.}}{M(T)^*_{opt.}} \times 100 \\ L_1 &= \frac{M(T_1)^*_{opt.} - M(T_{CR})^*_{opt.}}{M(T_1)^*_{opt.}} \times 100 \\ L_2 &= \frac{M(T_2)^*_{opt.} - M(T_{CR})^*_{opt.}}{M(T_2)^*_{opt.}} \times 100 \end{aligned}\right\} \qquad (25)$$

5. Numerical Illustrations and Simulation

5.1. Empirical Study

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

Empirical validation of theoretical results has been elaborated by means of a natural population. The population I consist of N=51 states of United States. Let y_i be net summer capacity during 2008 in the i^{th} state of U. S., x_i denote the net summer capacity during 2000 in the i^{th} state of U. S., z_{1_i} denote the residential consumption of electric power during 2000 and z_{2_i} denote the residential consumption of electric power during 2008.

The empirical analysis of the proposed estimators has been shown in Table 1 for various choices for fraction of non-response over the successive occasions.

Table 1: Empirical results when the proposed estimators T, T_1 and T_2 have been compared to the estimator T_{CR}.

t_1=0.30, t_2=0.30						
μ^*	μ_0	μ_1	μ_2	L_0	L_1	L_2
0.6773	0.6018	0.7347	0.4312	19.80	3.03	10.31
t_1=0.20, t_2=0.20						
μ^*	μ_0	μ_1	μ_2	L_0	L_1	L_2
0.6773	0.6050	0.6968	0.5062	11.86	1.03	6.41
t_1=0.25, t_2=0.15						
μ^*	μ_0	μ_1	μ_2	L_0	L_1	L_2
0.6773	0.6540	0.7158	0.5387	10.56	2.03	4.25

5.2. Generalization of empirical study

A more generalized study has also been done to show the impact of the proposed estimators under different fractions of non-response and choices of correlation coefficients of study and auxiliary variables. The results obtained are shown in Table 2. Here for the sake of convenience, we have considered $\rho_{yz_1} = \rho_{yz_2} = \rho_1$ and $\rho_{xz_1} = \rho_{xz_2} = \rho_2$.

Table 2: Generalized empirical results while the proposed estimators T, T_1 and T_2 have been compared to the estimator T_{CR} .

t_1=0.10 and t_2=0.10															
$\rho_{yx}, \rho_{z_1z_2}$		**0.5**							**0.6**						
ρ_2	ρ_1	μ^*	μ_0	μ_1	μ_2	L_0	L_1	L_2	μ^*	μ_0	μ_1	μ_2	L_0	L_1	L_2
0.4	**0.4**	0.81	0.63	0.50	0.59	11.49	5.62	11.99	0.91	0.83	0.81	0.81	9.85	3.03	9.89
	0.5	0.78	0.58	0.49	0.54	12.54	6.88	13.34	0.87	0.68	0.65	0.65	1.93	4.83	11.21
	0.6	0.75	0.55	0.47	0.50	13.71	8.20	14.87	0.83	0.61	0.57	0.57	12.35	6.60	12.95
0.5	**0.4**	0.84	0.63	0.50	0.59	11.90	6.07	12.40	0.93	0.83	0.81	0.81	9.95	3.14	9.99
	0.5	0.81	0.58	0.49	0.54	13.14	7.51	13.93	0.89	0.68	0.65	0.65	11.18	5.09	11.46
	0.6	0.78	0.55	0.47	0.50	14.51	9.05	15.66	0.85	0.61	0.57	0.57	12.79	7.07	13.38
0.6	**0.4**	0.83	0.63	0.50	0.59	12.21	6.40	12.71	0.94	0.83	0.81	0.81	10.02	3.21	10.07
	0.5	0.83	0.58	0.49	0.54	13.59	7.99	14.37	0.90	0.68	0.65	0.65	11.36	5.28	11.63
	0.6	0.80	0.55	0.47	0.50	15.12	9.70	16.26	0.87	0.61	0.57	0.57	13.11	7.41	13.70
0.7	**0.4**	0.87	0.63	0.50	0.59	12.45	6.65	12.95	0.94	0.83	0.81	0.81	10.07	3.27	10.12
	0.5	0.85	0.58	0.49	0.54	13.94	8.36	14.72	0.91	0.68	0.65	0.65	11.49	5.43	11.77
	0.6	0.82	0.55	0.47	0.50	15.60	10.21	16.74	0.88	0.61	0.57	0.57	13.35	7.67	13.95
t_1=0.25 and t_2=0.20															
$\rho_{yx}, \rho_{z_1z_2}$		**0.5**							**0.6**						
ρ_2	ρ_1	μ^*	μ_0	μ_1	μ_2	L_0	L_1	L_2	μ^*	μ_0	μ_1	μ_2	L_0	L_1	L_2
0.4	**0.4**	0.81	0.81	0.58	0.75	18.83	4.13	18.92	0.91	**	0.60	**	-	2.41	-
	0.5	0.78	0.74	0.57	0.66	18.72	4.87	19.66	0.87	0.90	0.59	0.87	18.85	3.73	18.93
	0.6	0.75	0.69	0.56	0.59	19.07	5.59	20.75	0.83	0.78	0.58	0.71	1902	4.94	19.59
0.5	**0.4**	0.84	0.81	0.58	0.75	18.91	4.58	19.30	0.93	**	0.60	**	-	2.52	-
	0.5	0.81	0.74	0.57	0.66	19.27	5.52	20.21	0.89	0.90	0.59	0.87	19.08	4.00	19.16
	0.6	0.78	0.69	0.56	0.59	19.82	6.46	21.48	0.85	0.78	0.58	0.71	19.42	5.41	19.99
0.6	**0.4**	0.83	0.81	0.58	0.75	19.20	4.92	19.59	0.94	**	0.60	**	-	2.60	-
	0.5	0.83	0.74	0.57	0.66	19.69	6.01	20.62	0.90	0.90	0.59	0.87	19.24	4.19	19.32
	0.6	0.80	0.69	0.56	0.59	20.39	7.13	22.04	0.87	0.78	0.58	0.71	19.72	5.76	20.29
0.7	**0.4**	0.87	0.81	0.58	0.75	19.42	5.18	19.81	0.94	**	0.60	**	-	2.66	-
	0.5	0.85	0.74	0.57	0.66	20.01	6.39	20.94	0.91	0.90	0.59	0.87	19.36	4.34	19.44
	0.6	0.82	0.69	0.56	0.59	20.84	7.66	22.49	0.88	0.78	0.58	0.71	19.94	6.02	20.51

Note: The values for μ^*, μ_0, μ_1 and μ_2 have been rounded off up to two places of decimal for presentation.

5.3. Monte Carlo Simulation

The population II comprise of N = 51 states of United States. Let y_i be the net electric power generation during 2008 in the i^{th} state of U. S., x_i be the net electric power generationduring 2000 in the i^{th} state of U. S., z_{1_i} denote the net summer capacity during 2000 in the i^{th} state of U. S. and z_{2_i} denote the net summer capacity during 2008 in the i^{th} state of U. S.

Monte Carlo simulation has been performed on population II, for better analysis considering different choices of t_1 and t_2.

5.3.1. Simulation Algorithm

(i) Choose 5000 samples of size n=25 using simple random sampling without replacement on first occasion for both the study and auxiliary variable.

(ii) For $f_1 = 0.88$, choose $r_1 = 22$ responding units out of n=25 samples units.

(iii) Calculate sample mean $\bar{x}_{r_1|k}$ and $\bar{z}_{1|k}(r_1)$ for k =1, 2, - - -, 5000.

(iv) Retain m=15 units out of each $r_1 = 22$ sample units of the study and auxiliary variables at the first occasion.

(v) Calculate sample mean $\bar{x}_{m|k}$ and $\bar{z}_{1|k}(m)$ for k= 1, 2, - - -, 5000.

(vi) Select u=10 units using simple random sampling without replacement from N-n=26 units of the population for study and auxiliary variables at second (current) occasion.

(vii) For $f_2 = 0.90$, choose $r_2 = 9$ responding units out of u=10 samples units.

(viii) Calculate sample mean $\bar{y}_{r_2|k}$, $\bar{y}_{m|k}$, $\bar{z}_{2|k}(m)$ and $\bar{z}_{2|k}(r_2)$ for k = 1, 2, - - -, 5000.

(ix) Iterate the parameter α from 0.1 to 0.9 with a step of 0.2.

(x) Iterate ξ from 0.1 to 0.9 with a step of 0.1 within (ix).

(xi) Calculate the percent relative loss in efficiencies of the proposed estimator T, T_1 and T_2 with respect to estimator T_{CR} as

$$L(T) = \frac{\sum_{k=1}^{5000}\left[T_{|k} - T_{CR|k}\right]^2}{\sum_{k=1}^{5000}\left[T_{|k}\right]^2} \times 100 , \qquad L(T_1) = \frac{\sum_{k=1}^{5000}\left[T_{1|k} - T_{CR|k}\right]^2}{\sum_{k=1}^{5000}\left[T_{1|k}\right]^2} \times 100$$

and $$L(T_2) = \frac{\sum_{k=1}^{5000}\left[T_{2|k} - T_{CR|k}\right]^2}{\sum_{k=1}^{5000}\left[T_{2|k}\right]^2} \times 100 , \quad k=1, 2, ..., 5000.$$

Table 3: Simulation result when the proposed estimator T is compared with the estimator T_{CR} when non-response occurs on both the occasion

α \ ξ	SET	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.1	**I**	46.05	60.81	66.92	59.37	39.65	16.81	-20.54
	II	31.76	49.12	55.63	47.74	21.70	-8.40	-59.24
	III	25.14	43.67	49.85	41.75	22.09	-25.97	-75.52
0.2	**I**	41.00	57.08	62.83	55.13	34.92	6.12	-35.84
	II	22.79	43.01	50.56	4.20	14.35	-24.41	-79.91
	III	14.44	36.58	42.69	33.62	7.15	-43.63	-82.36
0.3	**I**	38.96	54.47	60.92	52.37	31.50	1.25	-40.43
	II	17.24	40.30	47.62	36.94	9.23	-32.24	-91.28
	III	7.94	31.92	38.42	27.87	0.211	-54.25	-100.57
0.4	**I**	38.93	55.18	60.81	52.50	32.26	1.46	-41.78
	II	18.38	4.72	47.74	36.94	10.57	-30.64	-89.35
	III	7.79	31.61	38.07	27.43	-1.01	-54.23	-114.80
0.5	**I**	43.25	58.25	62.87	55.46	36.25	6.35	-33.20
	II	24.81	44.96	51.11	40.89	16.44	-22.26	-76.86
	III	11.57	35.99	42.22	31.33	4.43	-45.92	-106.78
0.6	**I**	49.73	62.25	66.09	59.91	42.63	15.98	-19.59
	II	33.26	51.74	56.74	47.77	25.91	-7.99	-57.18
	III	21.53	43.08	48.56	38.85	14.57	-29.38	-83.80
0.7	**I**	56.23	67.33	70.48	65.19	49.70	26.75	-4.89
	II	42.93	58.71	63.18	55.27	36.78	7.32	-34.08
	III	32.87	51.25	55.59	47.57	26.18	-10.55	-56.75
0.8	**I**	62.35	72.22	75.10	70.49	57.24	37.69	10.94
	II	52.18	65.42	69.19	62.50	46.56	22.27	-12.58
	III	43.92	59.27	62.51	56.31	38.34	8.08	-31.46
0.9	**I**	68.17	76.43	78.73	75.08	64.14	46.73	24.08
	II	60.26	71.02	74.18	68.56	55.75	35.10	6.61
	III	53.71	65.97	68.74	63.54	48.86	23.89	9.51

I: $n=25, \mu = 0.40, t_1=0.28, t_2=0.30$, II: $n=25, \mu = 0.40, t_1=0.16, t_2=0.20$

III: $n=25, \mu = 0.40, t_1=0.12, t_2=0.10$

Table 4: Simulation result when the proposed estimator T_1 is compared with the estimator T_{CR} when non-response occurs only on first occasion

φ \ ξ	SET	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.1	**I**	50.48	62.74	66.33	60.63	42.76	22.14	-12.67
	II	39.01	52.59	58.29	50.28	32.93	5.16	-42.17
	III	15.55	38.43	46.11	35.20	13.02	-27.07	-78.92
0.2	**I**	42.35	57.49	62.20	55.69	38.16	12.18	-25.80
	II	27.35	47.05	51.74	45.44	23.25	-9.66	-63.11
	III	3.85	30.21	38.69	26.61	1.42	-46.30	-105.64
0.3	**I**	36.84	53.21	58.90	51.89	32.90	3.79	-39.54
	II	18.80	42.12	47.64	40.46	16.23	-21.37	-78.27
	III	-4.81	23.70	33.18	19.94	-8.18	-57.58	-125.97
0.4	**I**	33.01	51.10	56.58	49.03	29.03	-1.88	-47.21
	II	15.93	39.76	44.81	37.29	12.45	-28.47	-86.36
	III	-7.52	21.62	31.00	18.34	-11.53	-61.03	-132.10
0.5	**I**	31.89	50.84	55.96	48.63	28.12	-2.83	-48.72
	II	16.13	39.89	45.46	37.69	13.17	-27.62	-84.67
	III	-3.76	24.78	33.62	21.69	-6.90	-55.47	-123.98
0.6	**I**	33.77	52.42	57.31	50.32	30.42	0.322	-43.88
	II	21.25	43.15	49.17	41.38	18.03	-20.40	-72.92
	III	5.63	32.18	39.17	29.47	2.71	-41.13	-102.72
0.7	**I**	38.97	56.08	60.54	54.01	35.32	7.75	-33.70
	II	29.43	49.05	54.49	47.50	26.19	-8.31	-55.78
	III	18.00	41.42	47.01	38.46	15.71	-23.19	-75.88
0.8	**I**	45.31	60.59	64.55	58.66	42.39	17.01	-19.92
	II	38.56	55.46	60.40	54.26	35.62	6.25	-35.71
	III	30.99	50.40	55.73	48.25	28.19	-3.87	-49.95
0.9	**I**	52.39	65.24	68.83	63.44	49.50	26.71	-5.54
	II	47.67	61.64	66.10	60.73	45.13	19.88	-15.94
	III	42.67	58.49	63.03	56.84	39.97	13.00	-24.82

I: $n=25, \mu = 0.40, t_1=0.28$

II: $n=25, \mu = 0.40, t_1=0.16$

III: $n=25, \mu = 0.40, t_1=0.12$

Table 5: Simulation result when the proposed estimator T_2 is compared with the estimator T_{CR} when non-response occurs only on second occasion

ψ \ ξ	SET	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0.1	I	38.23	54.22	59.08	51.01	31.01	-2.48	-45.26
	II	37.14	55.19	57.99	50.39	28.97	-4.50	-48.73
	III	35.67	52.66	59.41	49.95	27.97	-9.61	-49.80
0.2	I	39.89	54.99	59.63	51.74	30.71	-2.04	-45.18
	II	36.79	53.49	57.84	50.60	27.28	-3.92	-52.52
	III	33.84	51.31	58.05	49.13	24.82	-11.68	-54.67
0.3	I	42.86	57.14	61.30	53.45	34.87	5.06	-37.21
	II	38.89	54.70	59.65	53.00	30.58	-1.79	-46.86
	III	35.46	52.08	58.75	49.42	27.50	-8.13	-50.78
0.4	I	47.48	61.68	65.36	58.22	41.25	14.06	-24.02
	II	43.37	58.28	62.44	55.83	35.71	6.75	-33.67
	III	38.81	55.31	60.30	52.09	31.41	-1.52	-42.55
0.5	I	54.24	66.44	69.72	63.64	48.85	25.30	-7.38
	II	49.43	62.50	66.35	60.47	42.67	17.82	-19.20
	III	44.53	58.89	63.44	56.26	36.95	7.02	-30.51
0.6	I	60.24	71.03	74.16	68.86	55.69	35.86	6.91
	II	56.03	66.98	70.84	65.15	50.14	27.81	-3.87
	III	50.54	63.22	67.35	61.33	43.65	17.16	-16.55
0.7	I	66.19	75.94	77.92	73.60	62.46	45.13	21.03
	II	61.97	71.64	74.79	70.09	56.88	37.70	9.71
	III	56.78	67.76	71.62	66.20	51.09	27.81	-2.07
0.8	I	71.42	78.94	81.29	77.52	68.36	53.28	32.87
	II	67.38	75.91	78.11	74.32	63.25	46.35	22.38
	III	62.51	72.01	75.22	70.49	57.45	37.23	11.42
0.9	I	75.87	82.18	83.97	80.90	73.02	60.05	42.92
	II	71.87	79.38	81.26	78.08	68.50	53.63	33.36
	III	67.34	75.86	78.70	74.25	63.38	45.71	23.12

I: n=25, $\mu = 0.40$, t_2=0.40

II: n=25, $\mu = 0.40$, t_2=0.30

III: n=25, $\mu = 0.40$, t_2=0.20

9. Rendition of Results

The performance of an estimator in successive sampling in the presence of non-response is generally judged on the basis of percent relative loss in efficiency (lesser is the loss better is the estimator) and in terms of optimum value of fraction of fresh sample to be drawn afresh on current(second) occasion which directly related to the cost of survey. Following interpretation can be drawn from Tables 1- 5,

1) Results based on empirical study

a) From Table 1 we can see that the values of μ_0, μ_1, and μ_2 exist for various choices of fraction of non-response over two successive occasions which completely signifies the utility of a dynamic natured auxiliary character.

b) Also from Table 1, we identify that percent relative loss L_0, L_1 and L_2 exist each combination of t_1 and t_2 and when non response occurs at both occasion the percent relative loss is more as compared to non-response on first or second occasion only.

c) We can also conclude from the Table 1 that, the percent relative loss in efficiency is not very much significant when the proposed estimators T, T_1 and T_2 are compared to estimator T_{CR}.

2) Results extracted from the generalized study for various combinations of correlation coefficients

a) In Table 2, we see that the values of $\mu_0, \mu_1, \mu_2, L_0, L_1$ and L_2 exist for almost every combination of coefficient of correlation of study and auxiliary characteristics considering various possibilities of non-response that can creep in a sample survey over successive occasions.

b) We also see that the proposed estimators work efficiently when auxiliary character which is dynamic in nature conceives a moderate or low correlation with the study character over the successive occasions.

c) We also identify that as the amount of correlation between study and auxiliary character is increased, the proposed estimators intend to provide a lesser fraction of sample to be drawn afresh at current occasion.

3) Results based on simulation study in Table 3, Table 4 and Table 5

a) We can see that for fixed choice of ξ, percent relative loss first decreases and then increases with increasing value of α and φ respectively while for all fixed choices of ξ, percent relative loss $L(T_2)$ increases as ψ increases.

b) The percent relative loss $L(T), L(T_1)$ and $L(T_2)$, for fixed choice of α, φ and ψ respectively, first increase as ξ increases and then decreases with increasing value of ξ.

(c) As we decrease the fraction of non-response in the sample on first and second occasion, the percent relative loss $L(T), L(T_1)$ and $L(T_2)$ decrease for all combinations of α, φ and ψ with ξ respectively.

10. Conclusion

The proposed estimators have been analysed considering a detailed study in the presence of non-response utilizing additional auxiliary information which is dynamic in nature over the successive occasions. Loss in efficiency is very much plebeian when non-response is encountered in the sample survey. Although percent relative loss is encountered for various fractions of non-response on two occasions using the proposed method of imputation while utilizing dynamic auxiliary character but a negative loss is also available for various choices of parameters. This signifies that the proposed estimators emerge to be better than the estimator due to Priyanka and Mittal (2016) for such combinations of parameters and hence the proposed method of imputation is fruitful to cope with the non-response. The proposed imputation techniques prove to be worthy from the point of cost as well when correlation between study and auxiliary character is considered moderate or even low. Hence, it is observed that the proposed imputation methods deal the sour effect of non-response excellently, therefore, the proposed estimators may be recommended for encouraging their practical use by survey practitioners.

UNIT - IV

SEARCH OF GOOD ROTATION PATTERNS ADDRESSING SENSITIVE ISSUES

CHAPTER – 12*

Scrambled Response Techniques in Two Wave Rotation Sampling for Estimating Population Mean of Sensitive Characteristics and its Applications

* Following is the publication based on the work of this chapter:--

1. Priyanka, K. and Mittal, R. (2016): Scrambled Response Techniques in Two Wave Rotation Sampling for Estimating Population Mean of Sensitive Characteristics and its Applications. Metrika. (Communicated).

Scrambled Response Techniques in Two Wave Rotation Sampling for Estimating Population Mean of Sensitive Characteristics and its Applications

1. Introduction

Analysis of sensitive issues like negligence of governmental rules, number of abortion before marriage, bribing for some entrance exam, sexual indulgence during teenage, status of extramarital relationship, employing child labourers, child-sexual abuse, voluntary prostitution, commencement of crime, honour killing, drug intake etc., usually lead to over or under reporting of the true facts due to social or moral inclinations and stigma. Thus a significant deviation occurs in the results owing to socially desirable answers which do not comply to real scenario subsisting in the society.

There are two approaches to estimate population proportion or population mean of a quantitative sensitive character. First approach is to reduce the stigma involved in answering such sensitive questions by providing certain privacy through a randomized response device following certain randomized response rule (Randomized Response Model). Warner (1965) was the first to provide such a randomizing model and later on extensive literature have been added by Horvitz et al. (1967), Greenberg et al. (1971), Gupta et al. (2002), Christofides (2003, 2005), Gupta and Shabbir (2004), Kim and Elam (2007), Wu et al. (2008), Yan et al. (2009), Arnab (2011), Dianna and Perri (2011), Arnab et al. (2012), Singh and Sedory (2012) and Sihm and Gupta (2015) etc.
All these authors have focused on estimation of population mean or proportion of sensitive characters using some randomised response models.

This approach becomes practically next to impossible when it comes to observe a very large sample since lifestyle has drastically changed and people are living a very fast life with certain time constraints so complete refusal to response is also encountered due to time consuming procedure. In such situations, second approach known as Scrambled Response Technique (SRT) which was introduced by Warner (1971) but was left for exploration and very first attempts were made by Pollock and Bek (1976) and Eichhorn and Hayre (1983), works as saviour. This technique reduces the impossibility of conducting a survey having large sample size with a sensitive issue to be addressed. In this technique to estimate the population mean of sensitive character the respondent is asked to answer freely about the stigmatizing character by adding or multiplying a corrective scrambling factor to his/her response hiding real response from the interviewer. In this line a rich literature is available from Saha (2007), Gupta et al. (2006), Gupta et al. (2010, 2012), Koyuncu et al. (2014) and Hussain and Al-Zhrani (201) etc.

Moreover, these above said issues have been addressed through a single time survey in the literature available on sensitive character analysis; instead these issues are required to be monitored continuously over time, since doing so will reflect the change of social scenario related to the sensitive issues as well as changed level of sensitivity of issue with respect to time. For example, any government of a county may be interested to record the mean number of rape cases in the country at starting of their ruling period. After recording them one time the government may interest to decrease these for ensuring the better society. For this government can make stricter laws against the rapist, more awareness of such laws can be spread amongst the females, it may also increase the level of security for females at work place and so on. After such precautious measures government may wish to see the changed level of the society by recoding the mean number of rape cases at the end of their ruling tenure.

In order to monitor such a variable more than once, statistical tool generally recommended in literature is successive or rotation sampling. Jessen (1942) started the theory of rotation sampling by utilising all the information collected from previous occasion. His pioneer work in this line has been followed by Patterson (1950), Sen (1973), Feng and Zou (1997), Singh and Priyanka (2008a), Bandyopadhyay and Singh (2014), Priyanka and Mittal (2014, 2015a, 2015b), Priyanka et al. (2015) and many others.

None of the above works in successive sampling analyses sensitive issues which change over time. Very few attempts namely Arnab and Singh (2013) and Yu et al. (2014) are found which dealt with sensitive issues on successive waves while using randomized response technique. As per our knowledge is concerned no attempt has been made to utilise a non-sensitive auxiliary variable to estimate a sensitive study variable on successive waves using scrambled response technique. Hence, motivated with this scope of study, the present article endeavours to propose two kinds of estimators to estimate population mean of a sensitive character, first is a modified Jessen's estimator under scrambled response using all information from previous wave without any auxiliary information and second are four exponential ratio type estimators accompanying a non-sensitive stable auxiliary character correlated to the sensitive-study character over two successive waves. All the above said estimators are studied under additive scrambled response model (ASRM) as well as multiplicative scrambled response model (MSRM) and properties of proposed estimators including the optimum rotation rates have been derived upto first order of approximations. Discussion has been made regarding the distribution of scrambling variable. A numerical illustration has been made to compare both the scrambled response models. Also an empirical study has been worked out for the best suited scrambled response model on two successive waves by the means of a case study of drug usage by undergraduate students in a college for the real life application of the proposed estimators. Simulation studies are rationalized to show the feasibility of proposed estimators. Mutual comparisons of the proposed estimators have also been illustrated. The model for optimum total cost of the survey has also been designed and discussed.

2. Survey Design and Analysis

2.1. Sample Structure and Notations

Let $\Theta = (\Theta_1, \Theta_2, \dots, \Theta_N)$ be the finite population of N units, which has been sampled over two successive waves. It is assumed that size of the population remains unchanged but values of units change over two successive waves. The sensitive character under study be denoted by x (y) on the first (second) waves respectively. It is assumed that information on non-sensitive auxiliary variable z, stable in nature over the successive waves with completely known population mean $\bar{Z}$, is readily available on both the successive waves and positively correlated to x and y respectively. Simple random sample (without replacement) of n units is taken at the first wave. A random subsample of $m = n\lambda$ units is retained for use at the second wave. Now at the current wave, a simple random sample (without replacement) of u= (n-m) = nμ units is drawn afresh from the remaining (N-n) units of the population so that the sample size on the second wave remains the same. Let μ and $\lambda\left(\mu + \lambda = 1\right)$ are the fractions of fresh and matched samples respectively at the second (current) successive wave. The following notations are considered here after:

$\bar{X}, \bar{Y}, \bar{Z}$: Population means of the variables x, y and z respectively.

$\bar{h}_u, \bar{g}_m, \bar{h}_m, \bar{g}_n$ and $\bar{h}_u^*, \bar{g}_m^*, \bar{h}_m^*, \bar{g}_n^*$: Sample mean of sensitive variate based on sample sizes shown in suffice under additive scrambled response model and multiplicative scrambled response model respectively.

$\bar{z}_u, \bar{z}_m, \bar{z}_n$: Sample mean of the auxiliary variate based on sample sizes shown in suffice.

$\rho_{yx}, \rho_{xz}, \rho_{yz}$: Correlation coefficient between the variables shown in suffices.

C_x, C_y, C_z : Coefficient of variance of variables shown in suffices.

$S_x^2, S_y^2, S_z^2, S_s^2$: Population mean squared of variables x, y, z and s respectively.

2.2. Additive Scrambled Response Model (ASRM)

Pollock and Bek (1976) were the first to discuss scrambling through additive model. In this model the respondent is asked to add his/her sensitive response X (Y) into a random (scrambling) variable S (independent of X(Y)) from a completely known distribution. Hence the observed response is given by

$G = X + S$ on the first wave ,

and $H = Y + S$ on the second wave,

The scrambling variable S may follow any distribution and which has been discussed in detail in further sections. If $E(S)=\bar{S}$, $V(S)=S_s^2$ then $\bar{G}=\bar{X}+\bar{S}$, $\bar{H}=\bar{Y}+\bar{S}$, $S_g^2=S_x^2+S_s^2$ and $S_h^2=S_y^2+S_s^2$, Also under ASRM $\rho_{hg}=\rho_{yx}$, $\rho_{hz}=\rho_{yz}$ and $\rho_{gz}=\rho_{xz}$.

2. 3. Multiplicative Scrambling Response Model (MSRM)

Multiplicative scrambling was also first studied by Pollock and Bek (1976) but a deep discussion was made by Eichhorn and Hayre (1983). In this model respondent is asked to multiply his/her sensitive response X(Y) by a scrambling variable S^* (independent of X(Y)) from a completely known distribution. So the observed response is given by

$G^*=XS^*$ on the first wave,

$H^*=YS^*$ on the second wave,

where $E(S^*)=\bar{S}^*$, $V(S^*)=S_s^{*2}$, $\bar{G}^*=\bar{X}\,\bar{S}^*$, $\bar{H}^*=\bar{Y}\,\bar{S}^*$, $S_g^{*2}=S_x^2S_s^{*2}+S_s^{*2}\bar{X}^2+S_x^2\bar{S}^{*2}$

and $S_h^{*2}=S_y^2S_s^{*2}+S_s^{*2}\bar{Y}^2+S_y^2\bar{S}^{*2}$ such that $\wp^2=\left(S_s^*/\bar{S}^*\right)^2$ should be as small as possible. Also under MSRM

$$\rho_{hg}^*=\frac{\left(\rho_{yx}C_yC_x\left(S_s^{*2}+\bar{S}^{*2}\right)+S_s^{*2}\right)}{\sqrt{C_y^2S_s^{*2}+C_y^2\bar{S}^{*2}+S_s^{*2}}\sqrt{C_x^2S_s^{*2}+C_x^2\bar{S}^{*2}+S_s^{*2}}},\ \rho_{gz}^*=\frac{\left(\rho_{xz}C_x\bar{S}^*\right)}{\sqrt{C_x^2S_s^{*2}+C_x^2\bar{S}^{*2}+S_s^{*2}}}\ \text{and}\ \rho_{hz}^*=\frac{\left(\rho_{yz}C_y\bar{S}^*\right)}{\sqrt{C_y^2S_s^{*2}+C_y^2\bar{S}^{*2}+S_s^{*2}}}$$

2.4. Design of the Proposed Estimators

Two kinds of estimators have been proposed to estimate population mean of sensitive characteristic on current wave under ASRM and MSRM. First is the modified Jessen's estimator under scrambled response which utilizes information from previous wave but doesn't accompany any auxiliary information on any non-sensitive character.

Table 1: The modified Jessen's estimators for scrambled response on two successive waves under ASRM and MSRM are proposed as

Estimator based on sample size u, drawn afresh at current wave		
Estimator	**Structure Under ASRM**	**Structure Under MSRM**
$\hat{\hat{¥}}_u(\tau)$	$\bar{h}_u$	$\bar{h}_u^*$
Estimator based on sample size m, retained from previous wave		
Estimator	**Structure Under ASRM**	**Structure Under MSRM**
$\hat{\hat{¥}}_m(\tau)$	$\bar{h}_m + k\left(\bar{g}_n - \bar{g}_m\right)$	$\bar{h}_m^* + k\left(\bar{g}_n^* - \bar{g}_m^*\right)$

The second kind of estimators are various exponential ratio type estimators which utilize information from previous wave as well as information on a non-sensitive auxiliary character which is stable over both the waves.

Table 2: The exponential ratio type estimators on two successive waves

Estimator based on sample size u, drawn afresh at current wave		
Estimator	**Structure Under ASRM**	**Structure Under MSRM**
$\check{\mathcal{T}}_{1u}(\tau)$	$\bar{h}_u\left(\dfrac{\bar{Z}}{\bar{z}_u}\right)$	$\bar{h}_u^*\left(\dfrac{\bar{Z}}{\bar{z}_u}\right)$
$\check{\mathcal{T}}_{2u}(\tau)$	$\bar{h}_u \exp\left(\dfrac{\bar{Z} - \bar{z}_u}{\bar{Z} + \bar{z}_u}\right)$	$\bar{h}_u^* \exp\left(\dfrac{\bar{Z} - \bar{z}_u}{\bar{Z} + \bar{z}_u}\right)$
Estimator based on sample size m, retained from previous wave		
Estimator	**Structure Under ASRM**	**Structure Under MSRM**
$\check{\mathcal{T}}_{1m}(\tau)$	$\bar{h}_m\left(\dfrac{\bar{g}_n}{\bar{g}_m}\right) \exp\left(\dfrac{\bar{Z} - \bar{z}_m}{\bar{Z} + \bar{z}_m}\right)$	$\bar{h}_m^*\left(\dfrac{\bar{g}_n^*}{\bar{g}_m^*}\right) \exp\left(\dfrac{\bar{Z} - \bar{z}_m}{\bar{Z} + \bar{z}_m}\right)$
$\check{\mathcal{T}}_{2m}(\tau)$	$\bar{h}_m^{\oplus}\left(\dfrac{\bar{g}_n^{\oplus}}{\bar{g}_m^{\oplus}}\right)$ where $\bar{h}_m^{\oplus} = \bar{h}_m \exp\left(\dfrac{\bar{Z} - \bar{z}_m}{\bar{Z} + \bar{z}_m}\right)$ $\bar{g}_m^{\oplus} = \bar{g}_m \exp\left(\dfrac{\bar{Z} - \bar{z}_m}{\bar{Z} + \bar{z}_m}\right)$ $\bar{g}_n^{\oplus} = \bar{g}_n \exp\left(\dfrac{\bar{Z} - \bar{z}_n}{\bar{Z} + \bar{z}_n}\right)$	$\bar{h}_m^{\dagger}\left(\dfrac{\bar{g}_n^{\dagger}}{\bar{g}_m^{\dagger}}\right)$ where $\bar{h}_m^{\dagger} = \bar{h}_m^* \exp\left(\dfrac{\bar{Z} - \bar{z}_m}{\bar{Z} + \bar{z}_m}\right)$ $\bar{g}_m^{\dagger} = \bar{g}_m^* \exp\left(\dfrac{\bar{Z} - \bar{z}_m}{\bar{Z} + \bar{z}_m}\right)$ $\bar{g}_n^{\dagger} = \bar{g}_n^* \exp\left(\dfrac{\bar{Z} - \bar{z}_n}{\bar{Z} + \bar{z}_n}\right)$

where $\tau = \begin{cases} \text{M} & \text{for Multiplicative Scrambled Response Model} \\ \text{A} & \text{for Additive Scrambled Response Model} \end{cases}$

Hence, considering the convex combination of the estimators based on sample size u and sample size m, we have the final estimators of the population mean at the current wave from Table 1 and Table 2 as:

Table 3: Final estimators of population mean of sensitive character at the current wave

Estimator	Structure Under ASRM	Structure Under MSRM
$\hat{\hat{\yen}}(\tau)$	$\xi(\text{A})\ \hat{\hat{\yen}}_{u}(\text{A}) + \left[1\text{-}\ \xi(\text{A})\right]\hat{\hat{\yen}}_{m}(\text{A})$	$\xi(\text{M})\ \hat{\hat{\yen}}_{u}(\text{M}) + \left[1\text{-}\ \xi(\text{M})\right]\hat{\hat{\yen}}_{m}(\text{M})$
$\check{\mathcal{T}}_{ij}(\tau)(\text{i, j=1, 2})$	$\varpi_{ij}(\text{A})\check{\mathcal{T}}_{iu}(\text{A}) + \left[1\text{-}\varpi_{ij}(\text{A})\right]\check{\mathcal{T}}_{jm}(\text{A})$	$\varpi_{ij}(\text{M})\check{\mathcal{T}}_{iu}(\text{M}) + \left[1\text{-}\varpi_{ij}(\text{M})\right]\check{\mathcal{T}}_{jm}(\text{M})$

where $\xi(\tau);\left[0 \le \xi(\tau) \le 1\right]$, and $\varpi_{ij}(\tau);\left[0 \le \varpi_{ij}(\tau) \le 1\right]$ are suitably chosen weights so as to minimize the variance and mean squared errors of the estimators $\hat{\hat{\yen}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau);(\text{i, j=1, 2})$ respectively.

2.5. Analysis of the proposed estimators

2.5.1. Bias and Mean Squared Errors of the Proposed Estimators

The properties of the proposed estimators are derived under the following large sample approximations

$$\bar{h}_u = \bar{H}(1 + e_0),\ \bar{h}_m = \bar{H}(1 + e_1),\ \bar{g}_m = \bar{G}(1 + e_2),\ \bar{g}_n = \bar{G}(1 + e_3),\ \bar{z}_u = \bar{Z}(1 + e_4),\ \bar{z}_m = \bar{Z}(1 + e_5)$$

and $\bar{z}_n = \bar{Z}(1 + e_6)$ such that $|e_i| < 1\ \forall\ i = 0,...,6$.

$$\bar{h}^*_u = \bar{H}^*(1 + \varepsilon_0),\ \bar{h}^*_m = \bar{H}^*(1 + \varepsilon_1),\ \bar{g}^*_m = \bar{G}^*(1 + \varepsilon_2),\ \bar{g}^*_n = \bar{G}^*(1 + \varepsilon_3),$$

such that $|\varepsilon_i| < 1\ \forall\ i = 0,...,3$.

Here $E(e_i)=0$ and $E(\varepsilon_i)=0$; $\forall$ i=1,...,6.

$E(e_0^2)=\alpha C_h^2$, $E(e_1^2)=\beta C_h^2$, $E(e_2^2)=\beta C_g^2$, $E(e_3^2)=\gamma C_g^2$, $E(e_4^2)=\alpha C_z^2$, $E(e_5^2)=\beta C_z^2$, $E(e_6^2)=\gamma C_z^2$,

$E(e_0e_4)=\alpha\rho_{hz}C_hC_z$, $E(e_1e_2)=\beta\rho_{hg}C_hC_g$, $E(e_1e_3)=\gamma\rho_{hg}C_hC_g$, $E(e_1e_5)=\beta\rho_{hz}C_hC_z$,

$E(e_1e_6)=\gamma\rho_{hz}C_hC_z$, $E(e_2e_3)=\gamma C_g^2$, $E(e_2e_5)=\beta\rho_{gz}C_gC_z$, $E(e_2e_6)=\gamma\rho_{gz}C_gC_z$, $E(e_3e_5)=\gamma\rho_{gz}C_gC_z$,

$E(e_3e_6)=\gamma\rho_{gz}C_gC_z$, $E(e_5e_6)=\gamma C_z^2$,

$E(\varepsilon_0^2)=\alpha C_{h^*}^2$, $E(\varepsilon_1^2)=\beta C_{h^*}^2$, $E(\varepsilon_2^2)=\beta C_{g^*}^2$, $E(\varepsilon_3^2)=\gamma C_{g^*}^2$, $E(\varepsilon_0e_4)=\alpha\rho_{hz}^*C_hC_z$, $E(\varepsilon_1\varepsilon_2)=\beta\rho_{hg}^*C_{h^*}C_g$,

$E(\varepsilon_1\varepsilon_3)=\gamma\rho_{hg}^*C_{h^*}C_{g^*}$, $E(\varepsilon_1e_5)=\beta\rho_{hz}C_{h^*}C_z$, $E(\varepsilon_1e_6)=\gamma\rho_{hz}^*C_{h^*}C_z$, $E(\varepsilon_2\varepsilon_3)=\gamma C_{g^*}^2$, $E(\varepsilon_2e_5)=\beta\rho_{gz}^*C_{g^*}C_z$,

$E(\varepsilon_2e_6)=\gamma\rho_{gz}^*C_{g^*}C_z$, $E(\varepsilon_3e_5)=\gamma\rho_{gz}^*C_{g^*}C_z$, $E(\varepsilon_3e_6)=\gamma\rho_{gz}^*C_{g^*}C_z$,

where $\alpha=\left(\frac{1}{u}-\frac{1}{N}\right)$, $\beta=\left(\frac{1}{m}-\frac{1}{N}\right)$, $\gamma=\left(\frac{1}{n}-\frac{1}{N}\right)$, $C_g^2=C_x^2+S_s^2/\bar{X}^2$, $C_h^2=C_y^2+S_s^2/\bar{Y}^2$,

$C_{g^*}^2=C_x^2+C_x^2C_{s^*}^2+C_{s^*}^2$ and $C_{h^*}^2=C_y^2+C_y^2C_{s^*}^2+C_{s^*}^2$.

Table 4: Bias of the proposed estimators under ASRM

Estimator	Expression of Bias
$\hat{\bar{¥}}_u(A)$	0
$\hat{\bar{¥}}_m(A)$	0
$\check{\mathcal{T}}_{1u}(A)$	$\frac{1}{u}\bar{H}\left(\frac{\mathbb{C}_{002}}{\bar{Z}^2}-\frac{\mathbb{C}_{011}}{\bar{H}\bar{Z}}\right)$
$\check{\mathcal{T}}_{2u}(A)$	$\frac{1}{u}\bar{H}\left(\frac{3}{8}\frac{\mathbb{C}_{002}}{\bar{Z}^2}-\frac{1}{2}\frac{\mathbb{C}_{011}}{\bar{H}\bar{Z}}\right)$
$\check{\mathcal{T}}_{1m}(A)$	$\bar{H}\left(\frac{1}{m}\left(\frac{\mathbb{C}_{200}}{\bar{G}^2}+\frac{3}{8}\frac{\mathbb{C}_{002}}{\bar{Z}^2}-\frac{\mathbb{C}_{110}}{\bar{G}\bar{H}}-\frac{1}{2}\frac{\mathbb{C}_{011}}{\bar{H}\bar{Z}}+\frac{1}{2}\frac{\mathbb{C}_{101}}{\bar{G}\bar{Z}}\right)+\frac{1}{n}\left(\frac{\mathbb{C}_{110}}{\bar{G}\bar{H}}-\frac{\mathbb{C}_{200}}{\bar{G}^2}-\frac{1}{2}\frac{\mathbb{C}_{101}}{\bar{G}\bar{Z}}\right)\right)$
$\check{\mathcal{T}}_{2m}(A)$	$\bar{H}\left(\frac{1}{m}\left(\frac{\mathbb{C}_{200}}{\bar{G}^2}-\frac{\mathbb{C}_{110}}{\bar{G}\bar{H}}\right)+\frac{1}{n}\left(\frac{3}{8}\frac{\mathbb{C}_{002}}{\bar{Z}^2}-\frac{\mathbb{C}_{200}}{\bar{G}^2}+\frac{\mathbb{C}_{110}}{\bar{G}\bar{H}}-\frac{1}{2}\frac{\mathbb{C}_{011}}{\bar{H}\bar{Z}}\right)\right)$

Table 5: Bias of the estimators under MSRM

Estimator	Expression of Bias
$\hat{\hat{\yen}}_{u}(M)$	0
$\hat{\hat{\yen}}_{m}(M)$	0
$\check{\mathcal{T}}_{1u}(M)$	$\frac{1}{u}\bar{H}^{*}\left(\frac{\mathbb{C}^{*}_{002}}{\bar{Z}^{2}} - \frac{\mathbb{C}^{*}_{011}}{\bar{H}^{*}\bar{Z}}\right)$
$\check{\mathcal{T}}_{2u}(M)$	$\frac{1}{u}\bar{H}^{*}\left(\frac{3}{8}\frac{\mathbb{C}^{*}_{002}}{\bar{Z}^{2}} - \frac{1}{2}\frac{\mathbb{C}^{*}_{011}}{\bar{H}^{*}\bar{Z}}\right)$
$\check{\mathcal{T}}_{1m}(M)$	$\bar{H}^{*}\left(\frac{1}{m}\left(\frac{\mathbb{C}^{*}_{200}}{\bar{G}^{*2}} + \frac{3}{8}\frac{\mathbb{C}^{*}_{002}}{\bar{Z}^{2}} - \frac{\mathbb{C}^{*}_{110}}{\bar{G}^{*}\bar{H}^{*}} - \frac{1}{2}\frac{\mathbb{C}^{*}_{011}}{\bar{H}^{*}\bar{Z}} + \frac{1}{2}\frac{\mathbb{C}^{*}_{101}}{\bar{G}^{*}\bar{Z}}\right) + \frac{1}{n}\left(\frac{\mathbb{C}^{*}_{110}}{\bar{G}^{*}\bar{H}^{*}} - \frac{\mathbb{C}^{*}_{200}}{\bar{G}^{*2}} - \frac{1}{2}\frac{\mathbb{C}^{*}_{101}}{\bar{G}^{*}\bar{Z}}\right)\right)$
$\check{\mathcal{T}}_{2m}(M)$	$\bar{H}^{*}\left(\frac{1}{m}\left(\frac{\mathbb{C}^{*}_{200}}{\bar{G}^{*2}} - \frac{\mathbb{C}^{*}_{110}}{\bar{G}^{*}\bar{H}^{*}}\right) + \frac{1}{n}\left(\frac{3}{8}\frac{\mathbb{C}^{*}_{002}}{\bar{Z}^{2}} - \frac{\mathbb{C}^{*}_{200}}{\bar{G}^{*2}} + \frac{\mathbb{C}^{*}_{110}}{\bar{G}^{*}\bar{H}^{*}} - \frac{1}{2}\frac{\mathbb{C}^{*}_{011}}{\bar{H}^{*}\bar{Z}}\right)\right)$

where $\mathbb{C}_{rtq} = E\left[\left(g_i - \bar{G}\right)^{r}\left(h_i - \bar{H}\right)^{t}\left(z_i - \bar{Z}\right)^{q}\right]$ and $\mathbb{C}^{*}_{rtq} = E\left[\left(g^{*}_i - \bar{G}^{*}\right)^{r}\left(h^{*}_i - \bar{H}^{*}\right)^{t}\left(z_i - \bar{Z}\right)^{q}\right]$; $(r, t, q) \geq 0$.

Since $\hat{\hat{\yen}}_{u}(\tau)$ and $\hat{\hat{\yen}}_{m}(\tau)$ are unbiased for population mean hence the estimator $\hat{\hat{\yen}}(\tau)$ is also unbiased for population mean. The bias of the estimators $\check{\mathcal{T}}_{ij}(\tau); (i, j=1, 2)$ to the first order of approximations are obtained as

$$B\left[\check{\mathcal{T}}_{ij}(\tau)\right] = \varpi_{ij}(\tau)\, B\left(\check{\mathcal{T}}_{iu}(\tau)\right) + \left[1 - \varpi_{ij}(\tau)\right] B\left(\check{\mathcal{T}}_{jm}(\tau)\right); (i, j=1, 2), \quad (1)$$

Table 6: Variance and mean squared errors of the estimators $\hat{\text{¥}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau);(\text{i, j=1, 2})$

Estimator	Expression of variance/ mean squared errors under ASRM	Expression of variance/ mean squared errors under MSRM
$\hat{\text{¥}}_u(\tau)$	$\frac{1}{u}A_1\,\bar{H}^2$	$\frac{1}{u}A_1^*\,\bar{H}^{*2}$
$\hat{\text{¥}}_m(\tau)$	$\left(\frac{1}{m}A_2+\frac{1}{n}A_3\right)\bar{H}^2$	$\left(\frac{1}{m}A_2^*+\frac{1}{n}A_3^*\right)\bar{H}^{*2}$
$\check{\mathcal{T}}_{1u}(\tau)$	$\frac{1}{u}B_1\,\bar{H}^2$	$\frac{1}{u}B_1^*\,\bar{H}^{*2}$
$\check{\mathcal{T}}_{2u}(\tau)$	$\frac{1}{u}B_2\,\bar{H}^2$	$\frac{1}{u}B_2^*\,\bar{H}^{*2}$
$\check{\mathcal{T}}_{1m}(\tau)$	$\left(\frac{1}{m}B_3+\frac{1}{n}B_4\right)\bar{H}^2$	$\left(\frac{1}{m}B_3^*+\frac{1}{n}B_4^*\right)\bar{H}^2$
$\check{\mathcal{T}}_{2m}(\tau)$	$\left(\frac{1}{m}B_5+\frac{1}{n}B_6\right)\bar{H}^2$	$\left(\frac{1}{m}B_5^*+\frac{1}{n}B_6^*\right)\bar{H}^2$

Hence final expression of the variance and mean squared errors of the estimators $\hat{\text{¥}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau);(\text{i, j=1, 2})$ are obtained as

$$V\left[\hat{\text{¥}}(\tau)\right]=\xi^2(\tau)\,V\left[\hat{\text{¥}}_u(\tau)\right]+\left[1-\xi(\tau)\right]^2V\left(\hat{\text{¥}}_m(\tau)\right)+2\,\xi(\tau)\left[1-\xi(\tau)\right]\mathrm{Cov}\left(\hat{\text{¥}}_u(\tau),\hat{\text{¥}}_m(\tau)\right) \quad (2)$$

$$M\left[\check{\mathcal{T}}_{ij}(\tau)\right]=\varpi_{ij}^2(\tau)\,M\left[\check{\mathcal{T}}_{iu}(\tau)\right]+\left[1-\varpi_{ij}(\tau)\right]^2M\left[\check{\mathcal{T}}_{jm}(\tau)\right]+2\,\varpi_{ij}(\tau)\left[1-\varpi_{ij}(\tau)\right]\mathrm{Cov}\left(\check{\mathcal{T}}_{iu}(\tau),\check{\mathcal{T}}_{jm}(\tau)\right) \quad (3)$$

where $A_1=C_h^2,\ A_2=\left(1-\rho_{hg}^2\right)C_h^2,\ A_3=\rho_{hg}^2C_h^2,\ A_1^*=C_{h^*}^2,\ A_2^*=\left(1-\rho_{hg}^{*2}\right)C_{h^*}^2,\ A_3^*=\rho_{hg}^{*2}C_{h^*}^2,$

$B_1=C_h^2+C_z^2-2\rho_{hz}C_hC_z,\ B_2=C_h^2+\frac{1}{4}C_z^2-\rho_{hz}C_hC_z,$

$B_3=C_h^2+C_g^2+\frac{1}{4}C_z^2-2\rho_{hg}C_hC_g-\rho_{hz}C_hC_z+\rho_{gz}C_gC_z,$

$B_4=2\rho_{hg}C_hC_g-C_g^2-\rho_{gz}C_gC_z,\ B_5=C_h^2+C_g^2-2\rho_{hg}C_hC_g,\ B_6=\frac{1}{4}C_z^2-C_g^2+2\rho_{hg}C_hC_g-\rho_{hz}C_hC_z$ and

$$B_1^* = C_{h^*}^2 + C_z^2 - 2\rho_{hz}^* C_{h^*} C_z,\; B_2^* = C_{h^*}^2 + \frac{1}{4}C_z^2 - \rho_{hz}^* C_{h^*} C_z,\; B_3^* = C_{h^*}^2 + C_{g^*}^2 + \frac{1}{4}C_z^2 - 2\rho_{hg}^* C_{h^*} C_{g^*} - \rho_{hz}^* C_{h^*} C_z + \rho_{gz}^* C_{g^*} C_z,$$

$$B_4^* = 2\rho_{hg}^* C_{h^*} C_{g^*} - C_{g^*}^2 - \rho_{gz}^* C_{g^*} C_z,\; B_5^* = C_{h^*}^2 + C_{g^*}^2 - 2\rho_{hg}^* C_{h^*} C_{g^*},\; B_6^* = \frac{1}{4}C_z^2 - C_{g^*}^2 + 2\rho_{hg}^* C_{h^*} C_{g^*} - \rho_{hz}^* C_{h^*} C_z,$$

$\text{Cov}\left(\hat{\text{¥}}_u(\tau), \hat{\text{¥}}_m(\tau)\right)=0$ and $\text{Cov}\left(\check{\mathcal{T}}_{iu}(\tau), \check{\mathcal{T}}_{jm}(\tau)\right)=0.$

2.5.2. Minimum Variance and Mean Squared Errors of the Proposed Estimators

Since the variance and mean squared errors of the estimators obtained in equation (2) and equation (3) are the functions of unknown constants $\xi(\tau)$ and $\varpi_{ij}(\tau);(i, j = 1, 2)$, therefore, they are minimized with respect to $\xi(\tau)$ and $\varpi_{ij}(\tau)$ respectively and subsequently the optimum values of $\xi(\tau)$, $\varpi_{ij}(\tau);(i, j = 1, 2)$ so obtained are given as

Table 7: Optimum values of $\xi(\tau)$ and $\varpi_{ij}(\tau);(i, j = 1, 2)$

$\xi(\tau)/\varpi_{ij}(\tau)$	**Optimum value under ASRM**	**Optimum value under MSRM**
$\xi(\tau)_{\text{opt.}}$	$\frac{\chi[\chi A_3 - (A_2 + A_3)]}{[\chi^2 A_3 - \chi(A_3 + A_2 - A_1) - A_1]}$	$\frac{\chi^*[\chi^* A_3^* - (A_2^* + A_3^*)]}{[\chi^2 A_3^* - \chi(A_3^* + A_2^* - A_1^*) - A_1^*]}$
$\varpi_{11}(\tau)_{\text{opt.}}$	$\frac{\mu_{11}[\mu_{11} B_4 - (B_3 + B_4)]}{[\mu_{11}^2 B_4 - \mu_{11}(B_3 + B_4 - B_1) - B_1]}$	$\frac{\mu_{11}^*[\mu_{11}^* B_4^* - (B_3^* + B_4^*)]}{[\mu_{11}^{*2} B_4^* - \mu_{11}^*(B_3^* + B_4^* - B_1^*) - B_1^*]}$
$\varpi_{12}(\tau)_{\text{opt.}}$	$\frac{\mu_{12}[\mu_{12} B_6 - (B_5 + B_6)]}{[\mu_{12}^2 B_6 - \mu_{12}(B_5 + B_6 - B_1) - B_1]}$	$\frac{\mu_{12}^*[\mu_{12}^* B_6^* - (B_5^* + B_6^*)]}{[\mu_{12}^{*2} B_6^* - \mu_{12}^*(B_5^* + B_6^* - B_1^*) - B_1^*]}$
$\varpi_{21}(\tau)_{\text{opt.}}$	$\frac{\mu_{21}[\mu_{21} B_4 - (B_3 + B_4)]}{[\mu_{21}^2 B_4 - \mu_{21}(B_3 + B_4 - B_2) - B_2]}$	$\frac{\mu_{21}^*[\mu_{21}^* B_4^* - (B_3^* + B_4^*)]}{[\mu_{21}^{*2} B_4^* - \mu_{21}^*(B_3^* + B_4^* - B_2^*) - B_2^*]}$
$\varpi_{22}(\tau)_{\text{opt.}}$	$\frac{\mu_{22}[\mu_{22} B_6 - (B_5 + B_6)]}{[\mu_{22}^2 B_6 - \mu_{22}(B_5 + B_6 - B_2) - B_2]}$	$\frac{\mu_{22}^*[\mu_{22}^* B_6^* - (B_5^* + B_6^*)]}{[\mu_{22}^{*2} B_6^* - \mu_{22}^*(B_5^* + B_6^* - B_2^*) - B_2^*]}$

Table 8: Optimum values of $V\left[\hat{\bar{¥}}(\tau)\right]$ and $M\left[\check{\mathcal{T}}_{ij}(\tau)\right]$ for $(i, j=1, 2)$

	Optimum value under ASRM	**Optimum value under MSRM**
$V\left[\hat{\bar{¥}}(\tau)\right]_{opt.}$	$\frac{1}{n}\frac{[\chi A_4 - A_5]}{\left[\chi^2 A_3 - \chi A_6 - A_1\right]}$	$\frac{1}{n}\frac{[\chi^* A_4^* - A_5^*]}{\left[\chi^{*2} A_3^* - \chi^* A_6^* - A_1^*\right]}$
$M\left[\check{\mathcal{T}}_{11}(\tau)\right]_{opt.}$	$\frac{1}{n}\frac{[\mu_{11} C_1 - C_2]}{\left[\mu_{11}^2 B_4 - \mu_{11} C_3 - B_1\right]}$	$\frac{1}{n}\frac{[\mu_{11}^* C_1^* - C_2^*]}{\left[\mu_{11}^{*2} B_4^* - \mu_{11}^* C_3^* - B_1^*\right]}$
$M\left[\check{\mathcal{T}}_{12}(\tau)\right]_{opt.}$	$\frac{1}{n}\frac{[\mu_{12} C_4 - C_5]}{\left[\mu_{12}^2 B_6 - \mu_{12} C_6 - B_1\right]}$	$\frac{1}{n}\frac{[\mu_{12}^* C_4^* - C_5^*]}{\left[\mu_{12}^{*2} B_6^* - \mu_{12}^* C_6^* - B_1^*\right]}$
$M\left[\check{\mathcal{T}}_{21}(\tau)\right]_{opt.}$	$\frac{1}{n}\frac{[\mu_{21} C_7 - C_8]}{\left[\mu_{21}^2 B_4 - \mu_{21} C_9 - B_2\right]}$	$\frac{1}{n}\frac{[\mu_{21}^* C_7^* - C_8^*]}{\left[\mu_{21}^{*2} B_4^* - \mu_{21}^* C_9^* - B_2^*\right]}$
$M\left[\check{\mathcal{T}}_{22}(\tau)\right]_{opt.}$	$\frac{1}{n}\frac{[\mu_{22} C_{10} - C_{11}]}{\left[\mu_{22}^2 B_6 - \mu_{22} C_{12} - B_2\right]}$	$\frac{1}{n}\frac{[\mu_{22}^* C_{10}^* - C_{11}^*]}{\left[\mu_{22}^{*2} B_6^* - \mu_{22}^* C_{12}^* - B_2^*\right]}$

Where

$A_4 = A_1A_3,\ A_5 = A_1(A_2 + A_3),\ A_6 = A_2 + A_3 - A_1,\ A_4^* = A_1^*A_3^*,\ A_5^* = A_1^*(A_2^* + A_3^*),$

$A_6^* = A_2^* + A_3^* - A_1^*,\ C_1 = B_1B_4,\ C_2 = B_1(B_3 + A_4),\ C_3 = B_3 + B_4 - B_1,\ C_4 = B_1B_6,\ C_5 = B_1(B_5 + B_6),$

$C_6 = B_5 + B_6 - B_1,\ C_7 = B_2B_4,\ C_8 = B_2(B_3 + B_4),\ C_9 = B_3 + B_4 - B_2,\ C_{10} = B_2B_6,\ C_{11} = B_2(B_5 + B_6)\ ,$

$C_{12} = B_5 + B_6 - B_2,\ C_1^* = B_1^*B_4^*,\ C_2^* = B_1^*(B_3^* + A_4^*),\ C_3^* = B_3^* + B_4^* - B_1^*,\ C_4^* = B_1^*B_6^*,\ C_5^* = B_1^*(B_5^* + B_6^*),$

$C_6^* = B_5^* + B_6^* - B_1^*,\ C_7^* = B_2^*B_4^*,\ C_8^* = B_2^*(B_3^* + B_4^*),\ C_9^* = B_3^* + B_4^* - B_2^*,\ C_{10}^* = B_2^*B_6^*,\ C_{11}^* = B_2^*(B_5^* + B_6^*),$

$C_{12}^* = B_5^* + B_6^* - B_2^*$, $\mu_{ij}(i, j = 1, 2)$ and $\mu_{ij}^*(i, j = 1, 2)$ are the fractions of the sample drawn afresh at the current(second) wave under ASRM and MSRM respectively.

2.5.3. Optimum Rotation Rate for the Proposed Estimators

Since the mean squared errors of the proposed estimators $\hat{\bar{¥}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau);(i, j=1, 2)$ are the functions of the μ_{ij} and $\mu_{ij}^*;(i, j = 1, 2)$ which are nothing but the rotation rates or the fractions of sample to be drawn afresh at current wave. Since less the sample need to be drawn afresh, less is the total cost of survey, hence to estimate population mean with maximum precision and minimum cost, the variance and mean squared errors of the estimators $\hat{\hat{¥}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau);(i, j=1, 2)$ respectively obtained in Table 8 have been optimized with respect to μ_{ij} and μ^{*}_{ij}; $(i, j = 1, 2)$ respectively. Hence optimum rotation rates have been obtained for each of the estimators $\hat{\hat{¥}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau);(i, j=1, 2)$ and are given as:

Table 9: Optimum Rotation Rate for the Proposed Estimators $\hat{\hat{¥}}(\tau)$ **and** $\check{\mathcal{T}}_{ij}(\tau);(i, j=1, 2)$

	Optimum Rotation Rates under ASRM		**Optimum Rotation Rates under MSRM**
$\hat{\chi}$	$\frac{A_8 \pm \sqrt{A_8^2 - A_7 A_9}}{A_7}$	$\hat{\chi}^*$	$\frac{A_8^* \pm \sqrt{A_8^{*2} - A_7^* A_9^*}}{A_9^*}$
$\hat{\mu}_{11}$	$\frac{D_2 \pm \sqrt{D_2^2 - D_1 D_3}}{D_1}$	$\hat{\mu}^*_{11}$	$\frac{D_2^* \pm \sqrt{D_2^{*2} - D_1^* D_3^*}}{D_1^*}$
$\hat{\mu}_{12}$	$\frac{D_5 \pm \sqrt{D_5^2 - D_4 D_6}}{D_4}$	$\hat{\mu}^*_{12}$	$\frac{D_5^* \pm \sqrt{D_5^{*2} - D_4^* D_6^*}}{D_4^*}$
$\hat{\mu}_{21}$	$\frac{D_8 \pm \sqrt{D_8^2 - D_7 D_9}}{D_7}$	$\hat{\mu}^*_{21}$	$\frac{D_8^* \pm \sqrt{D_8^{*2} - D_7^* D_9^*}}{D_7^*}$
$\hat{\mu}_{22}$	$\frac{D_{11} \pm \sqrt{D_{11}^2 - D_{10} D_{12}}}{D_{10}}$	$\hat{\mu}^*_{22}$	$\frac{D_{11}^* \pm \sqrt{D_{11}^{*2} - D_{10}^* D_{12}^*}}{D_{10}^*}$

where

$A_9 = A_4A_1 + A_5A_6,\ A_8 = A_3A_5,\ A_7 = A_3A_4,\ A_9^* = A_4^*A_1^* + A_5^*A_6^*,\ A_8^* = A_3^*A_5^*,\ A_7^* = A_3^*A_4^*$

$D_1 = B_4C_1,\ D_2 = B_4C_2,\ D_3 = B_1C_1 + C_2C_3,\ D_4 = B_6C_4,\ D_5 = B_6C_5,\ D_6 = B_1C_4 + C_5C_6$

$D_7 = B_4C_7,\ D_8 = B_4C_8,\ D_9 = B_2C_7 + C_8C_9,\ D_{10} = B_6C_{10},\ D_{11} = B_6C_{11}$ and $D_{12} = B_2C_{10} + C_{11}C_{12}$.

$D_1^* = B_4^*C_1^*,\ D_2^* = B_4^*C_2^*,\ D_3^* = B_1^*C_1^* + C_2^*C_3^*,\ D_4^* = B_6^*C_4^*,\ D_5^* = B_6^*C_5^*,\ D_6^* = B_1^*C_4^* + C_5^*C_6^*$

$D_7^* = B_4^* C_7^*,\ D_8^* = B_4^* C_8^*,\ D_9^* = B_2^* C_7^* + C_8^* C_9^*,\ D_{10}^* = B_6^* C_{10}^*,\ D_{11}^* = B_6^* C_{11}^*$ and $D_{12}^* = B_2^* C_{10}^* + C_{11}^* C_{12}^*$.

Substituting the optimum values $\hat{\chi}, \hat{\chi}^*, \hat{\mu}_{ij}$ and $\hat{\mu}_{ij}^*; (i, j = 1, 2)$ in the minimum variance and mean squared errors of the estimators $\hat{\hat{¥}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau)$ obtained in Table 8, the optimum values of the variance and mean squared errors of the estimators $\hat{\hat{¥}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau); (i, j = 1, 2)$ respectively with respect to $\xi(\tau)$ and $\varpi_{ij}(\tau)$ as well as μ_{ij} and $\mu_{ij}^*; (i, j = 1, 2)$ have been obtained and are given as

Table 10: The optimum variance and mean squared errors of the estimators $\hat{\hat{¥}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau); (i, j = 1, 2)$

	Optimum variance/mean squared errors under ASRM	Optimum variance/mean squared errors under MSRM
$V\left[\hat{\hat{¥}}(\tau)\right]_{opt.}^*$	$\frac{1}{n}\frac{\left[\chi^{(0)} A_4 - A_5\right]}{\left[\chi^{(0)2} A_3 - \chi^{(0)} A_6 - A_1\right]}$	$\frac{1}{n}\frac{\left[\chi^{*(0)} A_4^* - A_5^*\right]}{\left[\chi^{*(0)^2} A_3^* - \chi^{*(0)} A_6^* - A_1^*\right]}$
$M\left[\check{\mathcal{T}}_{11}(\tau)\right]_{opt.}^*$	$\frac{\left[\mu_{11}^{(0)} C_1 - C_2\right]}{n\left[\mu_{11}^{(0)2} B_4 - \mu_{11}^{(0)} C_3 - B_1\right]}$	$\frac{\left[\mu_{11}^{*(0)} C_1 - C_2\right]}{n\left[\mu_{11}^{*(0)^2} B_4 - \mu_{11}^{*(0)} C_3 - B_1\right]}$
$M\left[\check{\mathcal{T}}_{12}(\tau)\right]_{opt.}^*$	$\frac{\left[\mu_{12}^{(0)} C_4 - C_5\right]}{n\left[\mu_{12}^{(0)2} B_6 - \mu_{12}^{(0)} C_6 - B_1\right]}$	$\frac{\left[\mu_{12}^{*(0)} C_4 - C_5\right]}{n\left[\mu_{12}^{*(0)^2} B_6 - \mu_{12}^{*(0)} C_6 - B_1\right]}$
$M\left[\check{\mathcal{T}}_{21}(\tau)\right]_{opt.}^*$	$\frac{\left[\mu_{21}^{(0)} C_7 - C_8\right]}{n\left[\mu_{21}^{(0)2} B_4 - \mu_{21}^{(0)} C_9 - B_2\right]}$	$\frac{\left[\mu_{21}^{*(0)} C_7 - C_8\right]}{n\left[\mu_{21}^{*(0)^2} B_4 - \mu_{21}^{*(0)} C_9 - B_2\right]}$
$M\left[\check{\mathcal{T}}_{22}(\tau)\right]_{opt.}^*$	$\frac{\left[\mu_{22}^{(0)} C_{10} - C_{11}\right]}{n\left[\mu_{22}^{(0)2} B_6 - \mu_{22}^{(0)} C_{12} - B_2\right]}$	$\frac{\left[\mu_{22}^{*(0)} C_{10} - C_{11}\right]}{n\left[\mu_{22}^{*(0)^2} B_6 - \mu_{22}^{*(0)} C_{12} - B_2\right]}$

3. Modelling the Total Cost for the Survey

When a survey, constituting sensitive issues, is deigned, the focus is centred to the total cost of the survey. Hence the model for total cost including design and analysis over two successive waves is proposed as:

$$C_T(\tau) = nc_p(\tau) + mc_r(\tau) + uc_d(\tau), \tag{4}$$

where $\tau = \begin{cases} \text{M} & \text{for Multiplicative Scrambled Response Model} \\ \text{A} & \text{for Additive Scrambled Response Model} \end{cases}$

$C_T(\tau)$: The total cost of sample survey at current (second) wave;

$c_p(\tau)$: The average per unit cost of investigating and processing data at previous (first) wave,

$c_r(\tau)$: The average per unit cost of investigating and processing retained data at current wave,

$c_d(\tau)$: The average per unit cost of investigating and processing freshly drawn data at current wave.

Remark 3.1: $c_p(\tau) < c_r(\tau) < c_d(\tau)$, When a survey is conducted on successive waves, the cost of investigating a single unit involved in the survey sample should be greater than before (at previous wave) since as time passes by different commodities (software) and services (human resources, daily wages and conveyance) become expensive so the cost incurring at second wave increases in a considerable amount. Also the average cost of investigating a retained unit from previous wave should be lesser than investigating a freshly drawn sample unit since survey investigator has some experiences from the previous wave and hence the investigator can trace the retained sample units easily as compared to freshly drawn sample units which reduces the cost in investigating but on the other hand due to time lag between the successive waves, cost of investigating a retained sample unit rises as compared to the previous wave.

Theorem 3.1: The optimum total cost for the proposed estimators $\hat{\hat{¥}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau)$; $(\text{i, j}=1, 2)$ are obtained as

$$C_T\left[\hat{\hat{¥}}(\text{A})\right] = n\left[\left(c_p(\text{A}) + c_s(\text{A})\right) + \left(1 - \chi^{(0)}\right)\left(c_r(\text{A}) - c_s(\text{A})\right)\right] \quad (5)$$

$$C_T\left[\hat{\hat{¥}}(\text{M})\right] = n\left[\left(c_p(\text{M}) + c_s(\text{M})\right) + \left(1 - \chi^{*(0)}\right)\left(c_r(\text{M}) - c_s(\text{M})\right)\right] \quad (6)$$

$$C_T\left[\check{\mathcal{T}}_{ij}(A)\right]= n\left[\left(c_p(A)+c_s(A)\right)+\left(1-\mu_{ij}^{(0)}\right)\left(c_r(A)-c_s(A)\right)\right]\ \forall\ i, j=1, 2 \tag{7}$$

$$C_T\left[\check{\mathcal{T}}_{ij}(M)\right]= n\left[\left(c_p(M)+c_s(M)\right)+\left(1-\mu_{ij}^{*(0)}\right)\left(c_r(M)-c_s(M)\right)\right]\ \forall\ i, j=1, 2 \tag{8}$$

Remark 3.2: The optimum total costs obtained in equation (5) to (8) are dependent on the value of sample size (n). Therefore, if a suitable guess of sample size is available, it can be used for obtaining optimum total cost of the survey by above equation. However, in the absence of suitable guess, sample size may be estimated by following Cochran (1977).

4. Efficiency Comparison

4.1. Estimator $\hat{\bar{¥}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau)$ versus Estimator $\bar{h}_n(\tau)$

To evaluate the performance of the proposed estimators, the estimators $\hat{\bar{¥}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau)$ at optimum conditions, they are compared with the scrambled sample mean estimator $\bar{h}_n(\tau)$, when there is no matching from previous wave. Since the scrambled sample mean estimator $\bar{h}_n(\tau)$ is unbiased for population mean, so variance of the estimator $\bar{h}_n(\tau)$ is given by

$$V\left[\bar{h}_n(A)\right] = \frac{1}{n}\left(S_y^2+S_s^2\right), \tag{9}$$

$$V\left[\bar{h}_n(M)\right] = \frac{1}{n}\left(S_y^2 S_s^{*2}+S_s^{*2}\bar{Y}^2+S_y^2\bar{S}^{*2}\right) \tag{10}$$

The percent relative efficiencies $E_1(\tau)$ and $E_{ij}^1(\tau)$ of the estimator $\hat{\bar{¥}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau)$ (under optimum conditions) with respect to $\bar{h}_n(\tau)$ are given by

$$E_1(\tau)= \frac{V\left[\bar{h}_n(\tau)\right]}{V\left[\hat{\bar{¥}}(\tau)\right]^*_{opt.}} \times 100 \tag{11}$$

$$E^{1}_{ij}(\tau)= \frac{V\left[\bar{h}_n(\tau)\right]}{M\left[\check{\mathcal{T}}_{ij}(\tau)\right]^{*}_{opt.}} \times 100\,; \ (i, j=1, 2). \tag{12}$$

4.2. Estimator $\check{\mathcal{T}}_{ij}(\tau)$ versus Estimator $\hat{\hat{¥}}(\tau)$

The percent relative efficiencies $E^{2}_{ij}(\tau)$ of the estimator $\check{\mathcal{T}}_{ij}(\tau)$ (under optimum conditions) with respect to $\hat{\hat{¥}}(\tau)$ are obtained as

$$E^{2}_{ij}(\tau)= \frac{V\left[\hat{\hat{¥}}(\tau)\right]^{*}_{opt.}}{M\left[\check{\mathcal{T}}_{ij}(\tau)\right]^{*}_{opt.}} \times 100 \tag{13}$$

5: Choice of the Distribution of Scrambling Variable

5.1. Scrambling variable under ASRM

Pollock and Bek (1976) did not lay down certain assumption for choosing the distribution of the scrambling variable S, since S has to be generated before conducting the survey for collecting the response to ensure the privacy of the respondents. Additive scrambled response model still provides us certain freedom to apply it practically. For a quantitative sensitive character, response may either be positive or zero. If response is some positive quantity then adding a scrambling variable would not alter the response when the scrambling variable follows some certain prior known distribution. Even if the response is zero then also additive scrambled response model would be good to go as mean and variance of the scrambling variable is known and should have been chosen in such a way that the mean value of scrambling variable would not make a huge impact on mean value of sensitive character. So for applying an additive scrambled response model it should be kept in consideration that mean and variance of distribution of scrambling variable should not alter the mean value of sensitive character provided that the respondents agree to answer truthfully. So while conducting a survey related to the drug usage of undergraduate students of a college, we have assumed that scrambling variable S follows normal distribution with mean zero and variance 1. Here considering mean value of scrambling variable zero makes least impact on the mean value of the sensitive variable as mentioned above.

5.2. Scrambling variable under MSRM

When the multiplicative scrambled response model was first studied by Pollock and Bek (1976), no assumptions were set for generating distribution of scrambling variable but when Eichhorn and Hayre (1983) studied this model in depth they provided certain specifications to be followed for the generating the scrambling variable. They suggested that, to estimate the population mean of the sensitive character $X(Y) > 0$ the scrambling variable $S^* > 0$ with $E(S^*) = \bar{S}^*$, $V(S^*) = S_s^{*2}$ should be chosen such that $\wp = S_s^* / \bar{S}^*$ is as small as possible. Also it has been shown reasonable that $\text{median}(S^*) \approx 1$. The numerical illustration done in the next section assumes the scrambling variable S^* a normal variate with mean one and variance 0.36 this makes $\wp = 0.6$ which a small value. Various methods have also been suggested when the scrambling variable assumes negative and zero values, Eichhorn and Hayre (1983) may be cited for detailed procedures.

6. Numerical Illustrations and Monte Carlo Simulation

6.1. Empirical study

For practicing the use of the proposed estimators under two different scrambled response model namely ASRM and MSRM over two successive waves, numerical illustration has been worked out for a completely known population with following population parameters:

$N=51$, $n=20$, $S_x^2 = 4.3451\times10^6$, $S_y^2 = 4.1604\times10^6$, $S_z^2 = 4.2152\times10^6$, $\bar{X}=1923.3$, $\bar{Y}=1947.8$, $\bar{Z}=1923.3$, $\rho_{yx}=0.7$, $\rho_{xz}=0.7$, $\rho_{yz}=0.7$.

And also hypothetical input costs were considered to get an idea about the optimum total cost of the survey.

For using the ASRM it has been assumed that the scrambling variable $S \sim N(0,1)$ and using MSRM, it has been considered that scrambling $S^{*} \sim N(1, 0.6)$. Hence the results obtained have been represented in Table 11, Table 12 and Table 13.

Table 11: Variance of estimators $\bar{h}_n$ and $\hat{\hat{¥}}$ under ASRM and MSRM.

Estimator	Variance under ASRM	Variance under MSRM
$\bar{h}_n(\tau)$	4.1604×10^{6}	1.2115×10^{7}
$\hat{\hat{¥}}(\tau)$	3.5658×10^{6}	9.7583×10^{6}

Table 12: Empirical results when the proposed estimators $\hat{\hat{¥}}(\tau)$ and $\check{\mathcal{T}}_{ij}(\tau);(i, j=1, 2)$ have been compared to the scrambled sample mean estimator.

Under ASRM				**Under MSRM**			
Estimators	**Optimum rotation rate**	**Percent relative efficiency**	**Optimum Total cost**	**Estimators**	**Optimum rotation rate**	**Percent relative efficiency**	**Optimum Total cost**
$\hat{\hat{¥}}(A)$	**0.5834**	**116.67**	**₹2258.30**	$\hat{\hat{¥}}(M)$	**0.6207**	**124.14**	**₹2262.10**
$\check{\mathcal{T}}_{11}(A)$	**0.5511**	**150.30**	**₹2255.10**	$\check{\mathcal{T}}_{11}(M)$	**0.5313**	**130.76**	**₹2253.10**
$\check{\mathcal{T}}_{12}(A)$	**#**	**-**	**-**	$\check{\mathcal{T}}_{12}(M)$	**0.5651**	**136.65**	**₹2256.50**
$\check{\mathcal{T}}_{21}(A)$	**0.4370**	**159.99**	**₹2243.70**	$\check{\mathcal{T}}_{21}(M)$	**0.5591**	**132.45**	**₹2255.90**
$\check{\mathcal{T}}_{22}(A)$	**0.4837**	**177.11**	**₹2248.40**	$\check{\mathcal{T}}_{22}(M)$	**0.5843**	**138.42**	**₹2258.40**

Note: " # " represents that the optimum rotation rate does not exist.

Table 13: Empirical results when the proposed estimators $\check{\mathcal{T}}_{ij}(\tau);(i, j=1, 2)$ have been compared to the Estimator $\hat{\hat{¥}}(\tau)$

Under ASRM			Under MSRM		
Estimators	Optimum rotation rate	Percent relative efficiency	Estimators	Optimum rotation rate	Percent relative efficiency
$\check{\mathcal{T}}_{11}(A)$	0.5511	128.81	$\check{\mathcal{T}}_{11}(M)$	0.5313	105.33
$\check{\mathcal{T}}_{12}(A)$	#	-	$\check{\mathcal{T}}_{12}(M)$	0.5651	110.07
$\check{\mathcal{T}}_{21}(A)$	0.4370	137.12	$\check{\mathcal{T}}_{21}(M)$	0.5591	106.68
$\check{\mathcal{T}}_{22}(A)$	0.4837	151.80	$\check{\mathcal{T}}_{22}(M)$	0.5843	111.50

Note: " # " represents that the optimum rotation rate does not exist.

From the results obtained in Table 11, it is observed that scrambles sample mean estimator under ASRM is better than the scrambled sample mean estimator under MSRM. Also the modified Jessen's estimator under scrambled response under ASRM is better than the same estimator under MSRM. Also in Table 12 and Table 13, the proposed estimators $\check{\mathcal{T}}_{ij}(\tau);(i, j=1, 2)$ have an enhanced performance in terms of optimum rotation rate, optimum total cost of the survey and percent relative efficiency with respect to scrambled sample mean estimator and modified Jessen's estimator under scrambled response. It has also been seen in section 5 that applying ASRM is easier as compared to MSRM on successive waves due to less complication and restrictions involved in the selection of scrambling variable for ASRM. Also the above empirical results suggest that beside the conveniences in the application of ASRM, it is reasonably better in terms of cost and precision. Also any quantitative sensitive character may assume a zero value at any time, in such a situation MSRM involves many complexities which again turns time consuming. So to make a survey involving large size and zero valued sensitive responses, less time consuming for the respondents, Application of additive scrambled response is suggested. Therefore, to validate the theoretical results, a case study has been carried out to deal with a sufficiently sensitive issue where fake response is quit prone, hence, an attempt has been made to apply additive scrambled response approach to handle that.

6.2. Case Study: Usage of Drugs (Cigarette, Alcohol, Gutkha, Paan Masala etc.)

For practicing the literal feasibility of the proposed estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$, a case study has been designed for two waves and real data have been collected from 315 under graduate students of a College (University of Delhi), India through a survey conducted on two successive waves. For convenience 315 random numbers (S) have been generated assuming $S \sim N(0,1)$ to retain the mean value of population mean unaffected from mean value of scrambling variable while ensuring the privacy of the respondents. The respondents were presented a bag, full of ball with random number written on them, the respondent had to pick a ball and then he/she had to add his/her answer to that random number which was completely unknown to the interviewer. In this way the interviewer received scrambled response from each respondent. Followings are the sensitive and non-sensitive variables of the interest:

x_i : Average monthly expenditure on drug usage in July, 2015, by the i^{th} student.

y_i : Average monthly expenditure on drug usage in April, 2016, by the i^{th} student.

z_i : Average monthly pocket money from all sources in July, 2015 of the i^{th} student.

And hence the scrambled response was collected from the respondents in the form of $G = X + S$ and $H = Y + S$ with $\bar{S} = 0$ which makes $\bar{G} = \bar{X}$ and $\bar{H} = \bar{Y}$.

Therefore, the optimum rotation rate, percent relative efficiencies of the proposed estimators $\check{\mathcal{T}}_{ij}(A); (i, j=1, 2)$ with respect to scrambled sample mean estimator and modified Jessen's estimator under scrambled response under ASRM and optimum total costs of the survey have been obtained and shown in Table 14. The optimum bias of each proposed estimator has also been calculated and shown in Table 15. Following are the different costs incurred in conducting the survey at two different waves: c_p = ₹ 50.00, c_r = ₹ 60.00 and c_s = ₹ 65.00.

Table 14: Empirical results when the proposed estimators $\check{\mathcal{T}}_{ij}(A)$ have been compared to estimators $\bar{h}_n(A)$ and $\hat{\bar{¥}}(A)$.

Estimator	$\mu_{ij}^{(0)}(i, j=1, 2)$	$\mathbf{E}_{ij}^{1}(\mathbf{A})$	$\mathbf{E}_{ij}^{2}(\mathbf{A})$	$C_T(\check{\mathcal{T}}_{ij}(A))$
$\check{\mathcal{T}}_{11}(A)$	0.7236	182.11	131.90	₹5112.8
$\check{\mathcal{T}}_{12}(A)$	0.7269	**184.16**	**133.39**	₹5113.5
$\check{\mathcal{T}}_{21}(A)$	**0.6487**	162.42	117.64	**₹5096.0**
$\check{\mathcal{T}}_{22}(A)$	0.6563	164.32	119.02	**₹5097.7**

Table 15: Optimum absolute bias of the estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$.

$B(\check{\mathcal{T}}_{ij}(A))$	**n=35**	**n=45**	**n=50**
$\lvert B(\check{\mathcal{T}}_{11}(A))\rvert$	22.36	17.39	15.65
$\lvert B(\check{\mathcal{T}}_{12}(A))\rvert$	21.14	16.44	14.80
$\lvert B(\check{\mathcal{T}}_{21}(A))\rvert$	18.59	14.46	13.01
$\lvert B(\check{\mathcal{T}}_{22}(A))\rvert$	**17.21**	**13.38**	**12.05**

6.2.1. Monte Carlo Simulation Study

For the above said survey data, detailed simulation study has been carried out and thus the simulation results obtained are shown in Table 16.

6.2.1.1. Simulation Algorithm

(i) Choose 5000 samples of size n=45 using simple random sampling without replacement on first wave for both the study (sensitive character) and auxiliary variable (non-sensitive) out of 315.

(ii) Calculate sample mean $\bar{g}_{n|k}$ and $\bar{z}_{n|k}$ for k =1, 2, - - -, 5000.

(iii) Retain m=33 units out of each n=45 sample units of the study and auxiliary variables at the first wave.

(iv) Calculate sample mean $\bar{g}_{m|k}$ and $\bar{z}_{m|k}$ for k= 1, 2, - - -, 5000.

(v) Select u=12 units using simple random sampling without replacement from N-n=270 units of the population for study and auxiliary variables at second (current) wave.

(vi) Calculate sample mean $\bar{h}_{u|k}$, $\bar{h}_{m|k}$ and $\bar{z}_{u|k}$ for k = 1, 2, - - -, 5000.

(vii) Iterate the parameter $\varpi_{ij}(A);(i, j=1, 2)$ from 0.1 to 0.9 with a step of 0.1.

(viii) Calculate the percent relative efficiencies of the proposed estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$ with respect to the scrambled sample mean estimator $\bar{h}_n(A)$ as

$$E(ij) = \frac{\sum_{k=1}^{5000}\left[\bar{h}(A)_{n|k} - \bar{H}\right]^2}{\sum_{k=1}^{5000}\left[\check{\mathcal{T}}(A)_{ij|k} - \bar{H}\right]^2} \times 100\ ;\ (i, j=1, 2);\ k=1, 2, ..., 5000.$$

To exhibit the performance of the proposed estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$, Monte Carlo simulation has been performed for three different sets which are quoted below:

SET I: n=45, u=12, m=33, **SET II:** n=45, u=18, m=27, **SET III:** n=45, u=27, m=18.

Following above simulation algorithm, simulations results have been obtained for all the above three mentioned sets.

Table 16: Monte Carlo simulation results when the proposed estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$ are compared to the scrambled sample mean estimator.

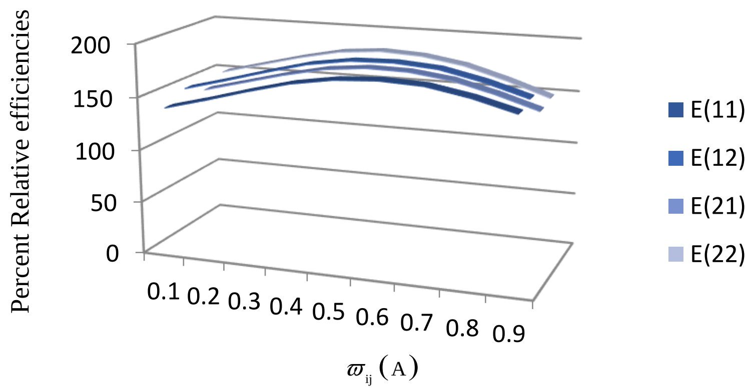
SET \ $\varpi_{ij}(A)$		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
I	E(11)	125.27	129.99	130.44	126.34	119.26	110.00	**	**	**
	E(12)	131.73	136.41	136.98	132.36	124.22	113.89	101.91	**	**
	E(21)	125.64	130.66	131.59	128.01	121.04	111.71	100.82	**	**
	E(22)	132.09	137.08	138.20	134.11	126.07	115.66	103.76	**	**
II	E(11)	101.71	113.80	125.54	136.03	144.43	149.49	150.06	145.89	137.01
	E(12)	118.73	132.03	144.56	154.92	162.21	165.38	163.42	156.57	145.11
	E(21)	101.88	113.89	125.27	135.06	142.30	146.01	144.97	139.41	129.54
	E(22)	118.89	132.01	144.00	153.40	159.24	160.80	157.10	148.87	136.55
III	E(11)	139.02	149.37	160.47	170.54	176.95	179.21	177.40	170.11	159.56
	E(12)	148.96	159.35	170.32	180.44	186.42	187.49	184.41	175.71	163.89
	E(21)	139.41	149.51	159.63	168.04	171.98	171.31	166.29	156.21	143.94
	E(22)	149.37	159.45	169.31	177.54	180.83	178.76	172.35	160.86	147.40

Note: "" represents no gain in the percent relative efficiency.**

7. Mutual Comparison of the estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$

The performances of the proposed estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$ have been elaborated empirically as well as through simulation studies in above section 6 and the results obtained are presented in Table 14 to Table 16. In this section the mutual comparison of the four proposed estimators has been elaborated pictorially given in Figure 7.1.

Figure 7.1: Mutual comparison of the proposed estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$ for set III.



8. Rendition of Results

1) From the numerical illustration of completely known population in Section 6.1, it has been observed that ASRM is better than MSRM under the standard supposition for the distribution of scrambling variables. ASRM is better than MSRM for scrambled sample mean estimator and also for the proposed estimators $\hat{\yen}(A)$ and $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$ in terms of total cost of the survey and precisions of estimates.

2) It has also been noted that the proposed estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$ under ASRM are the best suited estimators over the scrambled mean estimator and the modified Jessen's estimator under scrambled response in terms of cost and precision. The proposed estimator $\check{\mathcal{T}}_{21}(A);(i, j=1, 2)$ is the best performing estimator over the estimators $\bar{h}_n(A)$ and $\hat{\yen}(A)$

3) Results from the Empirical Study based on Case Study

a) From Table 14, we see that $\mu_{ij}^{(0)}(A);(i, j=1, 2)$ exist for each proposed estimators and $\mu_{21}^{(0)}$ is the least.

b) All the four proposed estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$ are efficient over the estimators $\bar{h}_n(A)$ and $\hat{\hat{¥}}(A)$ and the estimator $\check{\mathcal{T}}_{12}(A)$ is most efficient over the estimators $\bar{h}_n(A)$ and $\hat{\hat{¥}}(A)$. This justifies that using a positively correlated non-sensitive auxiliary character is highly rewarding in terms of efficiency.

c) The optimum total cost of the survey conducted on two successive waves has also been calculated while using all four proposed estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$. The optimum total cost of survey is least for the estimator $\check{\mathcal{T}}_{21}(A)$. The estimators $\check{\mathcal{T}}_{21}(A)$ and $\check{\mathcal{T}}_{22}(A)$ provide approximately same optimum total cost for the survey.

d) From Table 15, it is clear that the estimator $\check{\mathcal{T}}_{22}(A)$ is least biased amongst all other proposed estimators and also it is vindicated that for increasing size of sample (n), the bias of all proposed estimators decreases.

2) Results extracted from Monte Carlo simulation Study

a) In Table 16, it can be seen that all the proposed estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$ are efficient over the estimator $\bar{h}_n(A)$. All though for first set, for some choices of $\varpi_{ij}(A)$, all proposed estimators are not efficient.

b) While choosing the different sets for simulation study the empirical results have been taken care of, since empirically the optimum rotation rates have been suggested more than 50% for freshly drawn fraction of sample which has been clearly demonstrated by the simulation results. As the fraction of sample drawn afresh is increased gradually up to 60%, the performance of the proposed estimators has been enhanced.

c) There is no fixed pattern to choose that which one of the four estimators is best in terms of the efficiency but with a minute observation it can be understood that when the choice of $\mu_{ij}(A)$ is stretched near to their empirically optimum $\mu_{ij}^{(0)}(A);(i, j=1, 2)$, the estimator $\check{\mathcal{T}}_{22}(A)$ is most consistent amongst all the others.

3) Results from graphical mutual comparison

a) In Figure 7.1 it is clear that the estimator $\check{\mathcal{T}}_{22}(A)$ is most consistent and efficient over all other proposed estimators.

8. Ratiocination

The rendition of results leads the authors to conclude by assuring from section 5 and 6.1 that over two successive waves, use and application of additive scrambled response model is more feasible and beneficial in terms of cost and precision over the multiplicative scrambled response model. From Section 6 it is quite clear that two types of estimators; one which utilize a non-sensitive auxiliary information and second which doesn't utilize any non-sensitive auxiliary information, both are tremendously better than the scrambled sample mean estimator under ASRM as well as MSRM but since ASRM techniques prevails over MSRM hence further case study has been designed using ASRM approach.

The all four proposed estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$ are good enough to be practiced practically over the estimators $\bar{h}_n(A)$ and $\hat{¥}(A)$ while observing a sensitive character. From empirical results of the case study the estimator $\check{\mathcal{T}}_{12}(A)$ is best in terms of efficiency over $\bar{h}_n(A)$ and $\hat{¥}(A)$ and the estimator $\check{\mathcal{T}}_{21}(A)$ provides the least fraction of sample to be drawn afresh at current wave. This signifies that use of a non-sensitive auxiliary information is appreciable in enhancing the precision of estimates and cost of the survey while estimating population mean of sensitive character while using the ASRM. Here the estimators $\check{\mathcal{T}}_{21}(A)$ and $\check{\mathcal{T}}_{22}(A)$ provide approximately same optimum total cost of the survey and also $\mu_{21}^{(0)}(A)$ and $\mu_{22}^{(0)}(A)$ do not share a big difference in optimum values. The optimum bias of the estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$ has also been computed and the estimator $\check{\mathcal{T}}_{22}(A)$ comes out be the least biased estimator amongst the others that is may be due to the more utilization of non-sensitive auxiliary information and exponential structure as well, but $\check{\mathcal{T}}_{21}(A)$ is approximately equally biased as estimator $\check{\mathcal{T}}_{22}(A)$. In simulation study at closer values of optimum $\mu_{ij}(A)$, estimator $\check{\mathcal{T}}_{22}(A)$ is consistent enough to be considered over other proposed estimators in terms of efficiency. So looking at the overall performance of the estimators, the estimators $\check{\mathcal{T}}_{21}(A)$ and $\check{\mathcal{T}}_{22}(A)$ seems to outperform all other estimators in terms of bias, freshly drawn fraction sample of sample and optimum total cost, since for minimal advantage in precision, the cost of the survey cannot be put on stake in successive sampling while dealing with a sensitive issue. Hence the proposed estimators $\check{\mathcal{T}}_{ij}(A);(i, j=1, 2)$, especially the estimators $\check{\mathcal{T}}_{21}(A)$ and $\check{\mathcal{T}}_{22}(A)$ while accompanying a non-sensitive auxiliary variable, are recommended to survey statisticians for their practical use in surveys indulging sensitive issues pertaining large sample sizes.

UNIT-V

CONCLUSIONS AND FUTURE SCOPE

Conclusions and Future Scope

1. Conclusions

Entire work is stretched over four units on basis of population parameter to be estimated under different circumstances over two occasion successive sampling. The first unit deals with the estimation of population median. The second unit deals with estimation of population mean. Third unit proposes estimators to estimate population mean under non-response of respondents. Forth unit focuses on estimating population mean of sensitive characters where false response is plebeian due to sensitivity of the character to be addressed.

In the first unit an attempt has been made to propose new and different estimators to estimate population median of the study character in two occasion successive sampling since there is not large literature available for the estimation of population median of the study character in two occasion successive sampling.

In Chapter-1, MVLU estimator of population median has been suggested with the aid of completely known auxiliary information available over both occasions. It has been seen that the proposed estimator comes to be better with respect sample median estimator as well as the estimator utilizing no additional auxiliary information. The role of using an extra information has certainly been signified in enhancing the performance of proposed estimator in estimating the population median.

Chapter-2, makes an attempt to work out a factor type estimators to estimate population median without using any extra information on an auxiliary variable. It has been seen that the proposed factor type estimator with input parameter “d” becomes ratio type, product type and dual to ratio type in nature for different values of input parameter d. It has bias and mean squared error asymptotically equal to ratio type estimator for lager value of “d” over two successive waves. Also, it is better than the sample median estimator, ratio type estimator, product type estimator and dual to ratio type estimator at the optimum value of “d” in terms of efficiency and cost.

Bahl and Tuteja (1991) have shown that exponential ratio type estimators are better than the ratio type estimators and regression estimator at certain assumptions, hence an attempt has been made to work out exponential ratio type estimator in two occasion successive sampling. Therefore, four exponential ratio type estimators have been proposed utilizing information on a completely known stable auxiliary information, readily available over both the occasions. Also the proposed estimators turns better in terms of cost and efficiency with respect to the estimator due to Singh et al. (2007) for second quantile and the sample median estimator. Also a mutual comparison of the four proposed exponential ratio type estimators has also been done and it has been found out that the estimator T_{22} is best in terms of cost and efficiency while estimating population median.

In chapter-4, a multivariate generalization of the best performing estimator T_{22} has been done and in the availability of several auxiliary information. The increased level of precision has been shown by comparing the proposed estimator with respect to sample median estimator and estimator due to Singh et al (2007). Also it has been shown theoretically that increasing the number of auxiliary information lead to increased level of efficiency and it reduces the cost of survey as well.

In chapter-5, a possibility has been explored when the auxiliary information may not sustain to be stable and in such a case the four proposed exponential type estimators have been compared to sample median estimator and estimator due to Singh et al (2007) for second quantile and found to be dominant over the above said.

Further, by studying the increased level of precision of the four proposed exponential ratio type estimator in Unit –I, Unit-II has been devoted to the estimation of population mean. For this, four exponential ratio type estimator have been suggested while utilizing a stable and completely known auxiliary information available on both the occasions, to estimate population mean in chapter 6. It has been found out the proposed estimators also behave enormously better while estimating population mean. Their dominance has been shown by comparing them with respect to sample mean estimator and general successive sampling estimator due to Jessen (1942).

A multivariate generalization has been illustrated in chapter-7 for the estimator T_{22} while utilizing p- auxiliary information which are stable over two successive occasions and easily available on both the occasions. The multivariate weighted estimator has been shown dominant over two well-known recent estimators, Singh (2005) and Singh and Priyanka (2008a). It has been vindicated that the precision gradually increases as the number of auxiliary information is increased.

In chapter-8, the four exponential ratio type estimators for estimating population mean have been proposed while accompanying a dynamic auxiliary over two successive waves. The proposed estimators have been compared and shown to be better with respect to sample mean estimator and estimator due to Jessen (1942).

The Unit-III has been devoted to the estimation of population mean in the presence of non-response. An attempt has been made for the treatment of non-response in sampling over two successive occasions while using the technique of imputation. In chapter-9, methods of imputations have been proposed while using the above said estimator T_{22} utilizing stable auxiliary information over two successive occasions. The proposed estimator has been classified according to the presence of non-response at only first occasion, presence of non-response at only second and presence of non-response at both the occasions. Percent relative loss has been computed for above three possibilities of nonresponse as compared to the estimators proposed in chapter 6. It has been seen that the amount of loos is not significant in the presence of non-response. Hence the utilization of proposed estimators has been recommended to the survey statisticians under non-response while estimating population mean.

Chapter 10 crusades a multivariate weighted exponential ratio type estimator accompanying several auxiliary information in the presence of non-response for estimating population mean. Percent relative loss has been computed with respect to estimator proposed in chapter 7 when there is no non-response at any occasion and the amount of percent relative loss is found to be minimal in the presence of non-response. Also it is observed that more loss is observed when the non-response occurs at both the occasions.

In chapter 11, the estimator has been proposed under non-response while estimating the population mean of the study character and the auxiliary information tends to be dynamic due a large gap between the two successive occasions. The estimator has been compared to the estimator proposed in chapter 8 when there is no non-response at any occasion and the percent loss turns to be non-significant.

There is less literature available in the field of successive sampling while estimating any population parameter of a sensitive study character on successive occasions. In the available literature, the population parameter of sensitive character has been estimated using certain randomized response technique whose application turns next to impossible in large sample surveys since they may be time consuming and the new age fast life put constraints to the respondents of time. This leads to complete refusal also. So in Unit-IV an alternative approach known as scrambled response technique for estimating population mean of the sensitive character while utilizing a non-sensitive auxiliary information has been illustrated.

The scrambled response technique has been illustrated under two scrambled response models namely additive scrambled response model (ASRM) and multiplicative scrambled response model (MSRM). Various estimators have been proposed under ASRM and MSRM and they are compared with scrambled sample mean estimator under

ASRM and MSRM respectively. A comparison of two scrambled response models suggests that ASRM is plausibly better than MSRM in application over two successive occasions as well as increased precision and total cost of survey dealing a sensitive issue. Also it has been found that the estimators utilizing a non-sensitive auxiliary information (exponential ratio type estimators) out performs Jessen's estimator for scrambled response and scrambled sample mean estimator.

Hence looking at the consistent application of exponential type estimators under diverse situations as surveys troubling non-response, surveys having chances of false response and surveys having different population parameters to be estimated, the proposed exponential ratio type estimators are recommended to the survey statisticians for their practical applications in real time scenario.

2. Future Scope of Study

Survey sampling is vast area to be indulged in to. More and more possibilities are always hidden in the issue to be handled. The present work concerns sampling over two successive occasions using simple random sampling without replacement and considering all assumptions of SRSWOR. These work may be explored under many other sampling schemes like Stratified random sampling, Varying probability sampling, Double sampling, Cluster Sampling, Two stage sampling, or a ramification of any of them. The proposed estimators may also be tested under multiple other imputation techniques when the non-response occurs either considering successive sampling or any of the above said.

A wider range of experimentation includes the estimation of other population parameters using the same estimators and that too may include any sampling scheme.

Another field of possibility includes the testing of proposed estimators using various randomized response technique already available in literature. These estimators may also be worked out with different scrambled response models with a variety of assumption on the distribution of the scrambling variable.

UNIT-VI

REFERENCES

References

Ahmed, M. S., Al-Titi, O., Al-Rawi, Z. and Abu-Dayyeh, W. (2006): Estimation of a population mean using different imputation methods, Statist. Transit. 7, 1247-1264.

Allen, J., Saxena, S., Singh, H.P., Singh, S. and Smarandache, F. (2002): Randomness and optimal estimation in data sampling, Ameri. Res. Press, 26-43.

Arnab, R and Okafor, F.C. (1992): A note on double sampling over two occasions, Pakis. J. statis., 9-18.

Arnab, R. (2011): Alternative estimators for randomized response techniques in multi-character surveys, Comm. Statist.-Theo. Meth., 40, 1839-1848.

Arnab R, Singh S. and North, D. (2012): Use of two decks of cards in randomized response techniques for complex survey designs, Comm. Statist.- Theo. Meth., 41, 3198-3210.

Arnab R, and Singh S. (2013): Estimation of mean of sensitive characteristics for successive sampling, Comm. Statist.-Theo. Meth., 42, 2499-2524.

Artes, E. and Garcia, A. V. (2005): Multivariate indirect methods of estimation in successive sampling. J. Ind. Soci. Agri. Statis. 59(2), 97-103.

Bahl, S. and Tuteja, R.K. (1991): Ratio and Product type exponential estimator, Infor. and Optimi. Sci., 12, I 159-163.

Bandyopadhyay, A. and Singh, G. N. (2014): On the use of two auxiliary variables to improve the precision of estimate in two-occasion successive sampling. Inter. J. Math. Stati. 15(1), 73-88.

Biradar, R. S., Singh, H. P. (2001): Successive sampling using auxiliary information on both occasions. Cal. Statist. Assoc. Bull., 51, 234-251.

Chambers, R. L. and Dunstan, R. (1986): Estimating distribution functions from survey data. Biometrika 73, 597—604.

Chaturvedi, D. K., Tripathi, T. P. (1983): Estimation of population ratio on two occasions using multivariate auxiliary information. J. Ind. Statist. Assoc. 21, 113-120.

Christofides, T. C. (2003): A generalized randomized response technique, Metrika, 57(2), 195-200.

Christofides, T. C. (2005): randomized response in stratified sampling, J. statist. Plann. Infer. 128(1), 303-310.

Cochran, W. G. (1977): Sampling Techniques. New York, John Wiley & Sons.

Das, A. K. (1982): Estimation of population ratio on two occasions. J. Ind. Soc. Agr. Statist. 34, 1-9.

Des, R. (1965). On a method of using multi-auxiliary information in sample surveys. J. Amer. Statis. Asso. 60, 270-277.

Diana, G. and Perri, P. F. (2010). Improved estimators of the population mean for missing data. Comm. Stat.- Theo. Meth. 39, 3245-3251.

Dianna, G. and Perri, P. F. (2011), A class of estimators for quantitative sensitive data, Stat. Pap. 52, 633-650.

Eckler, A. R. (1955): Rotation Sampling, Ann. Math. Statist. 26(4), 664-685.

Eichhorn, B.H. and Hayre, L.S. (1983), Scrambled randomized response method for obtaining sensitive quantitative data. J. Statist. Plann. Infer. 7, 307-316.

Feng, S. and Zou, G. (1997): Sample rotation method with auxiliary variable. Commun. Statist. Theo-Meth. 26(6), 1497-1509.

Gordon, L. (1983): Successive sampling in finite populations, The Ann. Statis. 11(2), 702-706.

Greenberg, B.G., Kubler, R.R., Horvitz, D.G. (1971): Application of RR technique in obtaining quantitative data, J. Amer. Statist. Asso., 66, 243-250.

Gross, S. T. (1980): Median estimation in sample surveys, Proc. Surv. Res.Meth.Sect. Amer. Statist. Assoc., 181-184.

Gupta, P. C. (1979): Sampling on two successive occasions. J. Statist. Res, 13, 7-16.

Gupta, S., Gupta, B. and Singh, S. (2002),:Estimation of sensitivity level of personal interview survey question, J. Statist. Plann. Infer., 100, 239-247.

Gupta, S. and Shabbir, J. (2004), Sensitivity estimation for personal interview survey questions, Statistica, 64, 643-653.

Gupta, S. N., Thorton, B., Shabbir, J. and Singhal, S., (2006), A comparison of multiplicative and additive optional RRT models, J. Statist. Theo. Appl., 5, 226-239.

Gupta, S., Shabbir, J. and Ahmad, S. (2008): Estimation of median in two phase sampling using two auxiliary variables. Comm. in Statis.- Theory & Methods 37(11),1815-1822.

Gupta, S., Sabbir, J. and Sehra, S. (2010), Mean and sensitivity estimation in optional randomized response models, J. Statist. Plann. Infer., 140(10), 2870-2874.

Gupta, S., Shabbir, J., Sousa, R. and Real, P. C. (2012), Estimation of the mean of a sensitive variable in the presence of auxiliary information, Comm. Statist.-Theo. Meth., 41, 2394-2404.

Horvitz, D.G., Shah, B.V. and Simmons, W.R. (1967): The unrelated question randomized response model, Proc. Soci. Statist., Amer. Statist. Asso., 65-72.

Hussain, Z. and Al-Zahrani, B. (2016): Mean and sensitivity estimation of a sensitive variable through additive scrambling. Comm. Stat. – Theo. Meth., 45, 182-193.

Jessen, R. J. (1942): Statistical investigation of a sample survey for obtaining farm facts, Iowa Agricultural Experiment Station Road Bulletin No. 304, Ames, 1-104.

Kalton, G., Kasprzyk, D. and Santos R. (1981) Issues of non-response and imputation of income and program participation, In :Krevoski, D., Platek, R., Rao J. N. K. eds., Current Topic in Survey Sampling, Academics Press, New York, 455-480.

Kalton, G. and Kasprzyk, D. (1982): Imputating for missing survey responses, Proc. Sec. Surv. Res. Meth. Amer. Statist. Assoc. (1982), pp. 22-31.

Kim, J.-M. and Elam, M. E. (2007): A stratified unrelated question randomized response model, Statis. Pap., 48(2), 215-233.

Kuk, A. Y. C. and Mak, T. K. (1989): Median estimation in presence of auxiliary information, J. R. statit. Soc. B 51, 261-269.

Koyuncu, N., Gupta, S. and Sousa, R., (2014): Exponential-type estimators of the mean of a sensitive variable in the presence of a non-sensitive auxiliary information. Comm. Stat.- Sim. Comp., 43, 1583-1594.

Lee, H., Rancourt, E. and Särndall, C. E. (1994): Experiments with variance estimation from survey data with imputed values, J. official Statist. 10, 231-243.

Lee, H., Rancourt, E. and Särndall, C. E. (1995): Variance estimation in the presence of imputed data for generalized estimation system, Proc. Amer. Statist. Assoc., 384-389.

Martinez-Miranda, M. D., Rueda-Garcia, M., Arcos-cebrian, A., Roman-Montoya, Y. and Gonzaez-Aguilera, S. (2005): Quintile estimation under successive sampling, Compu. Statis. 20, 385-399.

Narain, R. D. (1953): On the recurrence formula in sampling on successive occasions, J. Indi. Soci. Agri. Statis. 5, 96-99, 1953.

Olkin, I. (1958): Multivariate ratio estimation for finite populations, Biometrika, 45, 154-165.

Patterson, H. D. (1950): Sampling on successive occasions with partial replacement of units, Jour. Royal Statist. Assoc., Ser. B 12, 241-255.

Pollock, K. H. and Bek, Y. (1976): A comparison of three randomized response models for quantitative data. J. Amm. Sta. Asso., 71, 884-886.

Priyanka, K. and Mittal, R. (2014): Effective rotation patterns for median estimation in successive sampling, Statis. Trans. new seri. Spring 15, 197 - 220.

Priyanka, K., Mittal, R. and Min-Kim, J. (2015): Multivariate Rotation Design for Population Mean in sampling on Successive Occasions. Comm. Statis. Appli. Meth. 22, 445-462.

Priyanka, K. and Mittal, R. (2015a): Estimation of Population Median in Two-Occasion Rotation. J. Stat. App. Prob. Lett. 2, 205-219.

Priyanka, K. and Mittal, R. (2015b): A Class of Estimators for Population Median in two Occasion Rotation Sampling. HJMS, 44, 189-202.

Priyanka, K. and Mittal, R. (2016): Searching Effective Rotation Patterns for Population Median using Exponential Type Estimators in Two-Occasion Rotation Sampling, Comm. Stat.-Theo. Meth.

Rao, J. N. K., Kovar, J. G. and Mantel, H. J. (1990): On estimating distribution functions and quantiles from survey data using auxiliary information. Biometrika, 77(2), 365-375.

Rao, J. N. K., Graham, J. E. (1964): Rotation Design for sampling on repeated occasions. J. Amer. Statist. Assoc, 59, 492-509.

Rao, C. R. (2006): Linear Statistical Inference and its Application, 2nd ed., John Wiley and Sons (Asia) Pvt. Ltd, Singapore.

Reddy, V. N. (1978): A study on the use of prior knowledge on certain population parameters in estimation. Sankhya, C 40, 29-37.

Rubin, R. B. (1976): Inference and missing data, Biometrika. 63, 581-592.

Rueda, M. D. M., Arcos, A. and Artes, E. (1998): Quantile interval estimation in finite population using a multivariate ratio estimator. Metrika 47: 203-213.

Rueda, M. D. M., and Munoz, J. F. (2008): Successive sampling to estimate quantiles with P-Auxiliary Variables, Quality and Quantity42, 427-443.

Saha, A. (2007), A simple randomized response technique in complex surveys. Metron, LXV, 59-66.

Sande, I. G. (1979): A personal view of hot-deck imputation procedure, Surv. Methodo. Statist. Can. 5, 238-258.

Sedransk, J. and Meyer, J. (1978): Confidence intervals for n quantiles of a finite population, simple random and stratified simple random sampling, J. R. Statist. Soc. B40, 239-252.

Sen, A. R. (1971): Successive sampling with two auxiliary variables. Sankhya, B 33, 371-378.

Sen, A. R. (1972): Successive sampling with p $(p \geq 1)$ auxiliary variables. Ann. Math. Statist., 43, 2031-2034.

Sen, A. R. (1973): Theory and application of sampling on repeated occasions with several auxiliary variables. Biometrics, 29, 381-385.

Silverman, B. W (1986). Density Estimation for Statistics and Data Analysis, (Chapman and Hall, London,

Sihm, J.-S. and Gupta, S. (2015), A two-stage binary optional randomized response model, Comm. Statist.-Sim. Comp., 44, 2278-2296.

Singh, G. N. and Singh, V. K. (2001): On the use of auxiliary information in successive sampling, Jour. Ind. Soc. Agri. Statist. 54(1), 1-12.

Singh, G. N. (2003): Estimation of population mean using auxiliary information on recent occasion in h occasions successive sampling. Statist. Transi., 6, 523-532.

Singh, G. N. (2005): On the use of chain-type ratio estimator in successive sampling. Statist. Transi., 7, 21-26.

Singh, G. N., Priyanka, K. (2006): On the use of chain-type ratio to difference estimator in successive sampling. Inter. Jour. Appli. Math. Stat., 5(S06), 41-49.

Singh, G. N., Priyanka, K. (2007a): On the use of auxiliary information in Search of good rotation patterns on successive occasions. Bull. of Stati. and Eco., 1(A 07), 42-60.

Singh, G. N., Priyanka, K. (2007b): Effect of non-response on current occasion in search of good rotation patterns on successive occasions, Statist. Transit. New Ser. 8, 273-292.

Singh, G. N. and Karna, J. P. (2009): Search of efficient rotation patterns in presence of auxiliary information in successive sampling over two occasions. Statist. Transi., 10(1), 59-73.

Singh, G. N. and Priyanka, K. (2008a): Search of good rotation patterns to improve the precision of estimates at current occasion, Comm. Statis.-Theo. Meth. 37(3), 337-348.

Singh, G. N. and Priyanka, K. (2008 b): On the use of guess value for the estimation of population median on current occasion in two occasions rotation patterns, Statis. Transi. 9, 215 -232.

Singh, G. N. and S. Prasad (2010): Some estimators of population mean in two-occasion rotation patterns. Assoc. of Adva. Mode. Sim.Tech. Enter., 12(1), 25-44.

Singh, G. N., Prasad, S, S. and Majhi, D. (2012): Best Linear Unbiased Estimators of Population Variance in successive Sampling, Mode. Assi. Statis. App. 7, 169-178.

Singh, G. N., Priyanka, K., Prasad, S., Singh, S. and Kim, J. M. (2013a): A class of estimators for estimating the population variance in two occasion rotation patterns. Comm. Statis. App. Meth., 20(4), 247-257.

Singh, G. N., Majhi, D., Prasad S. and Homa, F. (2013b): Assessment of Non-response under ratio method of imputation in two occasion successive sampling, J. statist. theo. Appl. 12, 403-418.

Singh, H. P., Tailor, R., Singh, S. and Jong-Min Kim (2007): Quintile Estimation in Successive Sampling, Journal of the Korean Statistical society 36(4), 543-556.

Singh, H. P., Kumar, S. and Bhougal, S. (2011). Multivariate ratio estimation in presence of non-response in successive sampling. Journal of Statistical Theory and Practice 5(4): 591-611.

Singh, H. P. and Solanki, R. S. (2013): Some Classes of estimators for population median using auxiliary information. Comm. Statis.-Theo. Meth., 42(23), 4222-4238.

Singh R. and Singh, N. (1991): Imputation methods in two-dimensional survey. Rece. Advan. Agri. Statis. Res., Wiley Eastern Ltd, New Delhi.

Singh, S. and Horn, S. (2002): Compromised imputation in survey sampling, Metrika 51, 266-276.

Singh, S. (2003): Advanced Sampling Theory with Applications, How Michael 'selected' Amy. Vol. 1 and 2, (Kluwer Academic Publishers, The Netherlands, 1-1247.

Singh, S. (2009): A new method of imputation in survey Sampling, Statist. 43, 499-511.

Singh, S. and Sedory, S. A. (2012): A true simulation study of three estimators at equal protection of respondents in randomized response sampling, Statistica Neer., 75(4), 442-451.

Singh, V.K. and Shukla, D. (1987): One parameter family of factor-type ratio estimators, Metron45, 1-2, 30, 273-283.

Singh, V. K., Singh, G. N., Shukla, D. (1991): An efficient family of ratio cum difference type estimators in successive sampling over two occasions. Jour. Sci. Res., 41 C, 149-159.

Smith, P. and Sedransk, J. (1983): Lower bounds for confidence coefficients for confidence intervals for finite population quantiles, Commun. Statist. – Theory Meth. 12, 1329-1344.

Sukhatme, P.V., Sukhatme B. V., Sukhatme, S., Asok, C. (1984): Sampling Theory of Surveys with applications. Ames, Iowa State University Press and New Delhi India, Ind. Soci. Agri. Statis.

Warner, S. L. (1965): Randomized response: a survey technique for eliminating evasive answer bias, J. Amer. Statist. Asso. 60, 63-69.

Warner, S. L. (1971): The linear randomized response model, J. Amer. Statist. Asso. 66, 884-888.

Wu, J.-W., Tian, G.-L.and Tang, M.-L. (2008): Two new models for survey sampling with sensitive characteristics: Design and Analysis, Metrika, 67, 251-263.

Yan, Z., Wang, J. and Lai, J. (2009): An efficiency and protection degree-based comparison among the quantitative randomized response strategies, Comm. Statist.- Theo. Meth., 38, 400-408.

Yu, B., Jin, Z., Tian, J. and Gao, G. (2015): Estimation of sensitive proportion by randomized response data in successive sampling, Compu. Mathemat. Meth. Med. DOI: 10.1155/2015/172918.

List of Publications

1. Priyanka, K. and Mittal, R. (2014): Effective Rotation Patterns for Median Estimation in Successive Sampling. ***Statistics in Transition-new series***, Vol. 15, No. 2, 197-220.

2. Priyanka, K. and Mittal, R. (2015): A Class of Estimators for Population Median in Two Occasion Rotation Sampling. ***HJMS***, Vol. 44, No. 1, 189 – 202.

3. Priyanka, K. and Mittal, R. (2015): Estimation of Population Median in Two-Occasion Rotation Sampling. ***Journal of Statistics Applications & Probability Letters***, Vol. 2, No. 3, 205-219.

4. Priyanka, K. Mittal, R. Kim, J. M. (2015): Multivariate Rotation Design for Population Mean in Sampling on Successive Occasions. ***Communications for Statistical Applications and Methods***, Vol. 22, No. 5, 445–462.

5. Priyanka, K. and Mittal, R. (2016): Longitudinal Analysis of Population Mean on Successive Occasions. ***International Journal of Mathematics and Statistics,*** Vol. 17, No. 2, 47-63.

6. Priyanka, K. and Mittal, R. (2016): Searching Effective Rotation Patterns for Population Median using Exponential Type Estimators in Two-Occasion Rotation Sampling. ***Communication in Statistics (Theory and Methods),*** DOI: 10.1080/03610926.2014.944661.

7. Priyanka, K. and Mittal, R. (2016): Intercession of Non-Response through Imputation in Longitudinal Surveys for Population Mean. ***Journal of Modern Applied Statistical Methods,*** (Accepted For Publication).

8. Priyanka, K. and Mittal, R. (2016): Cogitation of Incompleteness in the midst of Imputation in Longitudinal Surveys for Population Mean. Proceedings of National Conference on ***RSCTA***, held in Ramanujan College, University of Delhi on 11-12th Mach, 2016. (Accepted For Publication)

9. Priyanka, K. and Mittal, R. (2016): Multivariate Analysis of Longitudinal Surveys for Population Median. ***Journal of Applied Statistics***, (Communicated).

10. Priyanka, K. and Mittal, R. (2016): A fresh approach for Intercession of Non-Response in Multivariate Longitudinal Designs. ***Communication in Statistics (Theory and Methods),*** (Accepted For Publication).

11. Priyanka, K. and Mittal, R. (2016): New Approaches using Exponential Type Estimator with Cost Modelling for Population Mean on Successive Waves. ***Statistics in Transition-new series,*** (Accepted For Publication).

12. Priyanka, K. and Mittal, R. (2016): Scrambled Response Techniques in Two Wave Rotation Sampling for Estimating Population Mean of Sensitive Characteristics and its Applications. ***Metrika,*** (Communicated).