Some Contribution to Robust regression Techniques for the Estimation of Finite Population Parameter using Auxiliary Information

By:

Nasir Ali
Reg. No. 1-FBAS/PHDST/F16

Department of Mathematics and Statistics
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad
Pakistan

2022

Mathematical models
Parameter estimation

Some Contribution to Robust regression Techniques for the Estimation of Finite Population Parameter using Auxiliary Information

By:

Nasir Ali
Reg. No. 1-FBAS/PHDST/F16

Supervised By:

Dr. Ishfaq Ahmad

Co-Supervised By:

Dr. Muhammad Hanif

Department of Mathematics and Statistics
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad
Pakistan

2022

Some Contribution to Robust regression Techniques for the Estimation of Finite Population Parameter using Auxiliary Information

By:

Nasir Ali

Reg. No. 1-FBAS/PHDST/F16

A Thesis

Submitted in the Partial Fulfillment

of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

IN

STATISTICS

Supervised By:

Dr. Ishfaq Ahmad

Co-Supervised By:

Dr. Muhammad Hanif

Department of Mathematics and Statistics
Faculty of Basic and Applied Sciences
International Islamic University, Islamabad
Pakistan

2022

<u>Author's Declaration</u>

I, **Nasir Ali** Reg. No. **1-FBAS/PHDST/F16** hereby state that my Ph.D. thesis titled: **Some Contribution to Robust regression techniques for the Estimation of Finite Population Parameter using Auxiliary Information** is my own work and has not been submitted previously by me for taking any degree from this university, **International Islamic University, Sector H-10, Islamabad, Pakistan** or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after my Graduation the university has the right to withdraw my Ph.D. degree.

Name of Student: ***(Nasir Ali)***
Reg. No. **1-FBAS/PHDST/F16**
Dated: **19/10/2022**

Plagiarism Undertaking

I solemnly declare that research work presented in the thesis titled: **Some Contribution to Robust regression techniques for the Estimation of Finite Population Parameter using Auxiliary Information** is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and University, **International Islamic University, Sector H-10, Islamabad, Pakistan** towards plagiarism. Therefore, I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of Ph.D. degree, the university reserves the rights to withdraw/revoke my Ph.D. degree and that HEC and the University has the right to publish my name on the HEC/University Website on which names of students are placed who submitted plagiarized thesis.

Student/Author Signature: ______________

Name: **(Nasir Ali)**

Certificate of Approval

This is to certify that the research work presented in this thesis, entitled: **Some Contribution to Robust regression techniques for the Estimation of Finite Population Parameter using Auxiliary Information** was conducted by **Mr. Nasir Ali, Reg. No. 1-FBAS/PHDST/F16** under the supervision of **Dr. Ishfaq Ahmad** no part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the **Department of Mathematics & Statistics, FBAS, IIU, Islamabad** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Statistics, Department of Mathematics & Statistics, Faculty of Basic & Applied Science, International Islamic University, Sector H-10, Islamabad, Pakistan.**

Student Name: Nasir Ali ***Signature:*** ____________

Examination Committee:

a) **External Examiner 1:**
Name/Designation/Office Address **Signature:** ____________
Dr. Faqir Muhammad
Professor of Statistics, H. No. 186, St. No. 4, G-15/1, Islamabad.

b) **External Examiner 2:**
Name/Designation/Office Address) **Signature:** ____________
Dr. Tahir Mehmood
Professor of Statistics, School of Natural Sciences, NUST, Islamabad, Pakistan.

c) **Internal Examiner:**
Name/Designation/Office Address) **Signature:** ____________
Dr. Muhammad Akbar
Assistant Professor

Supervisor Name:
Dr. Ishfaq Ahmad **Signature:** ____________

Co-Supervisor Name:
Dr. Muhammad Hanif **Signature:** ____________

Name of HOD:
Prof. Dr. Nasir Ali **Signature:** ____________

Name of Dean:
Prof. Dr. Muhammad Irfan Khan **Signature:** ____________

DEDICATION

This work is completely dedicated to

Allah Almighty and Prophet (PBUH),

My lovely Parents for their prayers,

Respected Teachers for their inspiration,

and

Caring Siblings for their motivation and support.

Declaration

I hereby declare and affirm that this research work neither as a whole nor as a part has been copied out from any source. It is further declared that I have prepared this dissertation entirely on the basis of my personal efforts made under the sincere guidance of my worthy supervisor and co-supervisor. If any part of this thesis is proven to be copied out or found to be a reproduction of sum other, I shall stand by the consequences.

Moreover, no portion of the work presented in this thesis has been submitted in support of an application for other degree or qualification in this or any other university or institute of learning.

Name and Signature of the student________________

Nasir Ali

PhD (Statistics)

Reg. # 01/FBAS/PHDST/F16

FORWARDING SHEET BY RESEARCH SUPERVISOR

The thesis entitled **"Some Contributions to Robust Regression Techniques for the Estimation of Finite Population Parameter using Auxiliary Information"** submitted by **Nasir Ali** (Registration # 1-FBAS/PHDST/F16) in partial fulfillment of PhD degree in Statistics has been completed under my guidance and supervision. I am satisfied with the quality of his research work and allow him to submit this thesis for further process to graduate with Doctor of Philosophy degree from Department of Mathematics and Statistics, as per IIU Islamabad rules and regulations.

Dated: ____________________

____

Dr. Ishfaq Ahmad
Associate Professor
Department of Mathematics & Statistics
International Islamic University,
Islamabad.

Acknowledgements

First of all, I pay special thanks to Almighty Allah, who assisted me in all walks of life and also assisted and guided me throughout my PhD work and in the accomplishment of this thesis too. Many Salat-o-Salam on our Holy Prophet Hazrat Muhammad (PBUH) who is a source of knowledge and blessings for the entire creations.

I would like to acknowledge the worth mentioning supervision of **Dr. Ishfaq Ahmad** and Co-supervision of **Dr. Muhammad Hanif** who guided me and supported me during my whole research work. Moreover, their supervision enabled me to develop an understanding of the field. Without their sincere efforts I was unable to complete this hard task of my life. Really I am thankful to **Dr. Ishfaq Ahmad** and **Dr. Muhammad Hanif** for their inspiration and encouragement in every field of life especially in education and teaching. **Almighty Allah** may bless them with long life, health, happiness and knowledge. Moreover, I would like to express my sincere thanks to all faculty members of Department of Mathematics and Statistics IIU Islamabad especially to **Dr. Muhammad Akbar Awan.** I would also like to thank all other faculty member of my university for their sincere appreciation, comments and suggestions. I express my thanks to all staff of Mathematics and Statistics Department, IIU, for their various services. I am thankful to my all friends and would like to mention the names of some friends, Mr. Usman Shahzad, Dr. Abid Hussain Dr. Ehtasham ul Haq, Dr. Muhammad Shafeeq ul Rehman Khan, Dr. Muhammad Amjad, Mr. Atta Muhammad Asif, Mr. Muhammad Asim and Others, for their moral support. They always remain ready to assist me when I was needed. My beloved sons Mr. Ahmad Hassan, Mr. Hammad Hassan and my family members.

(Nasir Ali)

Preface

Robust regression provides an alternative to least squares regression that works with less restrictive assumptions. Specifically, it provides much better regression coefficient estimates when outliers are present in the data. Outliers violate the assumption of normality of residuals in least squares regression. They tend to misrepresent the least squares coefficients by having more influence than they deserve. Typically, you would expect that the weight attached to each observation would be about $1/n$ in a dataset with n observations. However, outlying observations may receive a weight of 10, 20, or even 50 %. This leads to serious distortions in the estimated coefficients. Because of this distortion, these outliers are difficult to identify since their residuals are much smaller than they should be. When only one or two independent variables are used, these outlying points may be visually detected in various scatter plots. However, the complexity added by additional independent variables often hides the outliers from view in scatter plots. Robust regression down-weights the influence of outliers. This makes residuals of outlying observations larger and easier to spot. Robust regression is an iterative procedure that seeks to identify outliers and minimize their impact on the coefficient estimates. Randomized response technique (RRT) allows respondents to mark their actual response by giving a scrambled response which makes the researcher at later to unscramble at an aggregate level but not at an individual level. In this thesis, we focus on generalized quantitative scrambled response methods and some generalized ratio-type estimators, which have been proposed for a finite population mean of a sensitive variable based on RRT by using sensitive auxiliary variable. These estimators have been proposed under different sampling designs, such as simple random sampling, stratified random sampling, systematic random sampling and two phase sampling.

Chapter 1 is the introductory chapter in which detailed explanations of robust regression tools is provided. Furthermore, we also discussed the concept of simple random sampling, stratified random sampling, systematic sampling and two phase sampling.

Chapter 2 presents the theoretical study of proposed robust regression estimators for sensitive study variable. The proposed robust regression estimators require less supplementary information as compared to robust ratio type estimators. The mean square error (MSE) equation for the estimators are also obtained. The superiority of the proposed class has been evaluated by simulation study utilizing both theoretically and empirically. One real and one artificial population is utilized for this purpose. The proposed class is recommended for survey practitioners as it strength expand the probabilities of getting increasingly efficient results of unknown population mean of sensitive study variable.

The results of this chapter are published in **Communications in Statistics-Theory and Methods. DOI: 10.1080/0 3610926.2019.1645857.**

Chapter 3 proposes two classes of estimators whenever data is contaminated with outliers under systematic random sampling scheme. At first we develop ratio type estimators based on robust regression tools. Secondly, we develop regression type estimator based on regression tools. It is seen that proposed estimator perform better as compare to ratio estimators. The results of this chapter are published in **Fresenius Environmental Bulletin (FEB).**

Chapter 4 Delivers a class of robust- regression type estimators for mean estimation under simple random sampling and two- stage sampling schemes when quantitative supplementary information is available. We also find MSE and minimum MSE expressions of the proposed class. The proposed class of estimators has been compared with prevailing ones. Based on empirical and theoretical percentage relative efficiency (PRE) results, it is clear that the proposed class perform better as compared to traditional mean estimator, traditional regression estimator.

The results of this chapter are published in, **The Electronic journal of Applied Statistical Analysis (EJASA).**

Chapter 5 is based on the information related to two auxiliary variables. A novel class of robust regression estimators for mean estimation is suggested. The proposed estimators are an extension of Abid et al. (2018) work and rely on robust regression tools. Three real life data sets in the presence of outliers have been considered in the numerical illustration. It is observed that the values of the PRE of the proposed estimators are much higher than those for the existing estimators.

The results of this chapter are submitted for publication in Journal titled **PLOS ONE.**

Chapter 6 focuses on modifying the combination of combined ratio and product estimators to estimate the population mean under stratified sampling scheme. For this purpose, different new estimators are proposed and compared with the existing estimators in stratified random sampling. Numerical results of the suggested and existing estimators are based on the MSE. All results indicate that the mean square error of the suggested modified estimators is lower than the MSE of existing estimators. Therefore, the suggested modified estimators are the better and more efficient estimators as compared to the existing ones. The results of this chapter are published in **The Journal of Science and Arts (JOSA).**

Table of Contents

Lists of Tables

Lists of Figures

Research Profile

We are published three articles and one is accepted and one is submitted from this thesis. The details of these articles are as follows.

Ali, N., Ahmad, I., Hanif, M., & Shahzad, U. (2021). Robust-regression-type estimators for improving mean estimation of sensitive variables by using auxiliary information, ***Communications in Statistics - Theory and Methods***, Vol. 50(4), 979-992.

Ali, N., Ahmad, I., Hanif, M., Shahzad, U., & Masood, M. (2021). Robust estimators for mean estimation in systematic sampling with the numerical application in Forestry. ***Fresenius Environmental Bulletin***, Vol. 30(6), 5635-5644.

Ali, N., Ahmad, I., Hanif, M., & Usman, U. (2022). Proposed a class of robust regression type estimators for mean estimation under simple random sampling and two stage sampling scheme. ***The Electronic Journal of Applied Statistical Analysis (EJASA).*** Vol. 15(2), 357-373.

Ali, N., Ahmad, I., Shahzad, U. & Al-Noor, H. N. (2022). Some improved estimators for the mean estimation under stratified sampling by using transformations. ***Journal of Science and Arts (JOSA).*** Vol. 22(2), 265-288.

Ali, N., Ahmad, I., Almanjahie, I. M., & Shahzad. U. (2022). Population mean estimation with two auxiliary variables using robust regression methods. ***PLOS ONE***, journal. (Submitted).

Nomenclature

N	Population size
n	Sample size
N_h	Population of h^{th} stratum
n_h	Sample size of h^{th} stratum
L	Number of strata
X	Auxiliary variable
$f = \frac{1}{N}$	Sampling fraction
$\bar{x}, \bar{y}$	Sample means
S_x, S_y	Population standard deviations
x, y	Sample totals
C_x, C_y	Coefficient of variation
$B(.)$	Bias of the estimator
$MSE(.)$	Mean squared error of the estimator
$\bar{Y}_i$	Existing estimators
$\bar{Y}_j$	Proposed estimators
M_d	Median of X
$MR = \frac{X_{(1)}+X_{(N)}}{2}$	Mid- range
$TM = \frac{Q_1+2Q_2+Q_3}{4}$	Tri- mean
ρ_{zxh}	Coefficient of correlation between scrambled and auxiliary variables h^{th} stratum
$\beta_1(x)$	Coefficient of skewness of auxiliary variable
$\beta_2(x)$	Coefficient of kurtosis of auxiliary variable
$\hat{\bar{Y}}_i$	Existing modified ratio estimator of $\bar{Y}$
$\hat{\bar{Y}}_{pj}$	Proposed modified ratio estimator of $\bar{Y}$
Subscripts	
i	For existing estimators
j	For proposed estimators

List of Abbreviations

EULFS:	European Union labor force survey
PSS:	Profiling segment survey
NCS:	National Co-morbidity survey
OLS:	Ordinary least square
SDB:	Social desirability bias
BPL:	Bogus pipeline
RRT:	Randomized response techniques
SRSWR:	Simple random sampling with replacement
SRSWOR:	Simple random sampling without replacement
LAD:	Least absolute deviations
LTS:	Least trimmed square
LMS:	Least median of square
REWLSE:	Robust and efficient weighted least squares estimator
MSE:	Mean square error
PRE:	Percentage relative efficiency
H.M:	Huber - M estimate
HPL:	Hample - M estimate
TKY:	Tukey - M estimate
H.MM:	Huber – MM estimate
POP:	Population

Chapter 1

Introduction

1.1 Background

We are living in an "information era". By using this term, we need not only the sufficient information but also speedy work of data collection. So, the question is this, if we need information on urgent basis, then how it possible that our collected information is accurate? Also that gathered information is also useful for the future decision-making processes. We are facing these types of challenges on daily basis, e.g. government tasks, business decisions, clinical investigations, etc. This is not surprising for us that how minimal information, can be useful for these tasks. For all these required information, sampling is the core field in research literature. For instance, the European Union Labor Force Survey (EULFS) provides the labor strength on quarterly basis, which conducted by Euro stat, to understand the market share behavior, the Profiling Segment Survey (PSS), provides us very relevant information. Also, to assess the level of anxiety, the National Co-morbidity Survey (NCS) gives very useful information. Finally the main objective of the sampling is to estimate the prevalence features of the population parameters.

Most of the time information about a study variable is difficult to collect but it can be with the help of another correlated variable (supplementary information). In statistical language, the additional or extra information, which is used for more efficient results of the study variable, usually called auxiliary variable. The early evidence to use an auxiliary variable as a helper variable, for more information, someone can consult to Lohr (1999), where he discussed how Laplace used auxiliary variable to estimate the total population of France in 18th century. Specifically, Laplace stated that how one can determine the total population of great empire by just maintaining the birth record.

Wright (1928) provided us another interesting example that how can we used the auxiliary variable to estimate the import tariff of butter with vegetable oils, whereas dairy area used as auxiliary information in United states. Some time, we have very low information about study variable and then in this situation, King et al. (2007) suggested that how an auxiliary information can be helpful in the analysis of aggregate data. In survey sampling, the mean estimation is the prime concerns and this can be improved by using auxiliary information's, see, for example, Koyuncu and Kadilar (2009); Koyuncu (2012); Shahzad et al. (2019) and Hanif and Shahzad (2019).

In survey sampling, the outliers are frequently occurred and create problem to estimate the population parameters, especially in highly skewed economic populations. For simplicity, we can define the outliers as those sampling units which have large deviations from their respective expectations. If there are outliers in the data, then they can be misleading from the true estimates. For example, one or more sample observation can be contributed a large proportion to estimate total population. The presence of outlier even mild outlier can be de- track the original population total estimates. Hence for ultimate estimates, someone can use some robust based approaches to control the estimates which effected due to outliers. In classical multiple regression analysis, the ordinary least square (OLS) estimates are best one basic assumption are fulfilled. On the other hand, when one or more assumptions are not fulfilled then OLS estimates are not usually accurate. Especially, the normality assumption can be violates in the presence of outliers. In general, the results are affected when outliers involved in the study variable or in the auxiliary variables. In the present study, we will also discuss that how handle the outliers throughout the estimation of mean.

1.2 Sensitive variables

In simple words, the sensitive survey is the survey where the respondent feels unsecure to provide his/her information. This types of survey, usually comes across the social, personal and health related phenomena. For example, if we put a question to a student about his/her marks based on cheating? Similarly, survey related to drug uses, abortions and assaulting someone are also sensitive in nature.

There are survey studies in literature that response rate is between 40 to 50 percent can be considered as an excellent output. Someone can consult to some of the related studies, such as, Fan and Yan (2010) and Miller and Dillman (2011). But if few or more sensitive questions are the part of the survey questions, then it is really difficult to get responses from the respondent.

In literature, there are a lot of approaches have been used to deal with the problem of non-response or false-response. Here, researchers are just happy when they have a non-response than a false-response. Reynolds (1982) suggested 13-points based SDB scale to quantify the strength about the response by respondent. This score is very much important to use for group comparisons. One of the well-known approach is also used with the help of psychologists *i.e* Bogus pipeline (BPL). Sigall (1971) reviewed about BPL, *i.e.* a fake lie detector test based on machine, where respondent's figure touches on the machine to get information from respondent. This fear complex may be the source to get accurate information.

Warner (1965) suggested a most scientific approach based on probabilistic phenomena. This approach is called Randomized Response Techniques (RRT) and we will discuss it with details in next section. In simple words, these models give enough privacy to the respondent to overcome the false response rate. On the other hand, a study about measurement error is still important in the models of RRT, where we need more accurate confidence intervals for unknown parameters like population mean and proportion etc.

1.2.1 Randomized response techniques

Randomized response techniques (RRT) build the respondent confidence about his/her privacy and allow an alternative way to show his/her responses through random devices. The basic purpose of RRT is to mask the respondent with responses.

For personal or sensitive questions, it is difficult or even impossible to take response from the respondent. For this type of survey, some time, and the respondent do not feel to response or if by force take his/her response, then proportion of false responses may be increases. Initially, Warner (1965) proposed this RRT in a very simple but in meaningful way and after that there a lot variants in literature. Some of them are: Greenberg et al. (1971); Gupta et al. (2006, 2010) and Huang et al. (2010), etc. Basically, the RRT model was based on binary response phenomena. It was done by directly asking a question to a randomly selected person. Few respondents are asked 'did you put on a wrong income tax return in previous year?' In this scenario, the researcher does not judge that which reply come from which respondent, because of scrambled or randomized base responses. For compiling, all these responses can be unscrambled as a whole but not at individual level. Hence, one can easily estimate that how much proportion of the tax payers involved with accurate taxes during the last year without knowing their identities.

In literature, there are many versions of RRT but we can easily differentiate it into two major branches, such as, quantitative and qualitative. In qualitative response, we need to collect the information based on binary responses or simply estimate the proportion for some specific behaviors. For example, we want to estimate the proportion of the people, who drink coffee today. For this, we put two questions to the respondent based on RRT phenomena.

The quantitative response models are being used to tackle the expected value for some population's behavior. They are mostly divided into two major Classes: additive and multiplicative models. For example, the standard deck of cards can be used to estimate "daily consumption for average number of cups of the coffee". For this response the respondent pick a card randomly and provide his/her response by using the sum of card value based on additive models and use product of card value for his/her response based on multiplicative models. In this research, we only focus quantitative response based randomized response models. For how can use multiplicative scrambling models, someone can consult to Eichhorn and Hayre (1983).

1.3 Simple random sampling

The simplest design in probability sampling designs is simple random sampling where all units have equal probability for selection. In this procedure, the sample may be drawn in two different ways: simple random sampling with replacement (SRSWR) and simple random sampling without replacement (SRSWOR). In SRSWR, a sample with unis of n be chosen from the target population one by one and the selected unit is replaced before the next trial. In this situation, the probability of drawing any unit in any draw is $\frac{1}{N}$ and for a sample of size n will be $\frac{1}{N^n}$. On the other hand, for SRSWOR, the unit is not being replaced for the next trial and probabilities are $\frac{1}{N}$, $\frac{1}{N-1}$ and $\frac{1}{(N-r-1)}$ for the first, second and r^{th} draws, respectively etc. Hence, $\frac{1}{N}$ is the probability to draw a specified unit at any draw $\frac{n}{N}$ is included in the sample. We will focus on SRSWOR throughout the thesis.

1.3.1 Non-sensitive variable

In a finite population with size N, we draw a sample of size n by using SRSWOR. Let (y_i, x_i) be the selected values from a bi-variate population (Y_i, X_i).

Let S_Y^2 and S_X^2 be the variances, C_Y and C_X as coefficient of variation, and ρ_{YX} be the coefficient of correlation for the study variable (Y) and the auxiliary variable (X).

1.3.2 Study variable as sensitive and auxiliary variable as non-sensitive

Suppose, we have a bivariate population where study variable Y is sensitive and auxiliary variable X is non-sensitive variable with a positive correlation. Let we have another scrambling variable S, which is also independent with our study and auxiliary variables. The respondent is asked to reply his/her response through $Z = Y + S$ along a true response for variable X, see, for example, Gupta et al. (2002).

For this, we can choose a random sample with size n by using SRSWOR from a population with size N. The observed values are in the form of (y_i, x_i, z_i) whereas, (Y_i, X_i, Z_i) are the true values for the study, auxiliary and scrambled variables. The population variance of scrambled variable is denoted by S_z^2 along coefficient of variation with C_z.

1.4 Stratified random sampling

The main objective in the problem of estimation is to choose an accurate or reliable estimator for the population parameter, which is used to keep all the desired features of the population as well. When all the units in the population are homogenous with respect to a specific characteristic, then simple random sampling (SRS) yielding a good estimate to estimate the population, etc. In this situation, the sample by using simple random sampling is free of non-sampling error. On the other hand, when units are not homogeneous under a specific study, then SRS gives biased results due to highly non-sampling error. In this situation, an alternative sampling design, i.e. stratified random sampling, can be used to avoid the above mentioned review problems due to SRS. The stratified random sampling is used in the following way:

- First a heterogeneous population is divided into subgroups based on the homogenous characteristics. The groups are called strata, where the term stratum is used for a single group.
- In simple words, the units within group are homogenous and heterogeneous when comparing to other group's units.
- Each single stratum is used as a separate population and sample can be drawn by using SRS approach.

The following are the notations, which are used throughout the thesis for stratified random sampling.

N: Population size

L : No. of strata

N_h : Population of h^{th} stratum, where, $h = 1,2,\dots,L$.

n : Sample size

n_h : Sample size of h^{th} stratum, where, $n = \sum_{h=1}^{L} n_h$.

Advantages of stratified sampling:

1. This design is useful when population is heterogeneous.
2. Stratified sampling is not only helpful to estimate the population parameters but also estimates the subpopulations.
3. Most of the cases, the sampling frame is different for the whole population but after dividing into subpopulation, then it may be available.
4. For a large population, it is not an easy task to handle it, but after splitting the population, it can be convenient to handle it.

1.4.1 Non-Sensitive variables

Consider again a population of size N and divided this into L strata, where, $\sum_{h=1}^{L} N_h = N$. Now, a simple random sample of size n_h from stratum h and $\sum_{h=1}^{L} n_h = n$. Let (y_{hi}, x_{hi}) be the selected pair of values on the behalf of true values of (Y_{hi}, X_{hi}), where Y and X are the study and auxiliary variables, respectively from the i^{th} unit of the selected h^{th} stratum.

Also, S_{Yh}^2, C_{Yh} and S_{Xh}^2, C_{Xh} are the variances and coefficient of variation of study and auxiliary variables, respectively and ρ_{YXh} be the correlation coefficient in h^{th} stratum.

1.4.2 Study variable as sensitive and auxiliary variable as non-sensitive

If study variable, Y is sensitive but the auxiliary variable, X is non-sensitive and has strong correlation between them. For RRT, we assume a scrambling random variable S with known distribution but zero mean. The respondent use an additive scrambled response to report his/her response for study variable, as $Z = Y + S$, but on the other hand, his/her true response about auxiliary variables is required. Here, we suppose (y_{hi}, x_{hi}, z_{hi}) are the observed values and (Y_{hi}, X_{hi}, Z_{hi}) are the accurate values for study, auxiliary and scrambled variables, respectively, associated with i^{th} unit of the h^{th} stratum.

S^2_{Zh} and S^2_{Xh} are variances, C_{Zh} and C_{Xh} are coefficients of variations with ρ_{ZXh} is denoted as the coefficient of correlation between scrambled and auxiliary variables in h^{th} stratum.

1.5 Systematic random sampling

Systematic sampling is also a type of random sampling. In this approaches, first select a random unit of the population then after a fixed or periodic interval, the second unit is selected and so on to complete the required sample size, i.e. n. It can be used when a low risk of data manipulation. It can also be preferred than SRS when more study area needed to be covered.

In survey, the systematic sampling is a statistical procedure to select the units from an ordered sampling frame. The one of the most common method in this approach is equiprobability approach. In this method, progression through the list is used circularly. This approach is started with a randomly picked element and then after each K^{th}

element, select the every next unit as $K = \frac{N}{n}$ with K as sampling interval, N as population size and n as sample size.

Using this procedure each element in the population has a known and equal probability of selection. Through this approach, the systematic sampling works exactly similar to SRS. But actually this is not happened because not the each possible outcome has an equal probability to of being chosen.

Logically, the systematic random sampling can only be applied when the target population is homogeneous. Before using this, the researcher should aware about any pattern because, it happens then he/she should it will not possible to take desired results.

1.6 Two-Phase Sampling

In ratio and regression estimators, someone use the knowledge of auxiliary variable to estimate the required parameters for the study variable. It is not possible to collect the information about auxiliary variable, then one of the following two options can be adopted. One of them is very simple and straight forward select a sample only on the behalf of study variable and its estimator is used for the population parameter. The other one can be used when someone has an enough budget to collect the information about auxiliary variable on large scale and find a good estimator for the auxiliary variable.

The second approach is appropriate when the collected information is not in tabulated but in file cards only. After a large preliminary sample size n, choose a subsample of size n from this selected sample to collect the information for study variable. After this, these two estimates are used to find the population parameter for study variable. This whole procedure is called double sampling or two-phase sampling. It is useful when cheap and quick results recorded but a high correlation required between study and auxiliary variables. Because of a large sample is required preliminary, the two-phase sampling is a cost-effective design.

1.7 The robust regression

In general the ordinary least square (OLS) gives optimal estimates in regression estimation when its basic assumptions are fulfilled. On the other hand, when one or more assumptions are invalid, then OLS gives poor estimates. To diagnose the breakdown in assumptions, someone can plot its residuals. The residual diagnostics are sometime difficult to capture as well as time consuming procedure. The alternative approach which is less restriction about OLS assumptions is robust regression method. This approach is provided a better bit in most of the cases when data contain outliers' even mild cases. One of the simplest approaches to estimate the parameters through robust methods is the least absolute deviations (LAD). The LAD is less sensitive in the presence of outliers, but they can put a significant impact in the model. It may lead us to search more useful robust approaches.

In the mid of 1960s, Huber (1964) provided M-estimation for regression, where M stands for "maximum likelihood type" estimation. This method is robust in the presence of outliers in the response variable only but not gives any resistance when outliers are in explanatory variables. To tackle this problem, in literature several alternative approaches have been proposed for M-estimation, for example, Rousseeuw and Ryan (1997, 2008).

Theil (1950) and Sen (1968) proposed a Theil-Sen estimator which is popular and statistical efficient but has lower breakdown points than least trimmed square (LTS). Another approach has been proposed in the literature with the name of S-estimation. By using this approach, someone finds a line on a plane or hyper plane which minimizes the robust estimate of the residuals. It is more likely to resistant in the leverage points with a robust behavior to tackle the outliers in the response. However, this approach has been declared as an inefficient in literature.

Another method, which is called MM-estimation carries to retains the robustness and gaining efficiency of M-estimation, whilst resistance of S-estimation. It proceeds by gaining a highly robust and resistant S-estimate which minimizes an M-estimate in residual's scale. The estimated scale is considered as constant.

The robust regression is an iterative process to tackle the problem with outlier's data and reduces their impact in the regression coefficients. The basic objective is to use robust regression to locate reliable estimators of the parameters when outliers are present in the data. By using robust techniques, the sum of squared residual, are handled through some functional observations instead of ordinary approach by using OLS. First of all, there methods directly apply on data to fit the regression and then locate the outliers. The robust techniques are required the basic three properties, i.e. efficiency, breakdown point and bounded influence. Here, the breakdown point is the least fraction of the outliers to bears the tendency by an estimator. The bounded property provides a resistance against outliers by the estimator. This gives opportunity to the OLS by allowing the leverage point to exhibit a great influence.

1.7.1 Least absolute deviation method

The least absolute deviation (LAD) regression was proposed Boschovish (1757), which is still used as an alternative of least squares approach. LAD was improved by Edgeworth (1887), which minimizes the sum of absolute error in the following way:

$$\min \sum_{i=1}^{n} |\varepsilon| .$$

LAD is helpful to decrease the influence of outliers in the prospective of y-variable by OLS. On the other hand, it can be sensitive to detect outliers from the auxiliary variable. This is happened because of the low breakdown point of LAD with ratios $\frac{1}{n}$ and $\lim_{n\to\infty}\frac{1}{n}=0$.

1.7.2 Least median of squares method

To detect the outliers, another robust approach called least median of squares regression (LMS) can be used. For more about this approach, someone can consult to Rousseeuw and Leroy (2005). In general, this method is using the median as error squares instead of mean as:

$$min.\,median(\varepsilon_i^2).$$

This method is declared as robust in the presence of outliers in the direction of both Y and X with 0.5 as a breakdown point see, for example, Rousseeuw and Leory (2005).

1.7.3 Least trimmed square method

In least trimmed squares method (LTS), the square error term is sorted in ascending order, then sum of the first Z -observations are taken and minimized the following equation:

$$\min \sum \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2.$$

1.8 M- Estimators

For the parametric estimation in regression model, Huber (1964) introduced another estimations approach, i.e. M-estimation. This approach is based on maximum likelihood estimate but with efficient performance than OLS. Fox (2002) extended this estimation technique and suggested an alternative method which is mostly used in robust regression nowadays. In the presence of outliers or even mild outliers, this approach gives good estimates of parameters. We can obtain the parametric estimates by using the minimizing the residual function. Hence, the objective function of M-estimate can be written as:

$$\hat{\beta}_M = \min \sum_{i=1}^{n} \rho \left(Y_i - \sum_{j=1}^{k} x'_{ij}\beta_j\right) \qquad (1.1)$$

The above function gives a system of normal equations which is required to be solved. For solving this system, someone can differentiate this normal equation by using partial derivatives with equating them to zero. The following is the final form of this estimation technique:

$$\hat{\beta}_M = (x'wX)^{-1}(x'wY), \qquad (1.2)$$

where, w is the matrix having diagonal values of the weight matrix.

1.8.1 Huber-M estimation function

The class of M-estimation was suggested Huber (1973). Which is based on any function of outlier rather than the error square? Also, these methods are robust only when outliers in the direction of y -variable. The objective function of M-estimator is:

$$\min \sum_{i=1}^{n} \rho(e_i) ,$$

which is symmetric.

Huber's function ρ can be written as:

$$\rho(e)=\begin{cases}\dfrac{e^2}{2} & |e|\le k.\\ k|e|-\dfrac{k^2}{2}\end{cases}$$

1.8.2 Hample-M estimation function

The Hample- M estimation function was proposed by Hample (1971) and can be designed as:

$$\rho(y)=\begin{cases}\dfrac{y^2}{2}\\ a|y|-\dfrac{y^2}{2}\\ \dfrac{-a}{2(c-b)}(c-y)^2+\dfrac{a}{2}(b+c-a)\\ \dfrac{a}{2}(b+c-a)\end{cases}$$

where, $a = 1.7$, $b = 3.4$ and $c = 8.5$.

1.8.3 Tukey-M estimation function

It was proposed by Tukey (1977) and can be written in the following way:

$$\rho(y)=\begin{cases}\dfrac{1}{6}\left[1-\left[1-\left(\dfrac{y}{K}\right)^2\right]^3\right], & |y|\le K\\ \dfrac{1}{6}\end{cases}$$

where, K = 5 or K = 6.

1.9 Objectives of the Study

The main purpose of this study is to develop randomized responsemodels and estimator population means of sensitive characteristics using non-sensitive auxiliary variables.

The main objectives are:

- To develop some new estimators under simple random sampling for estimating the population characteristic using robust regression tools.
- To develop some new estimators under systematic random sampling for estimating the population characteristic using robust regression tools.
- To develop some new estimators under two stage simple and stratified random sampling scheme for estimating the population characteristics using robust regression tools.
- To assess the merits of proposed class of estimators on behalf of simultaneous study and real life applications.
- To conduct a simulation study for the comprehensive assessment of proposed estimators.

1.10 Literature survey

In literature, a lot of techniques have been developed for the estimation of regression coefficients. The most basic and popular method is known as the name of ordinary least square (OLS) and coined with two well-known researchers, Carl Friedrich Gauss and Adrien Marie Legendre in the first decade of 1800s. The basic theme to use this approach is to minimize the sum of squared of errors. Since then because of its authenticity and explicit nature, it became the only choice in regression analysis for many decades.

After using the Gaussian distribution as the error distribution by Guass, it became optimal with very useful mathematical results. Till to date, the OLS procedure is being used on popular basis because of ease of its computation. The other method which is called least absolute deviation (LAD) was proposed by Edgeworth (1887). According to his views, the errors squared by OLS method have a significant effect on the parameter estimators, so LAD is better choice in the presence of outliers. In current literature, the other approaches are being utilized as the alternatives of OLS and LAD methods. The difficulty to tackle the real-life data and advancement of computer technologies, researchers are developing some reasonable approaches.

In literature, a lot of methods have been proposed to tackle the problems of outliers. Huber (2009) suggested a family of regression estimator, i.e. M-estimators with the aim to minimize: $\sum_{i=1}^{n} p(u_i)$, whereas $p(u_i)$ is some symmetric function of errors.

Rousseeuw (1984) proposed another class of estimators known as least median of squares (LMS). The LMS is a similar to OLS instead of mean deviation because in this approach median is used for squared deviation. The LTS has also developed by Rousseeuw (1984), which is used to minimize the h-ordered square residuals with h is consistent which is essential to be determined. In this all those summation are included which have largest squared residuals.

For efficiency and high breakdown point Yohai (1987) proposed another approach, which is called MM-estimation. After almost one and half decade, Gervini and Yohai (2002) suggested another class of estimators which are called robust and efficient weighted least squares estimators (REWLSE).

In recent past, Li et al. (2011) suggested a class of robust techniques which depend on the regularization in the desired parameters for each response. Their study further verified that M-estimation is the special case of the suggested approach. In this prospective, someone can find another estimator in Marona et al. (2006). Another study related to robust measure through regression analysis is known as, the robust coefficient of determination, was proposed by Renaud and Victoria-Faser (2010).

A comparison study is available in literature between four robust regression approaches and OLS method conducted by Alma (2011). The four used methods in Alma (2011) study were LTS, M-estimate, S-estimate and MM-estimate. In their conclusion remarks, the S-estimate and MM-estimate perform best with the presence of outliers. Mohebbi et al. (2007) also compared two robust methods, i.e. Huber M-estimate and LAD with some nonparametric methods. They concluded that LAD and Huber M-estimate are suitable in the heavy tailed distribution whereas nonparametric and LAD regression approaches are the good choice in the presence of skewed data.

Al-Noor and Mohammad (2013) compared some nonparametric techniques with three robust methods in simple linear regression models. In their study, they suggested that LAD and M-estimation are good approaches when need to compare with nonparametric and OLS methods in presence of outliers.

Since the introduction of RRT, a lot of versions have been proposed by several researchers in both setups, *i.e.* quantitative and qualitative. Throughout this study, we focused on the quantitative randomized response models, only. These models are used to handle the sensitive characteristic by the respondents. These approaches give an opportunity to provide their responses without showing their identity. They also provide a random behavior between asking the question and the individual's response. In this context, Warner (1965) provided the following model:

$$Z = Y_{\rho} + (1-\rho)(1-Y). \tag{1.3}$$

The quantitative additive version of Warner (1965) for full randomization model is specified by Warner (1971). It was completely randomization model with respect to their responses based on random device with known mean and variance. Mathematically, in additive model, let Z and Y are used as reported response and sensitive variables, with both μ_Y and σ_Y^2 are unknown, whereas S is scrambling variable (independent of Y) with known true mean μ_S and known variance σ_S^2. The model is in the following form:

$$Z = Y + S. \tag{1.4}$$

Taking expectation on both side,

$$\begin{aligned} E(Z) &= E(Y) + E(S) \\ &= \mu_Y + \mu_S \end{aligned} \tag{1.5}$$

By solving equation (1.2), we can obtain the following unbiased estimator of mean as:

$$\mu_{YW} = \overline{Z} - \mu_S, \tag{1.6}$$

The variance is given as,

$$\begin{aligned} Var(\hat{\mu}_{YW}) &= Var(\overline{Z}) = \frac{\sigma_Z^2}{n}, \\ Var(\hat{\mu}_{YW}) &= \frac{\sigma_Y^2}{n} + \frac{\sigma_S^2}{n}. \end{aligned} \tag{1.7}$$

Kumar et al. (2016) provided some ratio type estimators with sin few modifications in the mean by using non-conventional measures of dispersion, such as Gini's mean difference, Downtown's technique and moments based on probability weights suggested by Abid et al. (2016). These are with linear combination of skewness and kurtosis of auxiliary variable. In this study, up to first order of approximation of the large sample properties are studied for bias and mean squared error. A thoroughly study of the newly developed estimators along with the comparison with competition as well.

Abid et al. (2016) evaluated conventional measures of location which are commonly used in the development of ratio estimators. We attempt to use some non-conventional location measures, for example, Hodges-Lehmann, tri-mean and mid-range of the auxiliary variable in this study. To improve the efficiency of the newly proposed ratio estimators for population mean, the coefficient of variation, the correlation coefficient with the linear combinations of auxiliary variable is also exploited. For this, the underline assumptions which are also associated with newly developed estimators are evaluated with the help of bias and MSE.

Subzar et al. (2020) suggested that the robust regression methods for simple random sampling without replacement (SRSWOR) by using the Boweley's coefficient of skewness as supplementary information. For these newly developed estimators, the authors used the simple OLS, Mallows GM-estimate, Huber-M, SIS GM-estimate and Schweppes GM-estimate techniques for estimation of population parameters.

In recent past, Shahzad et al. (2021) suggested a quantile-based regression by using MCD based parametric location measures etc. In their study, they have also provided the mean squared errors for all the suggested estimators. Other than theory perspective, we also conduct a simulation study to measure the performance of their new estimators.

Zaman et al. (2021) improved the Shahzad et al. (2021) estimators by using a lot of new robust regression approaches. They compared their estimators with the several competitors' estimators which are presented in literature.

By the reviewed robust regression coefficients, we will also develop some new estimators for efficient estimation of population mean. Further, Additive and Bar-Lev models will be used for the case of sensitive study variable.

Chapter 2

Robust regression type-estimators for sensitive variables.

2.1 Background

In surveys concerning delicate inquiries, for example, betting, liquor addiction, sexual conduct, tax avoidance, illicit wage and else, coordinate procedures for gathering data may prompt talked with individuals to give untruthful or deceptive reactions (responses). To decrease non-respondents rates and one-sided reactions emerging from sensitive, humiliating, threatening questions, a few statistical procedures might be utilized to guarantee interviewee anonymity or, a higher level of certainty. Such procedures, known as randomized response strategies or techniques, utilize a randomization gadget, for example, a die or a deck of cards, instead of a true reaction to gather solid data on sensitive issues. Based upon the result created by the randomization gadget, the interviewee gives an answer concerning his/her actual status. Since the questioner is unconscious of the aftereffect of the gadget, the utilization of these strategies guarantees that respondents can't be identified based on their answers. Warner (1965) was the first one who introduced a randomized response method. After that many authors extend their work such as Pollock and Beck (1976) and Bar-Lev et al. (2004).

In case of sensitive research, estimation of mean is a major concern in survey studies and regression estimators utilizing traditional regression coefficient are the most favored choices for it. Recently, Zaman and Bulut (2018) suggested a new class of ratio-type estimators for the mean estimation of non-sensitive variable utilizing robust regression coefficients. In this Chapter, we have generalized their family of estimators to the case where the study variable refers to sensitive issues which produce measurement errors due to non-responses and/or untruthful reporting. These errors may be reduced by enhancing respondent cooperation through scrambled response methods that mask the true value of the sensitive variable. Hence, two scrambled response models by Pollock and Beck (1976) and Bar-Lev et al. (2004) are discussed for the purposes of this research. In case of sensitive research, we developed a new family of robust regression-type estimators. Some estimators belonging to the class are shown and the mean square errors are determined. Theoretical and empirical illustration is done through real and artificial data sets for assessing the performance of adapted and proposed class.

Many authors, such as Koyuncu (2012); Shahzad (2016) and Shahzad et al. (2017) have developed a family for simple random sampling's estimators by involving auxiliary information when study variable is non-sensitive. Similarly, Shahzad et al. (2018) have studied ratio, exponential and traditional regression estimators for mean estimation when study variable is sensitive. In case of positive correlation, ratio-type estimators are suitable for mean estimation. But when outliers are presented in data, these traditional estimators are not suitable and hence not provide much efficient results. Keeping this fact in mind, Zaman and Bulut (2018) introduced robust ratio type estimators for non-sensitive study variable. Taking motivation from their work, we have generalized their estimators defining a more general class of robust-ratio-type estimators for the sensitive setup. After that, we also contributed a new family of robust-regression-type estimators for sensitive setup under simple random sampling scheme.

The remaining part of this chapter is constructed as follows: In Section 2.2, we have introduced the basic terminology and generalized Zaman and Bulut (2018) robust-ratio-type estimators for the sensitive setup in presence of supplementary information. In Section 2.3, we have proposed the new family of robust regression-type estimators with MSE up to first-order of approximation. In Section 2.4, we have discussed the efficiency of proposed class over the adapted estimators through a real and an artificial data set. For sensitive setup, Pollock and Beck (1976) and Bar-Lev et al. (2004) randomized techniques are used. Both, theoretical and empirical illustrations are performed for assessing the efficiency of proposed class. Conclusion of the whole study is provided in Section 2.5.

2.2 Generalized family of estimators

In current section, following Zaman and Bulut (2018) we try to define a more general class of ratio-type-estimators utilizing robust regression methods namely: LAD, LMS, LTS, Huber-M, Hampel-M, Tukey-M, and Huber-MM for the estimation of population mean of sensitive study variable Z utilizing supplementary variable X under simple random sampling scheme. The most popular method for estimating the parameters is the least squares (LS). One of the simplest robust alternatives to the LS is the least absolute deviations (LAD) method which was introduced in 1757 by Roger Joseph Boscovich. Huber-M technique developed by Huber (1964) which was the next step towards robust regression.

Huber (1973) extended his own work and utilized this technique in regression modeling, hence a new robust regression estimate developed. The main theme of this technique is to replace squared error in ordinary least square by ρ , where ρ is some symmetric function. Many authors extended the work of Huber (1973) such as, Hampel (1971) named as Hampel-M estimate, Tukey (1977) named as Tukey-M estimate and Yohai (1987) named as Huber-MM estimate. We also utilize TLS, which is known as

Trimmed Least Squares robust regression method in presence of outliers, introduced by Rousseeuw and Yohai (1984).

The last robust regression technique included in this study is LMS, which is known as "Least Median of Squares" developed by Rousseeuw and Leroy (1987). This method based on minimization of median of error squares rather than mean of error squares. For details about all these robust regression methods, readers are referred to Zaman and Bulut (2018). Utilizing these measures Zaman and Bulut (2018) constructed a class of estimators when study variable was non-sensitive. Now we are considering Zaman and Bulut (2018) work for randomized response or sensitive setup. Randomized response methods are utilized to diminish refusal rates and one-sided reactions to delicate inquiries. Warner (1965) presented the scrambling device for the extent of a population portrayed by a delicate/threatening variable, which was based on the issues related to sex, xenophobia, abortion, drugs etc. The purpose of these devices to get truthful answers from respondents.

Quantitative randomized response techniques are utilized to estimate different measures of descriptive statistics such as mean, variance, mode etc., in a population. However, our study is limited to mean estimation. For instance, the delicate investigation variable might be the aggregate number of premature births a lady has had, the normal week after week liquor utilization, yearly income of individuals etc. These Randomized response methods are categorized as additive or multiplicative models. In additive model, the scrambling variable S (say) is added in the true response of respondent. The distribution of S is known such as Normal, Weibull, and Uniform etc. Similarly, in multiplicative model, the scrambling variable is multiplied with the true response of respondent. The whole procedure is done for avoiding refusal rates and getting more truthful responses from respondents. For more details, interested readers are referred to, Warner (1971); Pollock and Beck (1976) and Bar-Lav, Bobovitch and Boukai (2004). So we get here a motivation and extend the work utilizing robust regression measures for mean estimation in sensitive setup.

We have generalized Zaman and Bulut (2018) class for the case when study variable is sensitive in nature as

$$\bar{Z}_{zb_i} = \frac{\bar{z} + b_i\left(\bar{X} - \bar{x}\right)}{\left(F_i\bar{x} + G_i\right)}\left(F_i\bar{X} + G_i\right), \qquad \text{for,} \qquad i = 1, 2, \ldots, 35 \tag{2.1}$$

where, b_i is slope or regression coefficient, calculated from the above mentioned robust regression methods. Constants, $F_i \neq 0$ and G_i are either (0, 1) or known characteristics of the population such as, C_x, the coefficients of variation, $\beta_2(x)$, the coefficients of kurtosis from the population having N identifiable units. We can generate many new estimators using suitable variables for b_i, F_i and G_i as given in **Table (2.1).**

Further, $\left(\bar{X}, \bar{Z}\right)$ are population means, $\left(\bar{x}, \bar{z}\right)$ are their corresponding sample mean from the finite population with simple random sample. Some members of the family which are generated from $\bar{z}_{zb_i}$ are presented in **Table (2.1).**

Let, $\hat{\gamma}_z = \dfrac{\bar{z} + b_i\left(\bar{X} - \bar{x}\right)}{F_i\bar{x} + G_i}$.

Hence,

$$\bar{z}_{zb_i} = \hat{\gamma}_z\left(F_i\bar{X} + G_i\right). \tag{2.2}$$

Now we find MSE of $\bar{z}_{zb_i}$ through Taylor series, defined as:

$$h(\bar{x},\bar{z})=h(\bar{X},\bar{Z})+\left[\frac{\delta h(\bar{x},\bar{z})}{\delta \bar{x}}\right]_{\bar{X},\bar{Z}}(\bar{x}-\bar{X})+\left[\frac{\delta h(\bar{x},\bar{z})}{\delta \bar{z}}\right]_{\bar{X},\bar{Z}}. \tag{2.3}$$

Here,

$$h(\bar{x},\bar{z})=\hat{\gamma}_z \quad \text{and} \quad h(\bar{\mathrm{X}},\bar{\mathrm{Z}})=\gamma_z$$

so,

$$\hat{\gamma}_z=\gamma_z+\left[\frac{\delta h(\bar{x},\bar{z})}{\delta \bar{x}}\right]_{\bar{X},\bar{Z}}(\bar{x}-\bar{X})+\left[\frac{\delta h(\bar{x},\bar{z})}{\delta \bar{z}}\right]_{\bar{X},\bar{Z}}(\bar{z}-\bar{Z}),$$

$$\hat{\gamma}_z-\gamma_z=\left[\frac{\delta h(\bar{x},\bar{z})}{\delta \bar{x}}\right]_{\bar{X},\bar{Z}}(\bar{x}-\bar{X})+\left[\frac{\delta h(\bar{x},\bar{z})}{\delta \bar{z}}\right]_{\bar{X},\bar{Z}}(\bar{z}-\bar{Z}),$$

$$\hat{\gamma}_z-\gamma_z=\left[\frac{\delta\left(\frac{\bar{z}+b_t(\bar{X}-\bar{x})}{F_t\bar{x}+G_t}\right)}{\delta \bar{x}}\right]_{\bar{X},\bar{Z}}(\bar{x}-\bar{X})+\left[\frac{\delta\left(\frac{\bar{z}+b_t(\bar{X}-\bar{x})}{F_t\bar{x}+G_t}\right)}{\delta \bar{z}}\right]_{\bar{X},\bar{Z}}(\bar{z}-\bar{Z}).$$

Now, partially differentiating 1st term w. r. t. $\bar{x}$ and 2nd term w. r. t. $\bar{z}$.

$$\hat{\gamma}_z-\gamma_z=\left[-\frac{\bar{z}}{(F_t\bar{x}+G_t)^2}-\frac{b_t(F_t\bar{x}+G_t)}{(F_t\bar{x}+G_t)^2}\right]_{\bar{X},\bar{Z}}(\bar{x}-\bar{X})+\left[\frac{1}{\bar{x}+\tau}\right]_{\bar{X},\bar{Z}}(\bar{z}-\bar{Z}). \tag{2.4}$$

Now first squaring and then applying expectation. After that putting, $\bar{x}-\bar{X}$, $\bar{z}-\bar{Z}$, and $b_t=B_t$. We get MSE of generalized family of estimators in compressed form as given below:

$$MSE(\bar{Z}_{zb_t})=\left(\frac{1-f}{n}\right)\left[S_z^2+R_{zh_t}^2S_x^2+2B_tR_{zb_t}S_x^2+B_t^2S_x^2-2R_{zb_t}S_{xz}-2BS_{xz}\right]. \tag{2.5}$$

for, $i = 1, 2, \ldots, 35.$

where,

$$R_{zb_i} = \frac{F_i \bar{X}}{F_i \bar{x} + G_i}, \qquad \text{for,} \quad i = 1, 2, \ldots, 35.$$

Further,

$$S_z^2 = \frac{\sum_{i=1}^{N} (z_i - \bar{Z})^2}{N-1}, \quad S_x^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{X})^2}{N-1},$$ are the unbiased variance of Z and X respectively.

$$S_{zx} = \frac{\sum_{i=1}^{N} (z_i - \bar{Z})(x_i - \bar{X})}{N-1},$$ is the covariance between (X, Z).

2.3 Proposed class of robust regression-type estimators

Motivated by Zaman and Bulut (2018) traditional ratio estimators, we propose a new class of robust regression type estimators for sensitivity issue as:

$$\bar{Z}_{N_i} = \bar{z} + b_i \left(\bar{X} - \bar{x} \right), \qquad \text{for,} \qquad i = 1, 2, \ldots, 7. \tag{2.6}$$

The proposed class of estimators require less amount of information except $\bar{y}_{zb_1}$, $\bar{y}_{zb_6}$, $\bar{y}_{zb_{11}}$, $\bar{y}_{zb_{16}}$, $\bar{y}_{zb_{21}}$, $\bar{y}_{zb_{26}}$, and $\bar{y}_{zb_{31}}$. As we know from literature that mean of the sensitive study variate is unknown so the information of z is collected through sample. Two scrambled response models:

(i) The (Pollock and Beck model, 1976), $[\bar{Z} = \bar{Y} + \bar{S}]$

(ii) The (Bar Lev et al. model, 2004), $\bar{Z} = \bar{Y}[(1-P)\bar{S} + P]$ are utilized for collecting information about scrambled response. Our proposed class of estimators is based on simple the ratio (mean per unit) type component. All the seven proposed estimators rely on M, Tukey-M and Huber-MM methods for $i(i = 1, 2, \dots, 7)$, respectively, adapting Zaman and Bulut (2018). Interested readers may refer to Zaman and Bulut (2018).

for, $i = 1$; b_{lad} = Robust-regression coefficient computed from LAD method

for, $i = 2$; b_{lms} = Robust-regression coefficient computed from LMS method

for, $i = 3$; b_{lts} = Robust-regression coefficient computed from LTS method

for, $i = 4$; b_{hbm} = Robust-regression coefficient computed from Huber-M method

for, $i = 5$; b_{hpm} = Robust-regression coefficient computed from Hample-M method

for, $i = 6$; b_{tky} = Robust-regression coefficient computed from Tukey-M method

for, $i = 7$; b_{hmm} = Robust-regression coefficient computed from Huber-MM method.

Note that all the notations used in $\bar{Z}_N$, already described in previous section. Further, our new constructed class can be arranged in the framework of Mukhopadhyay (1998). But we are implementing their frame work in case of sensitivity. Therefore, taking the benefit of known results with some simple algebra, avoiding tedious or fruitless calculations, we mention the MSE expressions of the purpose class of estimators, up to order n^{-1} as:

$$MSE\left(\bar{Z}_{N_i}\right) = \left[V(\bar{z}) - 2B_i Cov(\bar{x}, \bar{z}) + B_i^2 V(\bar{x})\right]. \tag{2.7}$$

Now substituting,

$$V(\bar{z})=\left(\frac{1-f}{n}\right)S_z^2, \qquad Cov(\bar{x},\bar{z})=\rho_{xz}S_xS_z, \qquad V(\bar{x})=\frac{1-f}{n}S_x^2.$$

And get final expression for MSE as follows,

$$MSE\left(\bar{Z}_{N_i}\right)=\left(\frac{1-f}{n}\right)\left[\bar{Z}^2C_z^2-2B_i\bar{X}\bar{Z}\rho_{xz}C_xC_z+B_i^2\bar{X}^2C_x^2\right], \quad for\ i=1,2,...,7 \quad (2.8)$$

where, C_z is coefficient of variation, ρ_{xz} is the coefficients of correlation between X and Z. In **Table (2.2),** we present all the members of the proposed class. Note that the notations of descriptive measures (mean, variance, coefficient of variation etc.) provided in previous section are based on general frame work. So, in this section, we are providing the theoretical descriptive measures, keeping the fact in mind that data is collected through two randomized methods as mentioned above.

In case of Pollock and Beck (1976) model,

$$\bar{z}=\bar{Y}+\bar{S}$$

$$S_z^2=S_y^2+S_s^2$$

$$C_z=\sqrt{S_y^2+S_s^2\bar{Z}}$$

$$\rho_{xy}=\frac{S_{xy}}{S_x\sqrt{S_y^2+S_s^2}}.$$

where,

$$S_{xy}=\frac{\sum_{i=1}^{N}\left(y_i-\bar{Y}\right)\left(x_i-\bar{X}\right)}{N-1}, \qquad S_y^2=\frac{\sum_{i=1}^{N}\left(y_i-\bar{Y}\right)^2}{N-1}, \qquad S_s^2=\frac{\sum_{i=1}^{N}\left(s_i-\bar{S}\right)^2}{N-1}.$$

In case of Bar Lev et al. (2004) model,

$$\bar{Z} = \bar{Y}\left[(1-P)\bar{S} + P\right], \tag{2.9}$$

$$S_z^2 = \bar{Y}^2\left(C_y^2+1\right)\left[(1-P)\bar{S}^2\left(1+C_s^2\right)+P\right]-\bar{Z}^2, \tag{2.10}$$

$$C_z = \frac{\sqrt{\bar{Y}^2\left(C_y^2+1\right)\left[(1-P)\bar{S}^2\left(1+C_s^2\right)+P\right]-\bar{Z}^2}}{\bar{Y}\left[(1-P)\bar{S}+P\right]}, \tag{2.11}$$

$$\rho_{xz} = \frac{S_{xy}\left[(1-P)\bar{S}+P\right]}{S_x\sqrt{\bar{Y}\left(C_y^2+1\right)\left[(1-P)\bar{S}^2\left(1+C_s^2\right)+P\right]-\bar{Z}^2}}. \tag{2.12}$$

Note that these, theoretical measures are also valid/applicable for generalized family, which we construct on the lines of Zaman and Bulut (2018).

2.4 Efficiency comparisons

In current section, regarding the amount of benefits which can be achieved utilizing the proposed robust-regression estimators, we move towards simulation study. One real (Pop-1) and one artificial (Pop-2) population is considered for the purpose of this study.

Pop-1 is taken from Singh (2003) p. 1111,

where

X = Amount of non-real estate farm loans during 1977 and,

Y = Amount of real estate farm loans during 1977.

We consider this real population because it contains outliers. The size of population is $N = 100$. We take a sample of size $n = 20$

Pop-2 contains artificial data adapting Shahzad et al. (2018), having size of population $N = 1000$ generated from bivariate normal distributions for (Y, X) with mean vector $(\bar{Y}, \bar{X}) = (2,2)$ and variance-covariance matrix given by Pop-2.

$$\Sigma = \begin{bmatrix} 12 & 3 \\ 3 & 2 \end{bmatrix}.$$

From both real and artificial populations, $K' = 6000$ samples of size $n = 150$ are selected according to SRSWOR and for the $k'th$ sample the estimate $\hat{\theta}'^{(k')}$ of $\bar{Z}$ (mean of sensitive study variable) is computed $\hat{\theta}' = \bar{Z}_{zb_1} - \bar{Z}_{zb_{35}}, \bar{Z}_{N_1} - \bar{Z}_{N_7}$. for each adapted and proposed class of estimators, both theoretical and empirical MSEs are obtained. Following Koyuncu et al. (2014), the scrambling variate, $S \approx N(zero, \sigma)$, where, σ is the standard deviation equal to 10% of the standard deviation of auxiliary variable is utilized. Note that same scrambling device is used in both, (i) The (Pollock and Beck, 1976) $[Z = Y + S]$; (ii) The (Bar Lev et al., 2004) $[Z = (1-p)YS + pY]$ models. The steps of simulation by adapting Kadilar, Candan and Cingi (2007) and Abu-Dayyeh et al. (2003) are as follows:

Step 1: A SRSWOR selected of size n from both data sets.

Step 2: Use the data of above step to find the value of mean.

Step 3: Step (1) and (2) is repeated $\acute{K} = 6000$ (say) times.

Step 4: Empirical MSE is computed for each up-to $\acute{K}$ and then averaged as:

$$MSE = \frac{\sum_{i=1}^{\acute{K}} (z_i - \bar{Z})^2}{\acute{K}}.$$

Step 5: Theoretical MSE is computed for each up-to $\acute{K}$ and then averaged as:

$MSE = \frac{\sum_{i=1}^{\acute{K}}(MSE(z_i))}{\acute{K}}$, where z_i shows estimated mean of sensitive study variable for $i = 1, 2, \ldots, \acute{K}$ and $\bar{Z}$ is the population mean. It is worth mentioning that the values of z_i calculated from equation (2.1) and (2.6). Note that Zaman and Bulut (2018) also followed above mentioned simulation steps for empirical MSE in case of nonsensitive study variate. However, we have followed these steps for sensitive study variate. For theoretical $MSE(z_i)$ we adapted Koyuncu, Gupta and Sousa (2014) in sensitive setup.

The results of simulation study are provided in **Tables (2.3 - 2.5)**.

2.5 Summary of the chapter

In this chapter, beginning from some recent utilization of robust regression on design-based sampling from finite populations, we generalize Zaman and Bulut (2018) family of ratio estimators to the case in which the study variable is thought to be delicate/ sensitive issue and its values are gathered on survey units by means of scrambled responses in order to secure respondent protection, improving participation and diminishing nonresponse rate or potentially untruthful answers. For this purpose, two scrambled response models namely Pollock and Beck (1976) and Bar Lev et al. (2004) have been considered for perturbing the appropriate responses. Also, we develop new robust-regression type estimators$\left(\bar{Z}_{N_i}\right)$ for delicate study variable. The proposed new robust-regression type estimators require less supplementary information as compared to robust-ratio type estimators. The MSE equation for the new estimators are also obtained. The superiority of the proposed class has been evaluated by simulation study utilizing both theoretically and empirically. One real and one artificial population is utilized for this purpose. The numerical examinations well underline the predominance of the proposed class in sensitive setting, at least for the experimental circumstances considered. Thus, the proposed class is recommended for survey practitioners as it might expand the odds of getting progressively efficient results of unknown population mean of sensitive study variable.

Table 2.1: Generalized family of estimators.

Estimators	$b_{(i)}$	F_i	G_i
$\bar{Z}_{zb_1}$	$b_{(lad)}$	1	0
$\bar{Z}_{zb_2}$	$b_{(lad)}$	1	C_x
$\bar{Z}_{zb_3}$	$b_{(lad)}$	1	$\beta_2(x)$
$\bar{Z}_{zb_4}$	$b_{(lad)}$	$\beta_2(x)$	C_x
$\bar{Z}_{zb_5}$	$b_{(lad)}$	C_x	$\beta_2(x)$
$\bar{Z}_{zb_6}$	$b_{(lms)}$	1	0
$\bar{Z}_{zb_7}$	$b_{(lms)}$	1	C_x
$\bar{Z}_{zb_8}$	$b_{(lms)}$	1	$\beta_2(x)$
$\bar{Z}_{zb_9}$	$b_{(lms)}$	$\beta_2(x)$	C_x
$\bar{Z}_{zb_{10}}$	$b_{(lms)}$	C_x	$\beta_2(x)$
$\bar{Z}_{zb_{11}}$	$b_{(lts)}$	1	0
$\bar{Z}_{zb_{12}}$	$b_{(lts)}$	1	C_x
$\bar{Z}_{zb_{13}}$	$b_{(lts)}$	1	$\beta_2(x)$
$\bar{Z}_{zb_{14}}$	$b_{(lts)}$	$\beta_2(x)$	C_x
$\bar{Z}_{zb_{15}}$	$b_{(lts)}$	C_x	$\beta_2(x)$
$\bar{Z}_{zb_{16}}$	$b_{(hbm)}$	1	0
$\bar{Z}_{zb_{17}}$	$b_{(hbm)}$	1	C_x
$\bar{Z}_{zb_{18}}$	$b_{(hbm)}$	1	$\beta_2(x)$
$\bar{Z}_{zb_{19}}$	$b_{(hbm)}$	$\beta_2(x)$	C_x
$\bar{Z}_{zb_{20}}$	$b_{(hbm)}$	C_x	$\beta_2(x)$
$\bar{Z}_{zb_{21}}$	$b_{(hpm)}$	1	0
$\bar{Z}_{zb_{22}}$	$b_{(hpm)}$	1	C_x
$\bar{Z}_{zb_{23}}$	$b_{(hpm)}$	1	$\beta_2(x)$
$\bar{Z}_{zb_{24}}$	$b_{(hpm)}$	$\beta_2(x)$	C_x

$\bar{Z}_{zb_{25}}$	$b_{(hpm)}$	C_x	$\beta_2(x)$
$\bar{Z}_{zb_{26}}$	$b_{(tkz)}$	1	0
$\bar{Z}_{zb_{27}}$	$b_{(tkz)}$	1	C_x
$\bar{Z}_{zb_{28}}$	$b_{(tkz)}$	1	$\beta_2(x)$
$\bar{Z}_{zb_{29}}$	$b_{(tkz)}$	$\beta_2(x)$	C_x
$\bar{Z}_{zb_{30}}$	$b_{(tkz)}$	C_x	$\beta_2(x)$
$\bar{Z}_{zb_{31}}$	$b_{(hmm)}$	1	0
$\bar{Z}_{zb_{32}}$	$b_{(hmm)}$	1	C_x
$\bar{Z}_{zb_{33}}$	$b_{(hmm)}$	1	$\beta_2(x)$
$\bar{Z}_{zb_{34}}$	$b_{(hmm)}$	$\beta_2(x)$	C_x
$\bar{Z}_{zb_{35}}$	$b_{(hmm)}$	C_x	$\beta_2(x)$

Table 2.2: Family member proposed class with MSE's.

Estimators	MSE
$\bar{Z}_{N1} = \bar{z} + b_{lad}(\bar{X} - \bar{x})$	$MSE(\bar{Z}_{N1}) = \left(\frac{1-f}{n}\right)\left[\bar{Z}^2 C_z^2 - 2B_{lad}\bar{X}\bar{Z}\rho_{xz}C_xC_z + B_{lad}^2\bar{X}^2C_x^2\right]$
$\bar{Z}_{N2} = \bar{z} + b_{lms}(\bar{X} - \bar{x})$	$MSE(\bar{Z}_{N2}) = \left(\frac{1-f}{n}\right)\left[\bar{Z}^2 C_z^2 - 2B_{lms}\bar{X}\bar{Z}\rho_{xz}C_xC_z + B_{lms}^2\bar{X}^2C_x^2\right]$
$\bar{Z}_{N3} = \bar{z} + b_{lts}(\bar{X} - \bar{x})$	$MSE(\bar{Z}_{N3}) = \left(\frac{1-f}{n}\right)\left[\bar{Z}^2 C_z^2 - 2B_{lts}\bar{X}\bar{Z}\rho_{xz}C_xC_z + B_{lts}^2\bar{X}^2C_x^2\right]$
$\bar{Z}_{N4} = \bar{z} + b_{hhm}(\bar{X} - \bar{x})$	$MSE(\bar{Z}_{N4}) = \left(\frac{1-f}{n}\right)\left[\bar{Z}^2 C_z^2 - 2B_{hhm}\bar{X}\bar{Z}\rho_{xz}C_xC_z + B_{hhm}^2\bar{X}^2C_x^2\right]$
$\bar{Z}_{N5} = \bar{z} + b_{hpm}(\bar{X} - \bar{x})$	$MSE(\bar{Z}_{N5}) = \left(\frac{1-f}{n}\right)\left[\bar{Z}^2 C_z^2 - 2B_{hpm}\bar{X}\bar{Z}\rho_{xz}C_xC_z + B_{hpm}^2\bar{X}^2C_x^2\right]$
$\bar{Z}_{N6} = \bar{z} + b_{tky}(\bar{X} - \bar{x})$	$MSE(\bar{Z}_{N6}) = \left(\frac{1-f}{n}\right)\left[\bar{Z}^2 C_z^2 - 2B_{tky}\bar{X}\bar{Z}\rho_{xz}C_xC_z + B_{tky}^2\bar{X}^2C_x^2\right]$
$\bar{Z}_{N7} = \bar{z} + b_{hmm}(\bar{X} - \bar{x})$	$MSE(\bar{Z}_{N7}) = \left(\frac{1-f}{n}\right)\left[\bar{Z}^2 C_z^2 - 2B_{hmm}\bar{X}\bar{Z}\rho_{xz}C_xC_z + B_{hmm}^2\bar{X}^2C_x^2\right]$

Table 2.3: MSE Pop-1 (real) and Pop-2 (artificial) using the Additive scrambled response model, $\bar{Z} = \bar{Y} + \bar{S}$

$\hat{\theta}$	Pop-1		$\hat{\theta}$	Pop-2	
	Theoretical	**Empirical**		**Theoretical**	**Empirical**
$\bar{Z}_1$	17475.68	22656.98	$\bar{Z}_1$	0.05428	0.05641
$\bar{Z}_2$	17437.0	22578.42	$\bar{Z}_2$	0.04882	0.05052
$\bar{Z}_3$	17331.95	22366.09	$\bar{Z}_3$	0.04442	0.04590
$\bar{Z}_4$	4750.66	22639.92	$\bar{Z}_4$	0.04365	0.05374
$\bar{Z}_5$	17359.14	22420.9	$\bar{Z}_5$	0.04379	0.04525
$\bar{Z}_6$	17346.64	22475.48	$\bar{Z}_6$	0.05269	0.05478
$\bar{Z}_7$	17308.15	22397.43	$\bar{Z}_7$	0.04768	0.04937
$\bar{Z}_8$	17203.62	22186.49	$\bar{Z}_8$	0.04386	0.04534
$\bar{Z}_9$	4726.71	22458.53	$\bar{Z}_9$	0.04326	0.05231
$\bar{Z}_{10}$	17230.67	22240.94	$\bar{Z}_{10}$	0.04337	0.04482
$\bar{Z}_{11}$	44090.08	60247.65	$\bar{Z}_{11}$	0.05285	0.05493
$\bar{Z}_{12}$	44023.17	60077.99	$\bar{Z}_{12}$	0.04779	0.04948
$\bar{Z}_{13}$	43841.22	59619.16	$\bar{Z}_{13}$	0.04391	0.04539
$\bar{Z}_{14}$	15616.55	60210.81	$\bar{Z}_{14}$	0.04330	0.05245
$\bar{Z}_{15}$	43888.34	59737.63	$\bar{Z}_{15}$	0.04340	0.04486
$\bar{Z}_{16}$	18463.22	24046.42	$\bar{Z}_{16}$	0.05494	0.05710
$\bar{Z}_{17}$	18423.13	23963.99	$\bar{Z}_{17}$	0.04930	0.05101
$\bar{Z}_{18}$	18314.24	23741.17	$\bar{Z}_{18}$	0.04467	0.04615
$\bar{Z}_{19}$	4950.638	24028.52	$\bar{Z}_{19}$	0.04384	0.05435
$\bar{Z}_{20}$	18342.42	23798.68	$\bar{Z}_{20}$	0.04399	0.04545
$\bar{Z}_{21}$	18284.16	23794.44	$\bar{Z}_{21}$	0.05508	0.05724
$\bar{Z}_{22}$	18244.33	23712.71	$\bar{Z}_{22}$	0.04940	0.05111
$\bar{Z}_{23}$	18136.12	23491.77	$\bar{Z}_{23}$	0.04473	0.04621
$\bar{Z}_{24}$	4912.264	23776.69	$\bar{Z}_{24}$	0.04388	0.05447
$\bar{Z}_{25}$	18164.13	23548.8	$\bar{Z}_{25}$	0.04404	0.04549
$\bar{Z}_{26}$	18173.33	23638.48	$\bar{Z}_{26}$	0.05482	0.05697
$\bar{Z}_{27}$	18133.66	23557.18	$\bar{Z}_{27}$	0.04921	0.05092
$\bar{Z}_{28}$	18025.88	23337.42	$\bar{Z}_{28}$	0.04463	0.04611
$\bar{Z}_{29}$	4888.969	23620.83	$\bar{Z}_{29}$	0.04381	0.05424
$\bar{Z}_{30}$	18053.77	23394.14	$\bar{Z}_{30}$	0.04396	0.04541
$\bar{Z}_{31}$	18084.72	23513.79	$\bar{Z}_{31}$	0.05483	0.05698
$\bar{Z}_{32}$	18045.17	23432.84	$\bar{Z}_{32}$	0.04921	0.05093
$\bar{Z}_{33}$	17937.73	23214.01	$\bar{Z}_{33}$	0.04463	0.04611
$\bar{Z}_{34}$	4870.598	23496.21	$\bar{Z}_{34}$	0.04381	0.05424
$\bar{Z}_{35}$	17965.54	23270.5	$\bar{Z}_{35}$	0.04396	0.04541
$\bar{Z}_{N_1}$	4360.634	4158.825	$\bar{Z}_{N_1}$	0.04281	0.04419
$\bar{Z}_{N_2}$	4365.726	4164.278	$\bar{Z}_{N_2}$	0.04293	0.04432
$\bar{Z}_{N_3}$	4400.21	4210.11	$\bar{Z}_{N_3}$	0.04292	0.04430
$\bar{Z}_{N_4}$	4342.96	4137.782	$\bar{Z}_{N_4}$	0.04279	0.04416
$\bar{Z}_{N_5}$	4343.464	4138.975	$\bar{Z}_{N_5}$	0.04270	0.04411
$\bar{Z}_{N_6}$	4344.361	4140.28	$\bar{Z}_{N_6}$	0.04271	0.04410
$\bar{Z}_{N_7}$	4345.403	4141.64	$\bar{Z}_{N_7}$	0.04278	0.04418

Table 2.4: MSE Pop-1 using the Bar-Lev scrambled response model, $\bar{Z} = (1-P)\bar{Y}\bar{S} + \bar{Y}P$.

$\hat{\theta}$	P=0.40		$\hat{\theta}$	P=0.80	
	Theoretical	Empirical		Theoretical	Empirical
$\bar{Z}_1$	46645.32	48279.29	$\bar{Z}_1$	4370.54	4681.15
$\bar{Z}_2$	46638.24	48260.43	$\bar{Z}_2$	4370.02	4679.27
$\bar{Z}_3$	46619.01	48209.37	$\bar{Z}_3$	4368.62	4674.19
$\bar{Z}_4$	44358.65	48275.2	$\bar{Z}_4$	4301.44	4680.75
$\bar{Z}_5$	46623.98	48222.56	$\bar{Z}_5$	4368.99	4675.50
$\bar{Z}_6$	46833.83	48554.47	$\bar{Z}_6$	4359.23	4649.27
$\bar{Z}_7$	46826.56	48534.94	$\bar{Z}_7$	4358.70	4647.41
$\bar{Z}_8$	46806.79	48482.07	$\bar{Z}_8$	4357.26	4642.35
$\bar{Z}_9$	44339.59	48550.23	$\bar{Z}_9$	4282.28	4648.87
$\bar{Z}_{10}$	46811.91	48495.72	$\bar{Z}_{10}$	4357.63	4643.66
$\bar{Z}_{11}$	46574.73	48176.32	$\bar{Z}_{11}$	4366.22	4673.70
$\bar{Z}_{12}$	46567.72	48157.72	$\bar{Z}_{12}$	4365.71	4671.83
$\bar{Z}_{13}$	46548.69	48107.36	$\bar{Z}_{13}$	4364.30	4666.76
$\bar{Z}_{14}$	44231.31	48172.28	$\bar{Z}_{14}$	4296.43	4673.29
$\bar{Z}_{15}$	46553.62	48120.37	$\bar{Z}_{15}$	4364.66	4668.07
$\bar{Z}_{16}$	46360.44	47864.1	$\bar{Z}_{16}$	4390.95	4714.68
$\bar{Z}_{17}$	46353.69	47846.31	$\bar{Z}_{17}$	4390.44	4712.76
$\bar{Z}_{18}$	46335.36	47798.17	$\bar{Z}_{18}$	4389.05	4707.58
$\bar{Z}_{19}$	44261.62	47860.23	$\bar{Z}_{19}$	4224.31	4714.26
$\bar{Z}_{20}$	46340.11	47810.61	$\bar{Z}_{20}$	4389.41	4708.92
$\bar{Z}_{21}$	47776.56	49933.46	$\bar{Z}_{21}$	4759.06	5263.55
$\bar{Z}_{22}$	47768.51	49910.85	$\bar{Z}_{22}$	4758.58	5260.98
$\bar{Z}_{23}$	47746.63	49849.65	$\bar{Z}_{23}$	4757.28	5254.04
$\bar{Z}_{24}$	44850.51	49928.55	$\bar{Z}_{24}$	4609.64	5262.99
$\bar{Z}_{25}$	47752.30	49865.45	$\bar{Z}_{25}$	4757.62	5255.83
$\bar{Z}_{26}$	46789.23	48489.34	$\bar{Z}_{26}$	5348.63	6114.77
$\bar{Z}_{27}$	46782.01	48469.95	$\bar{Z}_{27}$	5348.18	6111.15
$\bar{Z}_{28}$	46762.35	48417.51	$\bar{Z}_{28}$	5346.94	6101.34
$\bar{Z}_{29}$	44319.51	48485.13	$\bar{Z}_{29}$	5213.2	6113.99
$\bar{Z}_{30}$	46767.43	48431.05	$\bar{Z}_{30}$	5347.26	6103.87
$\bar{Z}_{31}$	47540.92	49588.46	$\bar{Z}_{31}$	5321.08	6075.23
$\bar{Z}_{32}$	47533.04	49566.59	$\bar{Z}_{32}$	5320.62	6071.65
$\bar{Z}_{33}$	47511.64	49507.39	$\bar{Z}_{33}$	5319.39	6061.98
$\bar{Z}_{34}$	44711.00	49583.71	$\bar{Z}_{34}$	5185.11	6074.45
$\bar{Z}_{35}$	47517.18	49522.68	$\bar{Z}_{35}$	5319.71	6064.48
$\bar{Z}_{N_1}$	44151.97	44142.70	$\bar{Z}_{N_1}$	4196.60	4224.01
$\bar{Z}_{N_2}$	44203.18	44195.44	$\bar{Z}_{N_2}$	4175.27	4189.63
$\bar{Z}_{N_3}$	44136.59	44126.64	$\bar{Z}_{N_3}$	4191.39	4217.66
$\bar{Z}_{N_4}$	44106.86	44094.36	$\bar{Z}_{N_4}$	4220.15	4251.54
$\bar{Z}_{N_5}$	44194.77	44590.74	$\bar{Z}_{N_5}$	4110.23	4168.11
$\bar{Z}_{N_6}$	44189.88	44181.81	$\bar{Z}_{N_6}$	4217.65	4295.26
$\bar{Z}_{N_7}$	44181.82	44477.31	$\bar{Z}_{N_7}$	4189.41	4266.29

Table 2.5: MSE Pop-2 using the Bar-Lev scrambled response model, $\bar{Z} = (1-P)\bar{Y}\bar{S} + \bar{Y}P$.

$\hat{\theta}$	P=0.40		$\hat{\theta}$	P=0.80	
	Theoretical	**Empirical**		**Theoretical**	**Empirical**
$\bar{Z}_1$	0.00895	0.00903	$\bar{Z}_1$	0.03457	0.03447
$\bar{Z}_2$	0.00821	0.00822	$\bar{Z}_2$	0.03098	0.03075
$\bar{Z}_3$	0.00764	0.00759	$\bar{Z}_3$	0.02796	0.02764
$\bar{Z}_4$	0.00754	0.00866	$\bar{Z}_4$	0.02741	0.03284
$\bar{Z}_5$	0.00755	0.00750	$\bar{Z}_5$	0.02752	0.02718
$\bar{Z}_6$	0.00894	0.00902	$\bar{Z}_6$	0.03755	0.03749
$\bar{Z}_7$	0.00821	0.00822	$\bar{Z}_7$	0.03324	0.03302
$\bar{Z}_8$	0.00763	0.00759	$\bar{Z}_8$	0.02929	0.02898
$\bar{Z}_9$	0.00753	0.00865	$\bar{Z}_9$	0.02847	0.03555
$\bar{Z}_{10}$	0.00755	0.00749	$\bar{Z}_{10}$	0.02864	0.02831
$\bar{Z}_{11}$	0.00801	0.00803	$\bar{Z}_{11}$	0.03152	0.03138
$\bar{Z}_{12}$	0.00761	0.00757	$\bar{Z}_{12}$	0.02884	0.02858
$\bar{Z}_{13}$	0.00749	0.00746	$\bar{Z}_{13}$	0.02697	0.02663
$\bar{Z}_{14}$	0.00746	0.00780	$\bar{Z}_{14}$	0.02677	0.03012
$\bar{Z}_{15}$	0.00745	0.00737	$\bar{Z}_{15}$	0.02680	0.02644
$\bar{Z}_{16}$	0.00905	0.00913	$\bar{Z}_{16}$	0.03468	0.03458
$\bar{Z}_{17}$	0.00828	0.00830	$\bar{Z}_{17}$	0.03107	0.03083
$\bar{Z}_{18}$	0.00767	0.00763	$\bar{Z}_{18}$	0.02800	0.02768
$\bar{Z}_{19}$	0.00756	0.00875	$\bar{Z}_{19}$	0.02744	0.03294
$\bar{Z}_{20}$	0.00758	0.00753	$\bar{Z}_{20}$	0.02755	0.02722
$\bar{Z}_{21}$	0.00911	0.00920	$\bar{Z}_{21}$	0.03451	0.03442
$\bar{Z}_{22}$	0.00833	0.00835	$\bar{Z}_{22}$	0.03095	0.03071
$\bar{Z}_{23}$	0.00770	0.00766	$\bar{Z}_{23}$	0.02794	0.02762
$\bar{Z}_{24}$	0.00758	0.00881	$\bar{Z}_{24}$	0.02739	0.03279
$\bar{Z}_{25}$	0.00760	0.00755	$\bar{Z}_{25}$	0.02750	0.02717
$\bar{Z}_{26}$	0.00902	0.00911	$\bar{Z}_{26}$	0.03455	0.03446
$\bar{Z}_{27}$	0.00827	0.00828	$\bar{Z}_{27}$	0.03097	0.03074
$\bar{Z}_{28}$	0.00766	0.00762	$\bar{Z}_{28}$	0.02795	0.02763
$\bar{Z}_{29}$	0.00756	0.00872	$\bar{Z}_{29}$	0.02741	0.03282
$\bar{Z}_{30}$	0.00757	0.00752	$\bar{Z}_{30}$	0.02751	0.02718
$\bar{Z}_{31}$	0.00902	0.00910	$\bar{Z}_{31}$	0.03455	0.03446
$\bar{Z}_{32}$	0.00826	0.00827	$\bar{Z}_{32}$	0.03097	0.03074
$\bar{Z}_{33}$	0.00766	0.00762	$\bar{Z}_{33}$	0.02795	0.02763
$\bar{Z}_{34}$	0.00756	0.00872	$\bar{Z}_{34}$	0.02741	0.03282
$\bar{Z}_{35}$	0.00757	0.00752	$\bar{Z}_{35}$	0.02751	0.02718
$\bar{Z}_{N_1}$	0.00739	0.00736	$\bar{Z}_{N_1}$	0.02672	0.02632
$\bar{Z}_{N_2}$	0.00744	0.00731	$\bar{Z}_{N_2}$	0.02671	0.02630
$\bar{Z}_{N_3}$	0.00737	0.00733	$\bar{Z}_{N_3}$	0.02609	0.02634
$\bar{Z}_{N_4}$	0.00742	0.00735	$\bar{Z}_{N_4}$	0.02673	0.02631
$\bar{Z}_{N_5}$	0.00743	0.00734	$\bar{Z}_{N_5}$	0.02671	0.02624
$\bar{Z}_{N_6}$	0.00740	0.00730	$\bar{Z}_{N_6}$	0.02674	0.02628
$\bar{Z}_{N_7}$	0.00741	0.00732	$\bar{Z}_{N_7}$	0.02675	0.02629

Tables (2.3 - 2.5) report the MSE for the estimators involved in the empirical and theoretical illustration, some major observations are highlighted here.

In **Table (2.3)** all the estimators of proposed class have minimum MSE as compared to generalized class of estimators. $\bar{Z}_{N_4}$ has the least MSE in population -1 while $\bar{Z}_{N_5}$has the least MSE in population-2.

Table (2.4) and **Table (2.5)** reveal that all the estimators of proposed class have minimum MSE as compared to generalized class of estimators. $\bar{Z}_{N_4}$ has the least MSE in population-1 at $P = 0.40$ while $\bar{Z}_{N_5}$ has the least MSE in population-1 at $P = 0.80$. Similarly, $\bar{Z}_{N_3}$ and $\bar{Z}_{N_6}$ have least MSE with the same values of P in **Table (2.5)** for population-2.

According to the simulation results, we observe that in both populations, the new proposals outer perform as compared to adapted ones. In Bar-Lev model, real data set (Pop-2), we observe that by increasing P, MSE also increased. However, MSE results were opposite relative to artificial data set in (Pop-1) on Bar-Lev model. Moreover, new proposals performing out class as compared to adapted ones. Hence, we say, according to tabulated results, new proposals are the best ones for estimating $\bar{Z}$ because, every adapted estimator utilize the more or same supplementary information can be less efficient as compare to proposed robust-regression-type estimators.

Chapter 3

Mean estimation in systematic sampling: robust estimators

3.1 Background

To address the difficulty of accomplishing increasingly precise estimators, one of the most useful techniques is the mindful utilization of auxiliary information. The utilization of auxiliary information can be seen in a significant work of wright (1928) to describe the import levy on margarine and vegetable oil in the United States, where the normal participation in dairy region was considered as auxiliary information to appraise the strong harmony sought after and supply of dairy items. Mean estimation is one of the prime worry in review examining. The mean estimators can be improved by using helper data (see, e.g., Oral and oral, 2011; Abid et al., 2016; Subzar et al., 2019 and Shahzad et al., 2019).

In this chapter, we initially adapt ratio type estimators by replacing traditional OLS regression coefficient with their robust alternatives. After that, we propose robust regression type estimators for the estimation of population mean of the subject variable utilizing the supplementary information under systematic random sampling scheme by eliminating ratio part from robust ratio type estimators. We also obtain the MSE expressions for proposed estimators. The purpose of proposed estimators is to provide efficient estimate of population mean under systematic random sampling in presence of outliers. For this, we perform numerical illustration and find the superior results of proposed robust regression type estimators over adapted ones.

In literature, such as Kadilar and Cingi (2004); Koyuncu (2012) and Shahzad (2016), developed a family of estimators, which contains the supplementary information, when using simple random sampling design. In presence of outliers, these traditional estimators are not efficient so Kadilar et al. (2007) utilized robust technique namely Huber-M method for mean estimation. Further, Zaman and Bulut (2019) constructed some new estimators based on some different robust regression measures. Taking motivation from these studies, we define Kadilar et al. (2007) and Zaman and Bulut (2019) under systematic random sampling scheme. After that, we have also defined a new family of regression estimators based on robust-regression tools.

3.2 Adapted estimators in systematic sampling

In current section, following Zaman and Bulut (2019a) we attempt to characterize a progressively broad class of ratio type-estimators using robust regression techniques to be specific: LAD, LMS, LTS, Huber-M, Hampel-M, Tukey-M and Huber-MM for the estimation of population mean of objective variable Y using supplement, variable X under systematic random sampling design. The most mainstream technique for evaluating the parameters is the least squares (LS). One of the most straightforward vigorous option in contrast to the LS is the least absolute deviations (LAD) technique which was presented in 1757 by Roger Joseph Boscovich. Huber-M strategy created by Huber (1964) which was the subsequent stage towards powerful relapse.

Huber (1973) extended his own work and used this procedure in robust regression. Thus another vigorous robust regression technique created. The primary concern of this technique is to squared residual in Ordinary least square by p, where p is some symmetric $f(x)$. Many researchers extend the idea of Huber (1973) for example. Hampel (1971) named as Hample-M, estimate. Tukey (1977) named as Tukey-M estimate and Yohai (1987) named as Huber-MM estimate.

We likewise use TLS, which is known as Trimmed Least Squares robust regression strategy, presented by Rousseeuw and Yohai (1984). The last robust regression strategy consider for this study is LMS, which is known as "Least Median of Squares" created by Rousseeuw and Leroy (1987). This strategy dependent on minimization of median of residual squares as opposed to mean of residual squares. Kadilar et al. (2007) and Zaman and Bulut (2019a) utilized these tools and developed ratio type estimators in simple random sampling. We are adapting their estimators in systematic random sampling design in upcoming lines.

Taking motivation from Kadilar et al. (2007). We develop the following class of robust ratio type estimator:

$$T_{kc1} = \frac{\bar{y}_s + b_{h-rrc}\left(\bar{X} - \bar{x}_s\right)}{\left(\bar{x}_s\right)}\left(\bar{X}\right), \tag{3.1}$$

$$T_{kc2}=\frac{\bar{y}_s+b_{h-rrc}\left(\bar{X}-\bar{x}_s\right)}{\left(\bar{x}_s+G_c\right)}\left(\bar{X}+G_c\right)\left(\bar{X}\right), \tag{3.2}$$

$$T_{kc3}=\frac{\bar{y}_s+b_{h-rrc}\left(\bar{X}-\bar{x}_s\right)}{\left(\bar{x}_s+G_b\right)}\left(\bar{X}+G_b\right)\left(\bar{X}\right), \tag{3.3}$$

$$T_{kc4}=\frac{\bar{y}_s+b_{h-rrc}\left(\bar{X}-\bar{x}_s\right)}{\left(G_b\bar{x}_s+G_c\right)}\left(G_b\bar{X}+G_c\right)\left(\bar{X}\right), \tag{3.4}$$

$$T_{kc5}=\frac{\bar{y}_s+b_{h-rrc}\left(\bar{X}-\bar{x}_s\right)}{\left(G_c\bar{x}_s+G_b\right)}\left(G_c\bar{X}+G_b\right)\left(\bar{X}\right), \tag{3.5}$$

Let we find MSE of T_{kc1}, Suppose

$$\hat{\gamma}_{kc}=\frac{\bar{y}_s+b_{h-rrc}\left(\bar{X}-\bar{x}_s\right)}{F_{j1}\bar{x}_s+G_{1j1}}.$$

Hence,

$$T_{kc1}=\hat{\gamma}_{kc}\left(F_{j1}\bar{x}_s+G_{1j1}\right).$$

Now we find MSE of T_{kc1} through Taylor series, defined as:

$$h\left(\bar{x}_s,\bar{y}_s\right)=h\left(\bar{X},\bar{Y}\right)+\left[\frac{h\left(\bar{x}_s,\bar{y}_s\right)}{\bar{x}_s}\right]_{\bar{X},\bar{Y}}\left(\bar{x}_s-\bar{Y}\right)+\left[\frac{h\left(\bar{x}_s,\bar{y}_s\right)}{\bar{y}_s}\right]_{\bar{X},\bar{Y}}\left(\bar{y}_s-\bar{Y}\right).$$

Here,

$$h\left(\bar{x}_s,\bar{y}_s\right)=\hat{\gamma}_{kc} \qquad \text{and} \qquad h\left(\bar{X},\bar{Y}\right)=\hat{\gamma}_{kc}$$

$$\hat{\gamma}_{kc}=\gamma_{kc}+\left[\frac{h\left(\bar{x}_s,\bar{y}_s\right)}{\bar{x}_s}\right]_{\bar{X},\bar{Y}}\left(\bar{x}_s-\bar{Y}\right)+\left[\frac{h\left(\bar{x}_s,\bar{y}_s\right)}{\bar{y}_s}\right]_{\bar{X},\bar{Y}}\left(\bar{y}_s-\bar{Y}\right),$$

$$\hat{\gamma}_{kc}-\gamma_{kc}=\left[\frac{h(\bar{x}_s,\bar{y}_s)}{\bar{x}_s}\right]_{\bar{X},\bar{Y}}\left(\bar{x}_s-\bar{Y}\right)+\left[\frac{h(\bar{x}_s,\bar{y}_s)}{\bar{y}_s}\right]_{\bar{X},\bar{Y}}\left(\bar{y}_s-\bar{Y}\right).$$

$$\hat{\gamma}_{kc}-\gamma_{kc}=\left[\frac{\dfrac{\bar{y}_s+b_{h-rrc}\left(\bar{X}-\bar{x}_s\right)}{F_{j1}\bar{x}_s+G_{1j1}}}{\bar{x}_s}\right]_{\bar{X},\bar{Y}}\left(\bar{x}_s-\bar{Y}\right)+\left[\frac{\dfrac{\bar{y}_s+b_{h-rrc}\left(\bar{X}-\bar{x}_s\right)}{F_{j1}\bar{x}_s+G_{1j1}}}{\bar{y}_s}\right]_{\bar{X},\bar{Y}}\left(\bar{y}_s-\bar{Y}\right).$$

Now differentiating both terms w. r. t $\bar{x}_s$ and $\bar{y}_s$ respectively, squaring and then applying expectation. After that putting $\bar{x}_s-\bar{X}$ and $\bar{y}_s-\bar{Y}$ and $b_{h-rrc}=B_{h-rrc}$

$$MSE\left(T_{zb_{j1}}\right)=\left(\frac{1-f}{n}\right)\left[t_yS_y^2+\left(K_{zb_\mu}+B_{lad-rrc}\right)^2t_xS_x^2-2\left(K_{zb_\mu}+B_{lad-rrc}\right)t_xt^*S_{xy}\right],$$

where, $t_x=1+(n-1)r_x$, $t_y=1+(n-1)r_y$, $t^*=\sqrt{\frac{t_y}{t_x}}$. Further $\left(S_y^2,\ S_x^2\right)$ are the unbiased variances of $\left(\bar{X},\ \bar{Y}\right)$ and S_{xy} in representing covariance $\left(r_x,\ r_y\right)$ are the intra-class correlations of $\left(\bar{X},\ \bar{Y}\right)$ respectively. Similarly, the MSE of,

$$MSE\left(T_{kc_2}\right)=\left(\frac{1-f}{n}\right)\left[t_yS_y^2+\left(K_{zc_2}+B_{h-rrc}\right)^2t_xS_x^2-2\left(K_{zc_2}+B_{h-rrc}\right)t_xt^*S_{xy}\right],$$

$$MSE\left(T_{kc_3}\right)=\left(\frac{1-f}{n}\right)\left[t_yS_y^2+\left(K_{zc_3}+B_{h-rrc}\right)^2t_xS_x^2-2\left(K_{zc_3}+B_{h-rrc}\right)t_xt^*S_{xy}\right],$$

$$MSE\left(T_{kc_4}\right)=\left(\frac{1-f}{n}\right)\left[t_yS_y^2+\left(K_{zc_4}+B_{h-rrc}\right)^2t_xS_x^2-2\left(K_{zc_4}+B_{h-rrc}\right)t_xt^*S_{xy}\right],$$

$$MSE\left(T_{kc_5}\right)=\left(\frac{1-f}{n}\right)\left[t_yS_y^2+\left(K_{zc_5}+B_{h-rrc}\right)^2t_xS_x^2-2\left(K_{zc_5}+B_{h-rrc}\right)t_xt^*S_{xy}\right],$$

where

$$K_{kc_4} = \frac{F_{j1}\bar{Y}}{F_{j1}\bar{X} + G_{j1}} \;, \qquad for \quad i = j = 1,2,\dots,5$$

Taking motivation from Zaman and Bulut (2019a) we propose the following class of estimators,

$$T_{zb_{j1}} = \frac{\bar{y}_s + b_{lad-rrc}\left(\bar{X} - \bar{x}_s\right)}{\left(F_{j1}\bar{x}_s + G_{j1}\right)}\left(F_{j1}\bar{X} + G_{j1}\right) \qquad for \quad j = 1,2,\dots,5 \tag{3.7}$$

$$T_{zb_{j2}} = \frac{\bar{y}_s + b_{lad-rrc}\left(\bar{X} - \bar{x}_s\right)}{\left(F_{j2}\bar{x}_s + G_{j2}\right)}\left(F_{j2}\bar{X} + G_{j2}\right) \qquad for \quad j = 1,2,\dots,5 \tag{3.8}$$

$$T_{zb_{j3}} = \frac{\bar{y}_s + b_{lad-rrc}\left(\bar{X} - \bar{x}_s\right)}{\left(F_{j3}\bar{x}_s + G_{j3}\right)}\left(F_{j3}\bar{X} + G_{j3}\right) \qquad for \quad j = 1,2,\dots,5 \tag{3.9}$$

$$T_{zb_{j4}} = \frac{\bar{y}_s + b_{lad-rrc}\left(\bar{X} - \bar{x}_s\right)}{\left(F_{j4}\bar{x}_s + G_{j4}\right)}\left(F_{j4}\bar{X} + G_{j4}\right) \qquad for \quad j = 1,\ 2,\dots,5 \tag{3.10}$$

$$T_{zb_{j5}} = \frac{\bar{y}_s + b_{lad-rrc}\left(\bar{X} - \bar{x}_s\right)}{\left(F_{j5}\bar{x}_s + G_{j5}\right)}\left(F_{j5}\bar{x}_s + G_{j5}\right) \qquad for \quad j = 1,2,\dots,5 \tag{3.11}$$

$$T_{zb_{j6}} = \frac{\bar{y}_s + b_{lad-rrc}\left(\bar{X} - \bar{x}_s\right)}{\left(F_{j6}\bar{x}_s + G_{j6}\right)}\left(F_{j6}\bar{x}_s + G_{j6}\right) \qquad for \quad j = 1,2,\dots,5 \tag{3.12}$$

All the thirty family members of Zaman and Balut (2019a) are provided in **Table (3.1).**

Note that in $T_{Zb_{j1}}$ *to* $T_{Zb_{j6}}$, where, $j = 1,2,3,4,5$:

b_{h-rrc} = Huber regression coefficient,

$b_{lad-rrc}$ = LAD regression coefficient,

$b_{lts-rrc}$ = LTS regression coefficient,

$b_{lms-rrc}$ = LMS regression coefficient,

$b_{hpl-rrc}$ = Hample regression coefficient.

$b_{tky-rrc}$ = Tukey regression coefficient,

$b_{hmm-rrc}$ = Huber-MM regression coefficient,

G_c = Coefficient of variation,

G_b = Coefficient of variation.

The MSE of T_{zb} family as given below:

$$MSE\left(T_{zb_{j1}}\right)=\left(\frac{1-f}{n}\right)\left[t_y S_y^2+\left(K_{zb_{j1}}+B_{lad-rrc}\right)^2 t_x S_x^2-2\left(K_{zb_{j1}}+B_{lad-rrc}\right)t_x t^* S_{xy}\right],$$

$$MSE\left(T_{zb_{j2}}\right)=\left(\frac{1-f}{n}\right)\left[t_y S_y^2+\left(K_{zb_{j2}}+B_{lts-rrc}\right)^2 t_x S_x^2-2\left(K_{zb_{j2}}+B_{lts-rrc}\right)t_x t^* S_{xy}\right],$$

$$MSE\left(T_{zb_{j3}}\right)=\left(\frac{1-f}{n}\right)\left[t_y S_y^2+\left(K_{zb_{j3}}+B_{lms-rrc}\right)^2 t_x S_x^2-2\left(K_{zb_{j3}}+B_{lms-rrc}\right)t_x t^* S_{xy}\right],$$

$$MSE\left(T_{zb_{j4}}\right)=\left(\frac{1-f}{n}\right)\left[t_y S_y^2+\left(K_{zb_{j4}}+B_{hpl-rrc}\right)^2 t_x S_x^2-2\left(K_{zb_{j4}}+B_{hpl-rrc}\right)t_x t^* S_{xy}\right],$$

$$MSE\left(T_{zb_{j5}}\right)=\left(\frac{1-f}{n}\right)\left[t_y S_y^2+\left(K_{zb_{j5}}+B_{tky-rrc}\right)^2 t_x S_x^2-2\left(K_{zb_{j5}}+B_{tky-rrc}\right)t_x t^* S_{xy}\right],$$

$$MSE\left(T_{zb_{j6}}\right)=\left(\frac{1-f}{n}\right)\left[t_y S_y^2+\left(K_{zb_{j6}}+B_{hmm-rrc}\right)^2 t_x S_x^2-2\left(K_{zb_{j6}}+B_{hmm-rrc}\right)t_x t^* S_{xy}\right],$$

3.3 Proposed estimators in systematic sampling

Taking motivation from T_{kc} and T_{sb}, we define the following class of estimators under systematic random sampling.

$$T_{N_1} = \overline{y}_s + b_{h-rcc}(\overline{X} - \overline{x}_s),$$

$$T_{N_2} = \overline{y}_s + b_{lad-rcc}(\overline{X} - \overline{x}_s),$$

$$T_{N_3} = \overline{y}_s + b_{lts-rcc}(\overline{X} - \overline{x}_s),$$

$$T_{N_4} = \overline{y}_s + b_{lms-rcc}(\overline{X} - \overline{x}_s),$$

$$T_{N_5} = \overline{y}_s + b_{hpl-rcc}(\overline{X} - \overline{x}_s),$$

$$T_{N_6} = \overline{y}_s + b_{tky-rcc}(\overline{X} - \overline{x}_s),$$

$$T_{N_7} = \overline{y}_s + b_{hmm-rcc}(\overline{X} - \overline{x}_s).$$

The MSEs of T_N are given below:

$$MSE(T_{N_1}) = \left[V(\overline{y}_s) - 2B_{h-rcc}Cov(\overline{x}_s, \overline{y}_s) + B^2_{h-rcc}V(\overline{x}_s)\right],$$

$$MSE(T_{N_2}) = \left[V(\overline{y}_s) - 2B_{lad-rcc}Cov(\overline{x}_s, \overline{y}_s) + B^2_{lad-rcc}V(\overline{x}_s)\right],$$

$$MSE(T_{N_3}) = \left[V(\overline{y}_s) - 2B_{lts-rcc}Cov(\overline{x}_s, \overline{y}_s) + B^2_{lts-rcc}V(\overline{x}_s)\right],$$

$$MSE(T_{N_4}) = \left[V(\overline{y}_s) - 2B_{lms-rcc}Cov(\overline{x}_s, \overline{y}_s) + B^2_{lms-rcc}V(\overline{x}_s)\right],$$

$$MSE(T_{N_5}) = \left[V(\overline{y}_s) - 2B_{hpl-rcc}Cov(\overline{x}_s, \overline{y}_s) + B^2_{hpl-rcc}V(\overline{x}_s)\right],$$

$$MSE(T_{N_6}) = \left[V(\bar{y}_s) - 2B_{tky-rcc}Cov(\bar{x}_s, \bar{y}_s) + B^2_{tky-rcc}V(\bar{x}_s)\right],$$

$$MSE(T_{N_7}) = \left[V(\bar{y}_s) - 2B_{hmm-rcc}Cov(\bar{x}_s, \bar{y}_s) + B^2_{hmm-rcc}V(\bar{x}_s)\right].$$

Now putting the values of $V(\bar{y}_s)$, $V(\bar{x}_s)$ and $Cov(\bar{x}_s, \bar{y}_s)$, to get finalized MSE formulas as given below:

$$MSE\left(T_{N_1}\right) = \left(\frac{1-f}{n}\right)\left[t_y S_y^2 - 2B_{h-rcc}{}^2 t_x t^* S_{xy} + B^2_{h-rcc} t_x S_x^2\right],$$

$$MSE\left(T_{N_2}\right) = \left(\frac{1-f}{n}\right)\left[t_y S_y^2 - 2B_{lad-rcc}{}^2 t_x t^* S_{xy} + B^2_{lad-rcc} t_x S_x^2\right],$$

$$MSE\left(T_{N_3}\right) = \left(\frac{1-f}{n}\right)\left[t_y S_y^2 - 2B_{lts-rcc}{}^2 t_x t^* S_{xy} + B^2_{lts-rcc} t_x S_x^2\right],$$

$$MSE\left(T_{N_4}\right) = \left(\frac{1-f}{n}\right)\left[t_y S_y^2 - 2B_{lms-rcc}{}^2 t_x t^* S_{xy} + B^2_{lms-rcc} t_x S_x^2\right],$$

$$MSE\left(T_{N_5}\right) = \left(\frac{1-f}{n}\right)\left[t_y S_y^2 - 2B_{hpl-rcc}{}^2 t_x t^* S_{xy} + B^2_{hpl-rcc} t_x S_x^2\right],$$

$$MSE\left(T_{N_6}\right) = \left(\frac{1-f}{n}\right)\left[t_y S_y^2 - 2B_{tky-rcc}{}^2 t_x t^* S_{xy} + B^2_{tky-rcc} t_x S_x^2\right],$$

$$MSE\left(T_{N_7}\right) = \left(\frac{1-f}{n}\right)\left[t_y S_y^2 - 2B_{hmm-rcc}{}^2 t_x t^* S_{xy} + B^2_{hmm-rcc} t_x S_x^2\right].$$

3.4 Efficiency comparison

In current section, one real (Pop-1) population and one artificial (pop-2) population is considered for efficiency comparison.

Pop-1 is taken from Murthy (1967), where,

X = Strip length

Y = Volume of timber

The size of population is $N = 176$. Note that Murthy (1967) also provided some values of intra-class correlation i.e. $r_x = r_y = r_w$ (say) with respect to different sample size as follows:

$r_w = -0.1510$, $r_w = -0.1106$, $r_w = -0.0522$, $r_w = -0.0435$

So in this research we compare the estimators in light of all the above mentioned values.

Pop-2 contains simulation study. Where a random variable $X \sim G(2.66, 3.88)$ where, G denotes Gamma distribution and random variable Y_i is defined as $Y_i = h_1 + R_1 X_i + \in X_i^q$. Here we assume that the variables $q = 1.6$, $h_1 = 5$, $R_1 = 2$ and has normal distribution. The size of population is N=1000. Here, Systematic random sampling is considered for drawing samples with $n = 150$. The procedure of drawing samples replicated Q' =1000 times. After that empirical MSE's of existing estimators are calculated as follows:

$$.MSE = \frac{\sum_{i=1}^{Q}(\bar{y}-\bar{Y})^2}{Q}.$$

All results of MSE related to Pop-1 and Pop-2 available in **Tables (3.2 - 3.11)**. These results confirm the superiority of proposed estimators over existing ones.

3.5 Summary of the chapter

In this chapter, we propose two classes of estimators whenever data is contaminated with outliers under systematic random sampling scheme. At first we develop ratio type estimators based on robust regression tools. Secondly, we develop regression type estimator T_N based on regression tools. It is seen that T_N estimators performing better as compare to ratio estimators.

Table 3.1: Zaman and Bulut (2019a) adapted estimators.

Estimators	b_i	F_{j1}	G_{j1}
T_{zb11}	$b_{lad-rcc}$	1	0
T_{zb21}	$b_{lad-rcc}$	1	G_b
T_{zb31}	$b_{lad-rcc}$	1	G_C
T_{zb41}	$b_{lad-rcc}$	G_b	G_b
T_{zb51}	$b_{lad-rcc}$	G_C	G_C
T_{zb12}	$b_{lms-rcc}$	1	0
T_{zb22}	$b_{lms-rcc}$	1	G_b
T_{zb32}	$b_{lms-rcc}$	1	G_C
T_{zb42}	$b_{lms-rcc}$	G_b	G_b
T_{zb52}	$b_{lms-rcc}$	G_C	G_C
T_{zb13}	$b_{lts-rcc}$	1	0
T_{zb23}	$b_{lts-rcc}$	1	G_b
T_{zb33}	$b_{lts-rcc}$	1	G_C
T_{zb43}	$b_{lts-rcc}$	G_b	G_b
T_{zb53}	$b_{lts-rcc}$	G_C	G_C
T_{zb14}	$b_{hpl-rcc}$	1	0
T_{zb24}	$b_{hpl-rcc}$	1	G_b
T_{zb34}	$b_{hpl-rcc}$	1	G_C
T_{zb44}	$b_{hpl-rcc}$	G_b	G_b
T_{zb54}	$b_{hpl-rcc}$	G_C	G_C
T_{zb15}	$b_{tky-rcc}$	1	0
T_{zb25}	$b_{tky-rcc}$	1	G_b
T_{zb35}	$b_{tky-rcc}$	1	G_C
T_{zb45}	$b_{tky-rcc}$	G_b	G_b
T_{zb55}	$b_{hmm-rcc}$	G_C	G_C
T_{zb16}	$b_{hmm-rcc}$	1	0
T_{zb26}	$b_{hmm-rcc}$	1	G_b
T_{zb36}	$b_{hmm-rcc}$	1	G_C
T_{zb46}	$b_{hmm-rcc}$	G_b	G_b
T_{zb56}	$b_{lad-rcc}$	G_C	G_C

Table 3.2: MSE of adapted and proposed estimators with $n = 4$ and $r_w = -0.1510$ in Pop-1.

H-M	T_{kc_1}	T_{kc_2}	T_{kc_3}	T_{kc_4}	T_{kc_5}	T_{N_1}
	2613.4910	2406.2492	1646.0025	949.7094	1200.6353	783.1179
LAD	$T_{zb_{11}}$	$T_{zb_{21}}$	$T_{zb_{31}}$	$T_{zb_{41}}$	$T_{zb_{51}}$	T_{N_2}
	2558.4698	2354.4593	1608.3463	933.3642	1174.5552	784.7374
LTS	$T_{zb_{12}}$	$T_{zb_{22}}$	$T_{zb_{32}}$	$T_{zb_{42}}$	$T_{zb_{52}}$	T_{N_3}
	2747.9525	2533.0126	1739.0946	992.0309	1266.1487	782.6410
LMS	$T_{zb_{13}}$	$T_{zb_{23}}$	$T_{zb_{33}}$	$T_{zb_{43}}$	$T_{zb_{53}}$	T_{N_4}
	2709.7629	2496.9819	1712.5068	979.6814	1247.2952	782.2941
HPL	$T_{zb_{14}}$	$T_{zb_{24}}$	$T_{zb_{34}}$	$T_{zb_{44}}$	$T_{zb_{54}}$	T_{N_5}
	2653.2651	2443.7175	1673.3857	961.8858	1219.7581	782.4754
TKY	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	2617.3448	2409.8785	1648.6499	950.8762	1202.4784	783.0366
H-MM	$T_{zb_{16}}$	$T_{zb_{26}}$	$T_{zb_{36}}$	$T_{zb_{46}}$	$T_{zb_{56}}$	T_{N_7}
	2610.7097	2403.6301	1644.0927	948.8691	1199.3064	783.1792

Table 3.3: PRE of adapted and proposed estimators with $n = 4$ and $r_w = -0.1510$ in Pop-1.

H-M	T_{kc_1}	T_{kc_2}	T_{kc_3}	T_{kc_4}	T_{kc_5}	T_{N_1}
	124.0161	134.6972	196.9103	341.2780	269.9529	413.8776
LAD	$T_{zb_{11}}$	$T_{zb_{21}}$	$T_{zb_{31}}$	$T_{zb_{41}}$	$T_{zb_{51}}$	T_{N_2}
	126.6831	137.6600	201.5206	347.2545	275.9469	413.0234
LTS	$T_{zb_{12}}$	$T_{zb_{22}}$	$T_{zb_{32}}$	$T_{zb_{42}}$	$T_{zb_{52}}$	T_{N_3}
	117.9478	127.9563	186.3699	326.7186	255.9849	414.1298
LMS	$T_{zb_{13}}$	$T_{zb_{23}}$	$T_{zb_{33}}$	$T_{zb_{43}}$	$T_{zb_{53}}$	T_{N_4}
	119.6101	129.8027	189.2634	330.8371	259.8542	414.3134
HPL	$T_{zb_{14}}$	$T_{zb_{24}}$	$T_{zb_{34}}$	$T_{zb_{44}}$	$T_{zb_{54}}$	T_{N_5}
	122.1570	132.6319	193.6881	336.9578	265.7207	414.2174
TKY	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	123.8335	134.4943	196.5942	340.8592	269.5391	413.9205
H-MM	$T_{zb_{16}}$	$T_{zb_{26}}$	$T_{zb_{36}}$	$T_{zb_{46}}$	$T_{zb_{56}}$	T_{N_7}
	124.1482	134.8439	197.1391	341.5802	270.2520	413.8452

Table 3.4: MSE of adapted and proposed estimators with $n = 8$ and $r_w = -0.1106$ in Pop-1.

H-M	T_{kc_1}	T_{kc_2}	T_{kc_3}	T_{kc_4}	T_{kc_5}	T_{N_1}
	526.8760	485.0964	331.8318	191.4601	242.0464	157.8755
LAD	$T_{zb_{11}}$	$T_{zb_{21}}$	$T_{zb_{31}}$	$T_{zb_{41}}$	$T_{zb_{51}}$	T_{N_2}
	515.7839	474.6556	324.2403	188.1649	236.7887	158.2019
LTS	$T_{zb_{12}}$	$T_{zb_{22}}$	$T_{zb_{32}}$	$T_{zb_{42}}$	$T_{zb_{52}}$	T_{N_3}
	553.9833	510.6517	350.5990	199.9920	255.2538	157.7793
LMS	$T_{zb_{13}}$	$T_{zb_{23}}$	$T_{zb_{33}}$	$T_{zb_{43}}$	$T_{zb_{53}}$	T_{N_4}
	548.6821	505.6498	346.9058	198.2723	252.6326	157.7231
HPL	$T_{zb_{14}}$	$T_{zb_{24}}$	$T_{zb_{34}}$	$T_{zb_{44}}$	$T_{zb_{54}}$	T_{N_5}
	534.8944	492.6500	337.3522	193.9148	245.9015	157.7459
TKY	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	527.6530	485.8281	332.3655	191.6953	242.4179	157.8591
H-MM	$T_{zb_{16}}$	$T_{zb_{26}}$	$T_{zb_{36}}$	$T_{zb_{46}}$	$T_{zb_{56}}$	T_{N_7}
	526.3156	484.5686	331.4469	191.2907	241.7786	157.8878

Table 3.5: PRE of adapted and proposed estimators with $n = 8$ and $r_w = -0.1106$ in Pop-1.

H-M	T_{kc_1}	T_{kc_2}	T_{kc_3}	T_{kc_4}	T_{kc_5}	T_{N_1}
	124.0161	134.6972	196.9103	341.2780	269.9529	413.8776
LAD	$T_{zb_{11}}$	$T_{zb_{21}}$	$T_{zb_{21}}$	$T_{zb_{41}}$	$T_{zb_{51}}$	T_{N_2}
	126.6831	137.6600	201.5206	347.2545	275.9469	413.0234
LTS	$T_{zb_{12}}$	$T_{zb_{22}}$	$T_{zb_{12}}$	$T_{zb_{42}}$	$T_{zb_{52}}$	T_{N_3}
	117.9478	127.9563	186.3699	326.7186	255.9849	414.1298
LMS	$T_{zb_{13}}$	$T_{zb_{23}}$	$T_{zb_{33}}$	$T_{zb_{43}}$	$T_{zb_{53}}$	T_{N_4}
	119.0874	129.2221	188.3540	329.5524	258.6408	414.2773
HPL	$T_{zb_{14}}$	$T_{zb_{24}}$	$T_{zb_{34}}$	$T_{zb_{44}}$	$T_{zb_{54}}$	T_{N_5}
	122.1570	132.6319	193.6881	336.9578	265.7207	414.2174
TKY	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	123.8335	134.4943	196.5942	340.8592	269.5391	413.9205
H-MM	$T_{zb_{16}}$	$T_{zb_{26}}$	$T_{zb_{36}}$	$T_{zb_{46}}$	$T_{zb_{56}}$	T_{N_7}
	124.1482	134.8439	197.1390	341.5801	270.2518	413.8452

Table 3.6: MSE of adapted and proposed estimators with $n = 16$ and $r_w = -0.0522$ in Pop-1.

H-M	T_{kc_1}	T_{kc_2}	T_{kc_3}	T_{kc_4}	T_{kc_5}	T_{N_1}
	241.11540	221.99570	151.85687	87.61827	110.76819	72.24888
LAD	$T_{zb_{11}}$	$T_{zb_{21}}$	$T_{zb_{31}}$	$T_{zb_{41}}$	$T_{zb_{51}}$	T_{N_2}
	236.03926	217.21767	148.38278	86.11029	108.36209	72.39829
LTS	$T_{zb_{12}}$	$T_{zb_{22}}$	$T_{zb_{32}}$	$T_{zb_{42}}$	$T_{zb_{52}}$	T_{N_3}
	254.40678	234.52707	161.06400	91.81300	117.25257	72.21828
LMS	$T_{zb_{13}}$	$T_{zb_{23}}$	$T_{zb_{33}}$	$T_{zb_{43}}$	$T_{zb_{53}}$	T_{N_4}
	251.75459	232.02425	159.21457	90.94878	115.93831	72.18456
HPL	$T_{zb_{14}}$	$T_{zb_{24}}$	$T_{zb_{34}}$	$T_{zb_{44}}$	$T_{zb_{54}}$	T_{N_5}
	244.78488	225.45245	154.38319	88.74164	112.53242	72.18960
TKY	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	241.47095	222.33053	152.10111	87.72592	110.93823	72.24138
H-MM	$T_{zb_{16}}$	$T_{zb_{26}}$	$T_{zb_{36}}$	$T_{zb_{46}}$	$T_{zb_{56}}$	T_{N_7}
	240.85892	221.75416	151.68075	87.54078	110.64565	72.25453

Table 3.7: PRE of adapted and proposed estimators with $n = 16$ and $r_w = -0.0522$ in Pop-1.

H-M	T_{kc_1}	T_{kc_2}	T_{kc_3}	T_{kc_4}	T_{kc_5}	T_{N_1}
	124.0161	134.6972	196.9103	341.2780	269.9529	413.8776
LAD	$T_{zb_{11}}$	$T_{zb_{21}}$	$T_{zb_{31}}$	$T_{zb_{41}}$	$T_{zb_{51}}$	T_{N_2}
	126.6831	137.6600	201.5206	347.2545	275.9469	413.0234
LTS	$T_{zb_{12}}$	$T_{zb_{22}}$	$T_{zb_{32}}$	$T_{zb_{42}}$	$T_{zb_{52}}$	T_{N_3}
	117.5369	127.5000	185.6541	325.6858	255.0237	414.0529
LMS	$T_{zb_{13}}$	$T_{zb_{23}}$	$T_{zb_{33}}$	$T_{zb_{43}}$	$T_{zb_{53}}$	T_{N_4}
	118.7752	128.8753	187.8106	328.7805	257.9146	414.2463
HPL	$T_{zb_{14}}$	$T_{zb_{24}}$	$T_{zb_{34}}$	$T_{zb_{44}}$	$T_{zb_{54}}$	T_{N_5}
	122.1570	132.6319	193.6881	336.9578	265.7207	414.2174
TKY	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	123.8335	134.4943	196.5942	340.8592	269.5391	413.9205
H-MM	$T_{zb_{16}}$	$T_{zb_{26}}$	$T_{zb_{36}}$	$T_{zb_{46}}$	$T_{zb_{56}}$	T_{N_7}
	124.1482	134.8439	197.1390	341.5801	270.2518	413.8452

Table 3.8: MSE of adapted and proposed estimators with $n = 22$ and $r_w = -0.0435$ in Pop-1.

H-M	T_{kc_1}	T_{kc_2}	T_{kc_3}	T_{kc_4}	T_{kc_5}	T_{N_1}
	67.27898	61.94396	42.37297	24.44832	30.90790	20.15977
LAD	$T_{zb_{11}}$	$T_{zb_{21}}$	$T_{zb_{31}}$	$T_{zb_{41}}$	$T_{zb_{51}}$	T_{N_2}
	65.86257	60.61074	41.40358	24.02755	30.23652	20.20146
LTS	$T_{zb_{12}}$	$T_{zb_{22}}$	$T_{zb_{32}}$	$T_{zb_{42}}$	$T_{zb_{52}}$	T_{N_3}
	71.29765	65.73320	45.15867	25.72087	32.87165	20.15675
LMS	$T_{zb_{13}}$	$T_{zb_{23}}$	$T_{zb_{33}}$	$T_{zb_{43}}$	$T_{zb_{53}}$	T_{N_4}
	69.75730	64.27969	44.08498	25.21989	32.10906	20.13856
HPL	$T_{zb_{14}}$	$T_{zb_{24}}$	$T_{zb_{34}}$	$T_{zb_{44}}$	$T_{zb_{54}}$	T_{N_5}
	68.30288	62.90851	43.07789	24.76178	31.40018	20.14323
TKY	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	67.37818	62.03739	42.44112	24.47836	30.95535	20.15767
H-MM	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	67.20739	61.87655	42.32381	24.42670	30.87370	20.16135

Table 3.9: PRE of adapted and proposed estimators with $n = 22$ and $r_w = -0.0435$ in Pop-1.

H-M	T_{kc_1}	T_{kc_2}	T_{kc_3}	T_{kc_4}	c	T_{N_1}
	124.0161	134.6972	196.9103	341.2780	269.9529	413.8776
LAD	$T_{zb_{11}}$	$T_{zb_{21}}$	$T_{zb_{31}}$	$T_{zb_{41}}$	$T_{zb_{51}}$	T_{N_2}
	126.6831	137.6600	201.5206	347.2545	275.9469	413.0234
LTS	$T_{zb_{12}}$	$T_{zb_{22}}$	$T_{zb_{32}}$	$T_{zb_{42}}$	$T_{zb_{52}}$	T_{N_3}
	117.0259	126.9324	184.7635	324.3932	253.8259	413.9395
LMS	$T_{zb_{13}}$	$T_{zb_{23}}$	$T_{zb_{33}}$	$T_{zb_{43}}$	$T_{zb_{53}}$	T_{N_4}
	119.6101	129.8027	189.2634	330.8371	259.8542	414.3134
HPL	$T_{zb_{14}}$	$T_{zb_{24}}$	$T_{zb_{34}}$	$T_{zb_{44}}$	$T_{zb_{54}}$	T_{N_5}
	122.1570	132.6319	193.6881	336.9578	265.7207	414.2174
TKY	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	123.8335	134.4943	196.5942	340.8592	269.5391	413.9205
H-MM	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	124.1482	134.8439	197.1390	341.5802	270.2519	413.8452

Table 3.10: MSE of adapted and proposed estimators with simulation (Pop-2).

H-M	T_{kc_1}	T_{kc_2}	T_{kc_3}	T_{kc_4}	T_{kc_5}	T_{N_1}
	0.89621	0.26751	0.05208	0.06416	0.04551	0.04144
LAD	$T_{zb_{11}}$	$T_{zb_{21}}$	$T_{zb_{31}}$	$T_{zb_{41}}$	$T_{zb_{51}}$	T_{N_2}
	0.90294	0.27098	0.05284	0.06527	0.04599	0.04139
LTS	$T_{zb_{12}}$	$T_{zb_{22}}$	$T_{zb_{32}}$	$T_{zb_{42}}$	$T_{zb_{52}}$	T_{N_3}
	0.93925	0.28990	0.05737	0.07163	0.04899	0.04154
LMS	$T_{zb_{13}}$	$T_{zb_{23}}$	$T_{zb_{33}}$	$T_{zb_{43}}$	$T_{zb_{53}}$	T_{N_4}
	0.90717	0.27316	0.05334	0.06598	0.04630	0.04137
HPL	$T_{zb_{14}}$	$T_{zb_{24}}$	$T_{zb_{34}}$	$T_{zb_{44}}$	$T_{zb_{54}}$	T_{N_5}
	0.88715	0.26286	0.05109	0.06270	0.04490	0.04156
TKY	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	0.88904	0.26383	0.05129	0.06300	0.04502	0.04153
H-MM	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	0.88950	0.26406	0.05134	0.06307	0.04505	0.04153

Table 3.11: MSE of adapted and proposed estimators with simulation (Pop-2).

H-M	T_{kc_1}	T_{kc_2}	T_{kc_3}	T_{kc_4}	T_{kc_5}	T_{N_1}
	9.55632	32.01597	164.45267	133.49513	188.20856	206.64838
LAD	$T_{zb_{11}}$	$T_{zb_{21}}$	$T_{zb_{31}}$	$T_{zb_{41}}$	$T_{zb_{51}}$	T_{N_2}
	9.48507	31.60608	162.06970	131.22365	186.23868	206.92378
LTS	$T_{zb_{12}}$	$T_{zb_{22}}$	$T_{zb_{32}}$	$T_{zb_{42}}$	$T_{zb_{52}}$	T_{N_3}
	9.11847	29.54314	149.29140	119.57253	174.80349	206.17127
LMS	$T_{zb_{13}}$	$T_{zb_{23}}$	$T_{zb_{33}}$	$T_{zb_{43}}$	$T_{zb_{53}}$	T_{N_4}
	9.44086	31.35321	160.57213	129.81399	184.97252	207.02929
HPL	$T_{zb_{14}}$	$T_{zb_{24}}$	$T_{zb_{34}}$	$T_{zb_{44}}$	$T_{zb_{54}}$	T_{N_5}
	9.65394	32.58224	167.65079	136.60322	190.76093	206.06869
TKY	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	9.63341	32.46272	166.98505	135.95025	190.23857	206.20941
H-MM	$T_{zb_{15}}$	$T_{zb_{25}}$	$T_{zb_{35}}$	$T_{zb_{45}}$	$T_{zb_{55}}$	T_{N_6}
	9.62839	32.43348	166.82146	135.79030	190.10948	206.24236

Chapter 4

Robust-regression-type estimators: simple and two-stage sampling for mean

4.1 Background

Nowadays, a widely utilized phrase that we are living in the age of information. Utilizing this phrase, we are not just featuring the volume and speed of existing information yet in addition underlining the need of its exact stream. The later part of the above comprehension is legitimately connected with the true intention of the urge of gathering information. The intention is to empower ourselves of absolutely profiling our environment and in this way supporting the optimal decision making process. In fulfilling the need of multidisciplinary request interlocking government issues, business basic leadership, clinical examinations and mental profiling and so on, it is of nothing unexpected if the sampling theory and method remains at the core of applied research literature.

One of the most significant goal of practices in sampling stays with the estimation of mean of study variable. To meet the challenge of achieving more precise estimate of population mean, ratio method of estimation is the highly praised way utilizing supplementary information. Laplace in eighteenth century, as an early client/user of supplementary information in the estimation of total population of France, gave the method of utilization of supplementary information in an efficient way. Specifically, he referenced, "The register of births, which are kept with care in order to assure the condition of the citizens, can serve to determine the population of great empire without resorting a census of its inhabitants. Other than that, it is essential to know the ratio of the population to birth in an annual, see, e.g., Lohr (1999).

Mean is the most basic center of gravity of the data. Ratio and regression methods are significant tools for the estimation of the population mean. However, the mean estimation through ratio and regression based estimators are not suitable when outliers exist in data. Zaman and Bulut (2019a) provided the solution of this issue by utilizing some robust regression tools and develop a class of ratio type estimators under simple random sampling scheme. Recently, Zaman (2019), have suggested by extending the work of Zaman and Bulut (2019a) with a new class of ratio-type estimators.

This chapter proposed a new family of robust regression type estimators applying robust regression tools (LAD, LMS, LTS, Huber-M, Hampel-M, Tukey-M, and Huber-MM). The class is subsequently extended for the situation of two stage sampling where mean of the study variable is not available at first stage. So, we also developed reviewed and suggested some new estimators under two stage sampling scheme. It is worth mentioning that we consider two cases under two stage sampling scheme: (a) when second stage sample depends upon first stage sample and, (b) when second stage sample is independent of first stage sample. The mean square expressions of the proposed estimators have been determined through Taylor series method. A real life application and the simulation study are also provided to assess existing and proposed estimators. In the light of numerical results, someone can see that proposed findings are more efficient then existing techniques.

A huge literature is available about ratio and regression-type estimators. For example, Oral and Oral (2011); Koyuncu (2012); Abid et al. (2016a, 2016b); Shahzad el al. (2018); Hanif and Shahzad (2019); Bulut and Zaman (2019); Naz et al. (2019) and Irfan et al. (2019) have suggested a family of estimators, that contain auxiliary information under simple random sampling design. For more about ratio estimators see, Jemain et al. (2008); Al-Omari et al. (2008); Al-Omari et al. (2009); Al-Omari and Jaber (2010); Al-Omari (2012); Al-Omari and Bouza (2015); Bouza et al. (2017) and Al-Omari and Al-Nasser (2018). For positive correlation, the ratio estimators perform better for population mean estimation. For negative correlation, the product estimator is better for the estimation of population mean. The conventional regression estimator solves the issue related to the sign (positive/negative) of correlation and provides better results as compared to the ratio and product type estimators.

Note that conventional regression estimators based on conventional regression coefficient, i.e. known as Ordinary Least Square (OLS) regression coefficient. For example, OLS, but they are inefficient when data contain outliers. To handle this problem, Kadilar et al. (2007) incorporated Huber-M robust regression technique instead of OLS.

After that, Zaman and Bulut (2019a) extended the idea of Kadilar et al. (2007) and developed a class of ratio type estimators, utilizing some other robust regression tools namely: LAD (Least Absolute Deviations), LMS (Least Median of Squares), LTS (Least Trimmed Squares), Hampel-M, Tukey-M and Huber-MM. The basic purposes of LAD and LMS are to minimize error by incorporating absolute residual and squared median residual, respectively. The squared errors are arranged in LTS method and OLS is run by utilizing observations based on the first (smallest) z errors.

The purpose of M- estimation is to minimize the q functions with satisfying the necessary assumption, see, e.g., Zaman and Bulut (2019). There are many q functions are available in literature see, Huber (1964, 1973); Hampel (1971) and Tukey (1977).

Finally, Yohai (1987) presented MM robust regression tool which has high statistical efficiency and breakdown point. For more knowledge about these robust regression tools, interested readers may refer to Zaman and Bulut (2018) and Ali et al. (2021). Moreover, Zaman and Bulut (2019) defined another class utilizing robust regression estimates under stratified random sampling scheme. Zaman (2019) developed another class of estimators in the same context and achieved the results equivalent to traditional regression estimator. So in this research, taking inspiration from Zaman and Bulut (2018) and Zaman (2019), we introduce a new and improved class of robust-regression-type estimators for the mean estimation when response/study variable contaminated by outliers.

Outliers are those observations that misleading someone to a wrong track. When these are present in data, the mean estimation gives inappropriate results. Other than that, the mean estimation is the most useful choices for estimation purposes. With outliers, the results adding some wrong information. Hence, for population mean *i.e.* $(\bar{Y})$, based on OLS may indicate weak performance. Kadilar et al. (2007) and Zaman and Bulut (2019a) provided the solution of this issue by incorporating robust regression coefficients in this context. Robust regression is used when OLS assumptions are violated. In such circumstances, robust-regression tools such as, LAD, LMS, LTS, Huber-M, Hampel-M, Tukey-M, Huber-MM provide better results because they give less weight to outliers. Zaman and Bulut (2019a) introduced the following class of estimators utilizing robust regression tools for the estimation of mean as given by,

$$\bar{y}_{zb_i} = \frac{\bar{y} + b_{(i)}\left(\bar{X} - \bar{x}\right)}{\left(c\bar{x} + d\right)}\left(c\bar{X} + d\right), \qquad for \;\; i = 1,2,\ldots,35 \qquad (4.1)$$

Where, $(\bar{X},\bar{Y})$ and $(\bar{x},\bar{y})$ are population and sample means, respectively after using simple random sample to select a sample size n. The variances of these unbiased sample means, $(\bar{x},\bar{y})$ are $V(\bar{x})=\theta S_x^2$, and $V(\bar{y})=\theta S_y^2$. Further, c and d take the values from (0, 1) or any given population information, *i.e.* C_x, the coefficient of variation of X, $\beta_2(x)$, the coefficient of kurtosis of X, and $b_{(i)}$ are the robust regression coefficients. The family members of $\bar{y}_{zb_i}$ are provided in **Table (4.1).**

MSE of Zaman and Bulut (2019a) family of estimators is given below

$$MSE(\bar{y}_{zb_i})=\theta\left[S_y^2+g_i^2S_x^2+2B_ig_iS_x^2+B_i^2S_x^2-2g_iS_{xy}-2B_iS_{xy}\right], \qquad \text{for, } i = 1,2,\dots,35 \tag{4.2}$$

Where, $g_i=\dfrac{c\bar{Y}}{c\bar{X}+d}$ and, $\theta=\left(\dfrac{1-f}{n}\right)$. Further, S_y^2 and S_x^2 are the unbiased variances of Y and X, respectively. Note that $\bar{y}_{zb_{16}}-\bar{y}_{zb_{20}}$ belongs to Kadilar et al., (2007) in **Table (4.1).** Zaman (2019) introduced another class of estimators utilizing robust regression tools for the estimation of mean as follows:

$$\bar{y}_{z_i}=k\frac{\bar{y}+b_{(i)}(\bar{X}-\bar{x})}{(c\bar{x}+d)}(c\bar{X}+d)+(1-k)\frac{\bar{y}+b_{(i)}(\bar{X}-\bar{x})}{(c\bar{x}+d)}(c\bar{X}+d),$$

where, K is a constant such that it provides the minimum $MSE(\bar{y}_{z_i})$, the MSE of $\bar{y}_{z_i}$ is as follows:

$$MSE(\bar{y}_{z_i})=\theta\left[S_y^2-2\delta S_{xy}+\delta^2S_x^2\right]. \tag{4.3}$$

where, $\delta=\left[k\left(B_{(i)}+g_1\right)+(1-k)\left(B_{(i)}+g_i\right)\right]$

Zaman (2019) $(\delta = B)$ in above MSE expression, and get minimum MSE of $\bar{y}_{z_i}$ as follows:

$$MSE(\bar{y}_{z_i}) = \theta S_y^2 (1-\rho^2), \tag{4.4}$$

which, is the MSE of traditional regression estimator, i.e. $\bar{y}_{reg} = \bar{y} + b_{(i)}(\bar{X} - \bar{x})$. The rest of the chapter is constructed as follows: In Section 4.2, we have proposed a new class of robust-regression-type estimators. The theoretical mean squared error (MSE) of proposed class is also derived. Section 4.3 has been dedicated to two stage sampling scheme. We also provided some proposed estimators by using two-stage sampling design with their theoretical MSE expressions in Section 4.3.2. Results and discussion are provided in Section 4.4. The manuscript is ended with some concluding remarks in Section 4.5.

4.2 Proposed class of robust-regression-type-estimators

Taking motivation from Zaman (2019) and Zaman and Bulut (2019a), we propose the following class of estimators as given below:

$$\bar{y}_{N_i} = k_1 \left\{ \bar{y} + b_{(i)}(\bar{X} - \bar{x}) \right\} + k_2 (\bar{X} - \bar{x}), \qquad for \ \ i = 1,2,\ldots,7 \tag{4.5}$$

where, k_1 and k_2 are real constants. Further $\bar{x}, \bar{y}$ and $b_{(i)}$ have their usual meanings as defined in Section 4.1. The family members of proposed available in **Table (4.2).**

To obtain the MSE of equation (4.2), let us define $\bar{y} = (1+\eta_0)\bar{Y}$ and $\bar{x} = (1+\eta_1)\bar{X}$. Utilizing these notations $\eta_i \, (i = 0,1)$, we can write:

$$E(\eta_0) = E(\eta_1) = 0, \ E(\eta_0^2) = \theta C_y^2, E(\eta_1^2) = \theta C_x^2 \text{ and } E(\eta_0\eta_1) = \theta C_{yx}.$$

Now expanding $\bar{y}_{N_t}$ in terms of η_0 and η_1 as:

$$\bar{y}_{N_t} = k_1\bar{Y}\left\{1+\eta_0 - R'b_{(t)}\eta_1\right\} - k_2\bar{X}_{\eta_1}.$$

$$\bar{y}_{N_t} - \bar{Y} = k_1\bar{Y}\left\{1+\eta_0 - R'b_{(t)}\eta_1\right\} - k_2\bar{X}_{\eta_1} - \bar{Y}. \tag{4.6}$$

By taking square of equation (4.6), ignoring higher order terms and applying expectation, the MSE of $\bar{y}_{N_t}$ is given below:

$$MSE\left(\bar{y}_{N_t}\right) = \bar{Y}^2 + k_1^2\Phi_{AN} + k_2^2\Phi_{BN} + 2k_1k_2\Phi_{CN} - 2k_1\Phi_{DN}, \tag{4.7}$$

where,

$$\begin{aligned}
\Phi_{AN} &= \bar{Y}^2\left[1+\theta\left\{C_y^2 + R'b_{(t)}\left(R'b_{(t)}C_x^2 - 2C_{yx}\right)\right\}\right],\\
\Phi_{BN} &= \theta\bar{X}^2C_x^2,\\
\Phi_{CN} &= \theta\bar{X}\bar{Y}\left[R'b_{(t)}C_x^2 - 2C_{yx}\right],\\
\Phi_{DN} &= \bar{Y}^2,\\
R' &= \frac{\bar{X}}{\bar{Y}}.
\end{aligned}$$

The MSE is minimized when,

$$k_1^{opt} = \left[\frac{\Phi_{BN}\Phi_{DN}}{\Phi_{AN}\Phi_{BN} - \Phi_{CN}^2}\right],$$

$$k_2^{opt} = \left[\frac{\Phi_{CN}\Phi_{DN}}{\Phi_{AN}\Phi_{BN} - \Phi_{CN}^2}\right],$$

By substituting k_1^{opt} and k_2^{opt} in equation (4.7), we get minimum MSE of $\bar{y}_{N_t}$ as given below,

$$MSE_{\min}\left(\bar{y}_{N_t}\right)=\left[\bar{Y}^2-\frac{\Phi_{BN}\Phi_{DN}^2}{\Phi_{AN}\Phi_{BN}-\Phi_{CN}^2}\right], \tag{4.8}$$

Remarks

- By replacing ($k_1=1$, $b_{(t)}=0$, $k_2=0$), $\bar{y}_{N_t}$ becomes unbiased mean estimator.
- By replacing ($k_1=1$, $b_{(t)}=0$, $k_2=0$), $\bar{y}_{N_t}$ becomes regression estimator, and will be equally important as $\bar{y}_{z_t}$ or $\bar{y}_{reg}$.
- In light of above two points, we can say that $\bar{y}$, $\bar{y}_{reg}$, and $\bar{y}_{z_t}$, are the special cases of $\bar{y}_{N_t}$.

4.3 Two stage sampling scheme

At the point whenever the required information about population mean of auxiliary variable isn't accessible, one can utilize the two-stage sampling plan in acquiring the improved estimator as opposed to the past ones. Neyman (1938) was the first one, who suggested to estimating the parameters for population in two-stage sampling design. This design is financially understandable and simpler too. Also, it is utilized to get the information through auxiliary variable efficiently by selecting a greater sample from the initial or in stage one and a suitable size (comparatively small sample as compare to first stage) at stage two. Sukhatme (1962) utilized two-stage inspecting devise a general class of ratio-type estimators. For more details about two- stage sampling, interested readers may refer to Cochran (1977).

With two-stage sampling design, first we select a sample of size n_1 at stage-one by using SRSWOR. After this select another sample of size n_2 from the selected sample.

It is worth mentioning that we are considering two cases for second stage sample as follows:

Case I: The second stage sample of size, n_2 is a part of the first stage sample of size n_1.

Case II: The second stage sample, n_2 is independent of the first stage sample i.e. n_1.

For more details about these cases, interested readers may refer to Zaman and Kadilar (2019).

4.3.1 Adapted estimators: two stage sampling design

In this section, we have adapted the family of estimators of Kadilar et al. (2007) and Zaman and Bulut (2019a) for two-stage sampling plan as:

$$\bar{y}'_{zb_i} = \frac{\bar{y}_2 + b_{(i)}\left(\bar{x}_1 - \bar{x}_2\right)}{\left(c\bar{x}_2 + d\right)}\left(c\bar{x}_1 + d\right), \qquad for \;\; i = 1, 2, \ldots, 35 \qquad (4.9)$$

where, $\left(\bar{x}_2, \bar{y}_2\right)$ denoting the means at stage-two where $\bar{x}_1$ at stage-one. Whereas, c and d are described in preceeding section. The family members of $\bar{y}'_{zb_i}$ are same as $\bar{y}_{zb_i}$, available in **Table (4.1)**.

Zaman and Bulut (2019a) have used Taylor series method for $h\left(\bar{y}, \bar{x}\right) = \bar{y}_{zb_i}$ and obtained theoretical MSE. In current section, we are adapting their methodology for $h\left(\bar{y}_2, \bar{x}_1, \bar{x}_2\right) = \bar{y}'_{zb_i}$ and obtaining MSE for case-1 as follows:

$$MSE\left(\bar{y}'_{zb_i}\right)I = d\Sigma d', \qquad (4.10)$$

Where,

$$d=\left[\frac{\delta h(\bar{y}_2,\bar{x}_1,\bar{x}_2)}{\delta \bar{y}_2}\Big|\bar{Y},\bar{X} \quad \frac{\delta h(\bar{y}_2,\bar{x}_1,\bar{x}_2)}{\delta \bar{x}_1}\Big|\bar{Y},\bar{X} \quad \frac{\delta h(\bar{y}_2,\bar{x}_1,\bar{x}_2)}{\delta \bar{x}_2}\Big|\bar{Y},\bar{X}\right],$$

$$d=\left[1 \quad (g_t+B_t) \quad -(g_t+B_t)\right]$$

$$\Sigma=\begin{bmatrix} V(\bar{y}_2) & Cov(\bar{y}_2,\bar{x}_1) & Cov(\bar{y}_2,\bar{x}_2) \\ Cov(\bar{x}_1,\bar{y}_2) & V(\bar{x}_1) & Cov(\bar{x}_1,\bar{x}_2) \\ Cov(\bar{x}_2,\bar{y}_2) & Cov(\bar{x}_2,\bar{x}_1) & V(\bar{x}_2) \end{bmatrix},$$

with,

$$\begin{aligned}
&V(\bar{y}_2)=\gamma_2 S_y^2,\\
&V(\bar{x}_1)=\gamma_1 S_x^2,\\
&V(\bar{x}_2)=\gamma_2 S_y^2,\\
&Cov(\bar{y}_2,\bar{x}_1)=Cov(\bar{x}_1,\bar{y}_2)=\gamma_1 S_{yx},\\
&Cov(\bar{y}_2,\bar{x}_2)=Cov(\bar{x}_2,\bar{y}_2)=\gamma_2 S_{yx},\\
&Cov(\bar{x}_1,\bar{x}_2)=Cov(\bar{x}_2,\bar{x}_1)=\gamma_1 S_x^2.
\end{aligned}$$

By using the defined notations of the variance with co-variance, we substituting the values of d in equation (4.9), and get the MSE expressions of $\bar{y}'_{zb_t}$ for case-I as:

$$MSE\left(\bar{y}'_{zb_t}\right)I=\gamma_2 S_y^2+(\gamma_2-\gamma_1)\left[(g_t+B_t)^2 S_x^2-2(g_t+B_t)S_{yx}\right]. \tag{4.11}$$

To obtain MSE for case-II, all the notations will remain same except, $Cov(\bar{y}_2,\bar{x}_1)=Cov(\bar{x}_1,\bar{y}_2)=0$ and $Cov(\bar{x}_1,\bar{x}_2)=Cov(\bar{x}_2,\bar{x}_1)=0$. Hence the MSE of $\bar{y}'_{zb_t}$ for case-II is given by

$$MSE\left(\bar{y}'_{zb_t}\right)II=\gamma_2\left(S_y^2-2(g_t+B_t)S_{yx}\right)+(g_t+B_t)^2(\gamma_1+\gamma_2)S_x^2. \tag{4.12}$$

As the minimum MSE of Zaman (2019) class of estimators is equal to traditional regression estimator. So we are considering here traditional regression estimator for two phase sampling as follows:

$$\bar{y}'_{reg} = \bar{y}_2 + b_{(i)}(\bar{x}_1 - \bar{x}_2). \tag{4.13}$$

Note that, Pradhan (2005) only provide MSE expressions for $\bar{y}'_{reg}$, case-II. So, we incorporated their MSE expressions for case-II. We also find the MSE expressions for $\bar{y}'_{reg}$, case-I. The MSE of $\bar{y}'_{reg}$, for case-I and case-II respectively, as given below

$$\begin{aligned} MSE(\bar{y}'_{reg})I &= S_y^2\left[\gamma_2 - (\gamma_2 - \gamma_1)\rho^2\right], \\ MSE(\bar{y}'_{reg})II &= S_y^2\left[\gamma_2 + (\gamma_2 - \gamma_1)\rho^2\right]. \end{aligned} \tag{4.14}$$

Where, $\gamma_1 = \left(\frac{1}{n_1} - \frac{1}{N}\right)$ and $\gamma_2 = \left(\frac{1}{n_2} - \frac{1}{N}\right)$.

4.3.2 Proposed estimators: two-stage sampling design

In this section, we present a family of estimators by using two-stage sampling design as under:

$$\bar{y}'_{N_i} = k_1\left\{\bar{y}_2 + b_{(i)}(\bar{x}_1 - x)\right\} + k_2(\bar{x}_1 - \bar{x}_2). \qquad for \;\; i = 1, 2, \dots, 7 \tag{4.15}$$

The family members of $\bar{y}'_{N_i}$ is same as $\bar{y}_{N_i}$, available in **Table (4.2)**. To obtain MSE for case-1. Let us define $\eta_{y_2} = \frac{\bar{y}_2 - \bar{Y}}{\bar{Y}}, \eta_{x_1} = \frac{\bar{x}_1 - \bar{X}}{\bar{X}}$ and $\eta_{x_2} = \frac{\bar{x}_2 - \bar{X}}{\bar{X}}$.

Utilizing these notations, we can write,

$$E(\eta_{y_2}) = E(\eta_{x_2}) = 0, \quad E(\eta_{y_2}^2) = \gamma_2 C_y^2, \qquad E(\eta_{x_1}^2) = \gamma_1 C_x^2, \qquad E(\eta_{x_2}^2) = \gamma_2 C_x^2,$$

$$E(\eta_{y_2}\,\eta_{x_1}) = \gamma_1 C_{yx}\,,$$

with

$$E(\eta_{y_2}\eta_{x_1}) = \gamma_2 C_{yx},$$

and

$$E(\eta_{x_1}\eta_{x_2}) = \gamma_1 C_x^2.$$

Now, expending $\bar{y}'_{N_t}$ in terms of $\eta's$ as given below:

$$\bar{y}'_{N_t} = k_1\left\{\bar{Y}\left(1+\eta_{y_2}\right)+b_{(t)}\bar{X}\left(\eta_{x_1}-\eta_{x_2}\right)\right\}+k_2\bar{X}\left(\eta_{x_1}-\eta_{x_2}\right). \tag{4.16}$$

$$\bar{y}'_{N_t}-\bar{Y} = k_1\left\{\bar{Y}\left(1+\eta_{y_2}\right)+b_{(t)}\bar{X}\left(\eta_{x_1}-\eta_{x_2}\right)\right\}+k_2\bar{X}\left(\eta_{x_1}-\eta_{x_2}\right)-\bar{Y}.$$

Taking the expectation after squaring on both sides of eq. (4.12), up to the order n^{-1}, and we get

$$MSE\left(\bar{y}'_{N_t}\right) = \bar{Y}^2 + k_1^2\tau_{AN} + k_2^2\tau_{BN} + 2k_1k_2\tau_{CN} - 2k_1\tau_{DN}, \tag{4.17}$$

where

$$\tau_{AN} = \left[\bar{Y}^2\left(1+\gamma_2 C_y^2\right)+\left(\gamma_2-\gamma_1\right)b_{(t)}\bar{X}\left\{b_{(t)}\bar{X}C_x^2-2\bar{Y}C_{yx}\right\}\right],$$
$$\tau_{BN} = \bar{X}^2\left(\gamma_2-\gamma_1\right)C_x^2,$$
$$\tau_{CN} = \left(\gamma_1-\gamma_2\right)\bar{X}\left[\bar{Y}C_{yx}-b_{(t)}\bar{X}C_x^2\right],$$
$$\tau_{DN} = \bar{Y}^2$$

which is minimum for

$$k_1^{opt} = \left[\frac{\tau_{BN}\tau_{DN}}{\tau_{AN}\tau_{BN}-\tau_{CN}^2}\right],$$

and

$$k_2^{opt} = \left[\frac{\tau_{CN}\tau_{DN}}{\tau_{AN}\tau_{BN} - \tau_{CN}^2}\right].$$

$$MSE\left(\bar{y}'_{N_t}\right)I = \left[\bar{Y}^2 - \frac{\tau_{BN}\tau_{DN}^2}{\tau_{AN}\tau_{BN} - \tau_{CN}^2}\right]. \qquad (4.18)$$

To obtain MSE for case-II, all the notations will remain same except, $E\left(\eta_{y_2}\eta_{x_1}\right) = 0 = E\left(\eta_{x_1}\eta_{x_2}\right)$. Hence the MSE of $\bar{y}'_{N_t}$ for case-II as given below:

$$MSE\left(\bar{y}'_{N_t}\right) = \bar{Y}^2 + k_1^2\Psi_{AN} + k_2^2\Psi_{BN} - 2k_1\Psi_{DN}, \qquad (4.19)$$

where

$$\begin{aligned}
\Psi_{AN} &= \left[\bar{Y}^2\left(1+\gamma_2 C_y^2\right) + b_{(t)}^2\bar{X}^2\left(\gamma_1+\gamma_2\right)C_x^2 - 2b_{(t)}\bar{X}\bar{Y}\gamma_2 C_{yx}\right],\\
\Psi_{BN} &= \bar{X}^2\left(\gamma_2+\gamma_1\right)C_x^2,\\
\Psi_{CN} &= \bar{X}\left[b_{(t)}\bar{X}\left(\gamma_1+\gamma_2\right)C_x^2 - \bar{Y}\gamma_2 C_{yx}\right],\\
\Psi_{DN} &= \bar{Y}^2.
\end{aligned}$$

which is minimum for

$$k_1^{opt} = \left[\frac{\Psi_{BN}\Psi_{DN}}{\Psi_{AN}\Psi_{BN} - \Psi_{CN}^2}\right],$$

and

$$k_2^{opt} = \left[\frac{\Psi_{CN}\Psi_{DN}}{\Psi_{AN}\Psi_{BN} - \Psi_{CN}^2}\right].$$

$$MSE\left(\bar{y}'_{N_t}\right) = \left[\bar{Y}^2 - \frac{\Psi_{BN}\Psi_{DN}^2}{\Psi_{AN}\Psi_{BN} - \Psi_{CN}^2}\right]. \qquad (4.20)$$

4.4 Numerical illustration

For the assessment of the proposed and existing estimators. We consider a real life application as Pop.-1 and an artificial population as Pop.-2.

4.4.1 Real life application

In this section, we utilized data set available in Singh (2003 p. 1111). This data set is recently utilized by Ali et al. (2021) for sensitivity issue by adding scramble response in it. Here, we are considering this data in absence of sensitivity. As there is a non-negative correlation exist between the study and auxiliary variates, also, **Fig. (4.1)** and **Fig. (4.2)** present a graphical trends of non-normality data with presenting the outliers, respectively, by using the robust-regression tools. Results of percentage relative efficiency (PRE) are provided in **Table (4.3)**. Some major characteristics of the population are as given below:

X = Amount of non-real estate farm loans during 1977 and

Y = Amount of real estate farm loans during 1977.

N=50	$\bar{Y}$ = 555:4345	$\bar{X}$ = 878:1624	$b_{(lts)}$ = 0:3484253
n = 20	S_y = 584:826	S_x = 1084:678	$b_{(hbm)}$ = 0:4123359
ρ = 0:804	C_x = 4:617048	$b_{(OLS)}$ = 0:4334034	$b_{(hpm)}$ = 0:4267937
n_1 = 16	C_x = 1:235168	$b_{(lad)}$ = 0:3937749	$b_{(tkz)}$ = 0:4187815
n_2 = 20	C_y = 1:052916	$b_{(lms)})$ = 0:3396594	$b_{(hmm)}$ = 0:3480814

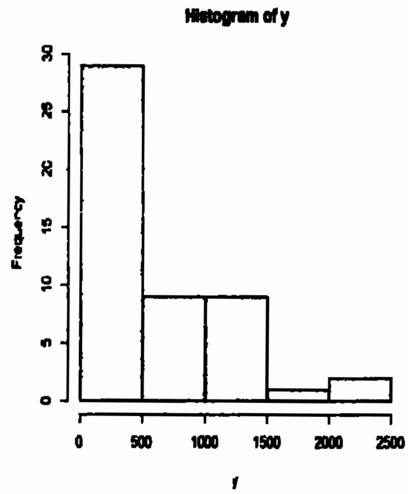


Fig. 4.1: Histogram Pop-1.

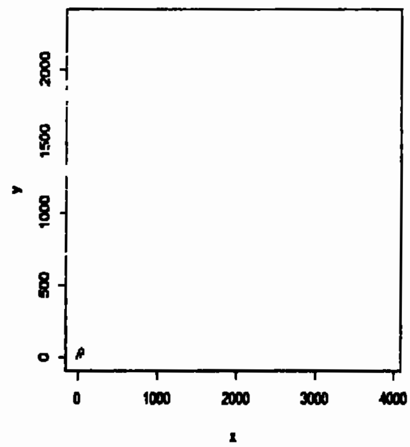


Fig. 4.2: Scatter Plot Pop-1.

4.4.2 Simulation study

In Section (4.2), an assessment of proposed and existing estimators performed with the assumption that all the population parameters are known. But in numerous genuine circumstances, these parameters are mostly obscure and can't be speculated based on past information assumption. Subsequently they should be evaluated. In such circumstances, an additional variability is presented in the evaluations that could invalid the hypothetical examinations. So in this sub-section, we are paying our attention regarding the PRE examinations at the point when obscure population parameters are assessed from the selected sample. For this purpose we are performing Monte Carlo simulation.

The simulation design is organized as follows: A random variable $X \backsim (2.6,\ 3.8)$ and random Variable Y_i is defined as $Y_i = h + RX_i + \in X_i^p$. Here we assume $p = 1.6, h = 5, R = 2$ and ϵ has standard normal distribution. We consider the population of size *N=1000*. Here, simple random sampling (SRS) is considered for n = *200*. The SRS sampling has been replicated 1000 times. We examine empirical MSE's of

$\bar{y}_{zb_i}, \bar{y}_{reg}$ and $\bar{y}_{N_i}$, as $MSE = \frac{\sum_{i=1}^{K'} (z_i - \bar{Z})^2}{K'}$. using the results of empirical MSE we calculate PRE of each estimator, available in **Table (4.4).**

We consider same simulation design for two stage sampling. A sample of size $n_1 = 200$ is selected from $(X_i; Y_i)$ at first stage and sample of size $n_2 = 160$ is selected at second stage. The Second stage sample is selected differently for case-I and case-II as per requirement of no-independence and independence with respect to initial stage sample i.e. n_1, respectively. **Fig. (4.3)** and **Fig. (4.4)** clearly show the applicability of robust-regression tools. The results of PRE for case-I and case-II are provided in **Table (4.4).** Note that PRE for each estimator is calculated with respect to $V(\bar{y})$ as:

$$PRE(\hat{\theta}) = \frac{Var(\bar{y})}{MSE(\hat{\theta})} \times 100. \quad (4.21)$$

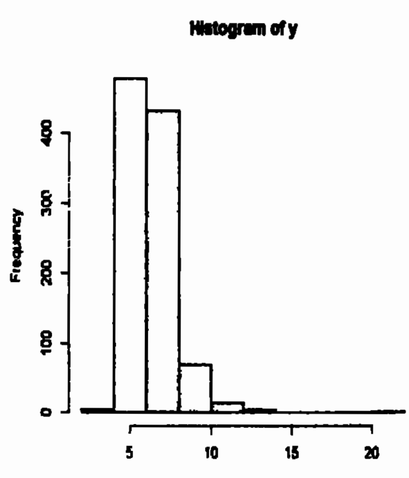


Fig. 4.3: Histogram Pop-2.

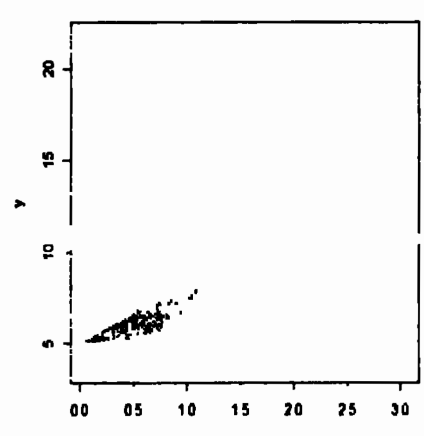


Fig. 4.4: Scatter Plot Pop-2.

4.5 Discussion

The results of numerical illustration are provided in **Table (4.3)** and **Table (4.4)**. Our findings are highlighted as given below:

- $\bar{y}_{zb_1}$, has the maximum PRE as compare to all the reviewed ratio type estimators under SRS and two stage sampling schemes.
- $\bar{y}_{reg}$, has the maximum PRE as compare to all the reviewed ratio type estimators under SRS and first case of two stage sampling schemes. However, $\bar{y}_{zb_1}$ is performing better than the usual regression estimator in case-II of two stage sampling scheme.
- By ignoring fractional values in proposed class, we observe that all the members of proposed class are equally important under SRS and first case of two stage sampling schemes. However for case-II, $\bar{y}'_{N_2}$ is performing outclass among the proposed class of estimators.
- All the estimators of proposed class have maximum PRE over sample mean estimator, $\bar{y}_{zb_1}$, $\bar{y}_{z_1}$ and $\bar{y}_{reg}$ under SRS and two stage sampling schemes.

According to the real life application and simulation results, we observed that, the new proposals out-perform over existing and adapting ones. Which clearly showed the superiority of proposed class over reviewed estimators?

4.6 Summary of the chapter

We have suggested a family of robust regression-type estimators for mean estimation under simple random sampling and two stage sampling schemes when quantitative supplementary information is available, in this chapter. We also find MSE and minimum MSE expressions of the proposed class. The proposed class of estimators has been compared with existing ones. Based on theoretical and empirical PRE results, it is clear that the proposed class performs better as compare to sample mean estimator, traditional regression estimator, Zaman (2019) estimators and Zaman and Bulut (2019a) estimators. Hence, it is recommended to utilize the proposed class of estimators in real life applications. In future studies, we hope to extend the proposed robust-regression class of estimators presented in this article to the sensitive issue, in light of Ali et al. (2021).

Table 4.1: Reviewed robust ratio type estimators.

Estimator	$b_{(i)}$	c	d
$\bar{y}_{zb_1}$	$b_{(lad)}$	1	0
$\bar{y}_{zb_2}$	$b_{(lad)}$	1	C_x
$\bar{y}_{zb_3}$	$b_{(lad)}$	1	$\beta_2(x)$
$\bar{y}_{zb_4}$	$b_{(lad)}$	$\beta_2(x)$	C_x
$\bar{y}_{zb_5}$	$b_{(lad)}$	C_x	$\beta_2(x)$
$\bar{y}_{zb_6}$	$b_{(lms)}$	1	0
$\bar{y}_{zb_7}$	$b_{(lms)}$	1	C_x
$\bar{y}_{zb_8}$	$b_{(lms)}$	1	$\beta_2(x)$
$\bar{y}_{zb_9}$	$b_{(lms)}$	$\beta_2(x)$	C_x
$\bar{y}_{zb_{10}}$	$b_{(lms)}$	C_x	$\beta_2(x)$
$\bar{y}_{zb_{11}}$	$b_{(lts)}$	1	0
$\bar{y}_{zb_{12}}$	$b_{(lts)}$	1	C_x
$\bar{y}_{zb_{13}}$	$b_{(lts)}$	1	$\beta_2(x)$
$\bar{y}_{zb_{14}}$	$b_{(lts)}$	$\beta_2(x)$	C_x
$\bar{y}_{zb_{15}}$	$b_{(lts)}$	C_x	$\beta_2(x)$
$\bar{y}_{zb_{16}}$	$b_{(hbm)}$	1	0
$\bar{y}_{zb_{17}}$	$b_{(hbm)}$	1	C_x
$\bar{y}_{zb_{18}}$	$b_{(hbm)}$	1	$\beta_2(x)$

$\bar{y}_{zb_{19}}$	$b_{(hbm)}$	$\beta_2(x)$	C_x
$\bar{y}_{zb_{20}}$	$b_{(hbm)}$	C_x	$\beta_2(x)$
$\bar{y}_{zb_{21}}$	$b_{(hpm)}$	1	0
$\bar{y}_{zb_{22}}$	$b_{(hpm)}$	1	C_x
$\bar{y}_{zb_{23}}$	$b_{(hpm)}$	1	$\beta_2(x)$
$\bar{y}_{zb_{24}}$	$b_{(hpm)}$	$\beta_2(x)$	C_x
$\bar{y}_{zb_{25}}$	$b_{(hpm)}$	C_x	$\beta_2(x)$
$\bar{y}_{zb_{26}}$	$b_{(tkz)}$	1	0
$\bar{y}_{zb_{27}}$	$b_{(tkz)}$	1	C_x
$\bar{y}_{zb_{28}}$	$b_{(tkz)}$	1	$\beta_2(x)$
$\bar{y}_{zb_{29}}$	$b_{(tkz)}$	$\beta_2(x)$	C_x
$\bar{y}_{zb_{30}}$	$b_{(tkz)}$	C_x	$\beta_2(x)$
$\bar{y}_{zb_{31}}$	$b_{(hmm)}$	1	0
$\bar{y}_{zb_{32}}$	$b_{(hmm)}$	1	C_x
$\bar{y}_{zb_{33}}$	$b_{(hmm)}$	1	$\beta_2(x)$
$\bar{y}_{zb_{34}}$	$b_{(hmm)}$	$\beta_2(x)$	C_x
$\bar{y}_{zb_{35}}$	$b_{(hmm)}$	C_x	$\beta_2(x)$

Table 4.2: Family member of proposed class.

SRS Estimators	$b_{(i)}$	**Two stage Sampling Estimators**
$\bar{y}_{N_1}$	$b_{(lad)}$	$\bar{y}'_{N_1}$
$\bar{y}_{N_2}$	$b_{(lms)}$	$\bar{y}'_{N_2}$
$\bar{y}_{N_3}$	$b_{(lts)}$	$\bar{y}'_{N_3}$
$\bar{y}_{N_4}$	$b_{(hbm)}$	$\bar{y}'_{N_4}$
$\bar{y}_{N_5}$	$b_{(hpm)}$	$\bar{y}'_{N_5}$
$\bar{y}_{N_6}$	$b_{(tkz)}$	$\bar{y}'_{N_6}$
$\bar{y}_{N_7}$	$b_{(hmm)}$	$\bar{y}'_{N_7}$

Table 4.3: PRE of Population-I.

SRS		Two Stage Case-I		Two Stage Case-II	
$\hat{\theta}$	PRE	$\hat{\theta}$	PRE	$\hat{\theta}$	PRE
$\bar{y}$	100	$\bar{y}'$	100	$\bar{y}'$	100
$\bar{y}_{zb1}$	2132.71	$\bar{y}'_{zb1}$	2018.69	$\bar{y}'_{zb1}$	571.04
$\bar{y}_{zb2}$	2137.66	$\bar{y}'_{zb2}$	2020.54	$\bar{y}'_{zb2}$	572.16
$\bar{y}_{zb3}$	2151.23	$\bar{y}'_{zb3}$	2025.57	$\bar{y}'_{zb3}$	575.22
$\bar{y}_{zb4}$	8625.92	$\bar{y}'_{zb4}$	2871.05	$\bar{y}'_{zb4}$	2198.24
$\bar{y}_{zb5}$	2147.70	$\bar{y}'_{zb5}$	2024.26	$\bar{y}'_{zb5}$	574.42
$\bar{y}_{zb6}$	2464.93	$\bar{y}'_{zb6}$	2132.02	$\bar{y}'_{zb6}$	645.14
$\bar{y}_{zb7}$	2470.94	$\bar{y}'_{zb7}$	2133.89	$\bar{y}'_{zb7}$	646.47
$\bar{y}_{zb8}$	2487.41	$\bar{y}'_{zb8}$	2138.99	$\bar{y}'_{zb8}$	650.10
$\bar{y}_{zb9}$	9252.26	$\bar{y}'_{zb9}$	2898.26	$\bar{y}'_{zb9}$	2580.40
$\bar{y}_{zb10}$	2483.13	$\bar{y}'_{zb10}$	2137.67	$\bar{y}'_{zb10}$	649.16
$\bar{y}_{zb11}$	2406.64	$\bar{y}'_{zb11}$	2113.57	$\bar{y}'_{zb11}$	632.26
$\bar{y}_{zb12}$	2412.46	$\bar{y}'_{zb12}$	2115.44	$\bar{y}'_{zb12}$	633.55
$\bar{y}_{zb13}$	2428.42	$\bar{y}'_{zb13}$	2120.53	$\bar{y}'_{zb13}$	637.08
$\bar{y}_{zb14}$	9179.14	$\bar{y}'_{zb14}$	2895.25	$\bar{y}'_{zb14}$	2515.99
$\bar{y}_{zb15}$	2424.27	$\bar{y}'_{zb15}$	2119.21	$\bar{y}'_{zb15}$	636.16
$\bar{y}_{zb16}$	2032.70	$\bar{y}'_{zb16}$	1980.27	$\bar{y}'_{zb16}$	548.39
$\bar{y}_{zb17}$	3237.34	$\bar{y}'_{zb17}$	1982.10	$\bar{y}'_{zb17}$	549.44
$\bar{y}_{zb18}$	2050.05	$\bar{y}'_{zb18}$	1987.09	$\bar{y}'_{zb18}$	552.33
$\bar{y}_{zb19}$	8332.40	$\bar{y}'_{zb19}$	2857.09	$\bar{y}'_{zb19}$	2077.35
$\bar{y}_{zb20}$	2046.75	$\bar{y}'_{zb20}$	1985.80	$\bar{y}'_{zb20}$	551.58
$\bar{y}_{zb21}$	1959.18	$\bar{y}'_{zb21}$	1950.56	$\bar{y}'_{zb21}$	531.61
$\bar{y}_{zb22}$	1963.60	$\bar{y}'_{zb22}$	1952.38	$\bar{y}'_{zb22}$	532.62
$\bar{y}_{zb23}$	1975.68	$\bar{y}'_{zb23}$	1957.34	$\bar{y}'_{zb23}$	535.38
$\bar{y}_{zb24}$	8084.92	$\bar{y}'_{zb24}$	2844.65	$\bar{y}'_{zb24}$	1987.31
$\bar{y}_{zb25}$	1972.54	$\bar{y}'_{zb25}$	1956.05	$\bar{y}'_{zb25}$	534.66
$\bar{y}_{zb26}$	1999.47	$\bar{y}'_{zb26}$	1967.01	$\bar{y}'_{zb26}$	540.82
$\bar{y}_{zb27}$	2004.01	$\bar{y}'_{zb27}$	1968.82	$\bar{y}'_{zb27}$	541.85
$\bar{y}_{zb28}$	2016.93	$\bar{y}'_{zb28}$	1973.80	$\bar{y}'_{zb28}$	544.68
$\bar{y}_{zb29}$	8223.43	$\bar{y}'_{zb29}$	2851.71	$\bar{y}'_{zb29}$	2036.75
$\bar{y}_{zb30}$	2013.20	$\bar{y}'_{zb30}$	1972.51	$\bar{y}'_{zb30}$	543.95
$\bar{y}_{zb31}$	2408.89	$\bar{y}'_{zb31}$	2114.29	$\bar{y}'_{zb31}$	632.76

$\bar{y}_{zb32}$	2414.72	$\bar{y}'_{zb32}$	2116.16	$\bar{y}'_{zb32}$	634.05
$\bar{y}_{zb33}$	2430.70	$\bar{y}'_{zb33}$	2121.25	$\bar{y}'_{zb33}$	637.58
$\bar{y}_{zb34}$	9182.23	$\bar{y}'_{zb34}$	2895.38	$\bar{y}'_{zb34}$	2518.50
$\bar{y}_{zb35}$	2426.54	$\bar{y}'_{zb35}$	2119.93	$\bar{y}'_{zb35}$	636.66
$\bar{y}_{reg}$	9420.16	$\bar{y}'_{reg}$	2905.02	$\bar{y}'_{reg}$	1977.18
$\bar{y}_{N1}$	9531.63	$\bar{y}'_{N1}$	3015.69	$\bar{y}'_{N1}$	18565.24
$\bar{y}_{N2}$	9531.26	$\bar{y}'_{N2}$	3015.93	$\bar{y}'_{N2}$	26053.66
$\bar{y}_{N3}$	9531.13	$\bar{y}'_{N3}$	3015.53	$\bar{y}'_{N3}$	25037.30
$\bar{y}_{N4}$	9531.07	$\bar{y}'_{N4}$	3015.48	$\bar{y}'_{N4}$	16030.02
$\bar{y}_{N5}$	9531.91	$\bar{y}'_{N5}$	3015.02	$\bar{y}'_{N5}$	14256.48
$\bar{y}_{N6}$	9531.55	$\bar{y}'_{N6}$	3015.41	$\bar{y}'_{N6}$	15215.86
$\bar{y}_{N7}$	9531.64	$\bar{y}'_{N7}$	3015.25	$\bar{y}'_{N7}$	25080.03

Table 4.4: PRE of Population-2.

SRS		Two Stage Case-I		Two Stage Case-II	
$\hat{\theta}$	PRE	$\hat{\theta}$	PRE	$\hat{\theta}$	PRE
$\bar{y}$	100	$\bar{y}'$	100	$\bar{y}'$	100
$\bar{y}_{zb1}$	2276.96	$\bar{y}'_{zb1}$	5992.13	$\bar{y}'_{zb1}$	998.62
$\bar{y}_{zb2}$	7237.82	$\bar{y}'_{zb2}$	12058.50	$\bar{y}'_{zb2}$	3026.99
$\bar{y}_{zb3}$	33694.82	$\bar{y}'_{zb3}$	19799.42	$\bar{y}'_{zb3}$	15053.22
$\bar{y}_{zb4}$	29157.60	$\bar{y}'_{zb4}$	19045.03	$\bar{y}'_{zb4}$	11884.51
$\bar{y}_{zb5}$	37168.84	$\bar{y}'_{zb5}$	20314.48	$\bar{y}'_{zb5}$	18254.59
$\bar{y}_{zb6}$	2247.27	$\bar{y}'_{zb6}$	5921.10	$\bar{y}'_{zb6}$	1047.90
$\bar{y}_{zb7}$	7081.52	$\bar{y}'_{zb7}$	11902.29	$\bar{y}'_{zb7}$	3282.46
$\bar{y}_{zb8}$	32938.70	$\bar{y}'_{zb8}$	19684.77	$\bar{y}'_{zb8}$	17047.55
$\bar{y}_{zb9}$	38160.15	$\bar{y}'_{zb9}$	20316.48	$\bar{y}'_{zb9}$	20472.14
$\bar{y}_{zb10}$	36514.21	$\bar{y}'_{zb10}$	20230.06	$\bar{y}'_{zb10}$	20469.01
$\bar{y}_{zb11}$	2227.07	$\bar{y}'_{zb11}$	5932.43	$\bar{y}'_{zb11}$	1027.16
$\bar{y}_{zb12}$	6976.16	$\bar{y}'_{zb12}$	11927.26	$\bar{y}'_{zb12}$	3173.54
$\bar{y}_{zb13}$	32410.50	$\bar{y}'_{zb13}$	19703.53	$\bar{y}'_{zb13}$	16202.11
$\bar{y}_{zb14}$	27874.02	$\bar{y}'_{zb14}$	18926.26	$\bar{y}'_{zb14}$	12798.11

$\bar{y}_{zb15}$	36044.12	$\bar{y}'_{zb15}$	20244.03	$\bar{y}'_{zb15}$	19548.56
$\bar{y}_{zb16}$	2299.48	$\bar{y}'_{zb16}$	5978.75	$\bar{y}'_{zb16}$	987.61
$\bar{y}_{zb17}$	7357.54	$\bar{y}'_{zb17}$	12029.12	$\bar{y}'_{zb17}$	2971.77
$\bar{y}_{zb18}$	34250.98	$\bar{y}'_{zb18}$	19778.35	$\bar{y}'_{zb18}$	14619.03
$\bar{y}_{zb19}$	29729.47	$\bar{y}'_{zb19}$	19018.77	$\bar{y}'_{zb19}$	11543.59
$\bar{y}_{zb20}$	37635.41	$\bar{y}'_{zb20}$	20299.14	$\bar{y}'_{zb20}$	17753.99
$\bar{y}_{zb21}$	2326.47	$\bar{y}'_{zb21}$	5921.55	$\bar{y}'_{zb21}$	1003.25
$\bar{y}_{zb22}$	7502.35	$\bar{y}'_{zb22}$	11903.28	$\bar{y}'_{zb22}$	3050.51
$\bar{y}_{zb23}$	34896.04	$\bar{y}'_{zb23}$	19865.52	$\bar{y}'_{zb23}$	15238.09
$\bar{y}_{zb24}$	30406.97	$\bar{y}'_{zb24}$	18904.15	$\bar{y}'_{zb24}$	12030.34
$\bar{y}_{zb25}$	38159.11	$\bar{y}'_{zb25}$	20230.62	$\bar{y}'_{zb25}$	18465.89
$\bar{y}_{zb26}$	2249.37	$\bar{y}'_{zb26}$	5905.88	$\bar{y}'_{zb26}$	1028.33
$\bar{y}_{zb27}$	7092.51	$\bar{y}'_{zb27}$	11868.73	$\bar{y}'_{zb27}$	3179.62
$\bar{y}_{zb28}$	32992.97	$\bar{y}'_{zb28}$	19659.31	$\bar{y}'_{zb28}$	16249.63
$\bar{y}_{zb29}$	28450.23	$\bar{y}'_{zb29}$	18872.06	$\bar{y}'_{zb29}$	12836.31
$\bar{y}_{zb30}$	36561.93	$\bar{y}'_{zb30}$	20211.01	$\bar{y}'_{zb30}$	19601.05
$\bar{y}_{zb31}$	2255.60	$\bar{y}'_{zb31}$	5907.31	$\bar{y}'_{zb31}$	1026.92
$\bar{y}_{zb32}$	7125.18	$\bar{y}'_{zb32}$	11871.89	$\bar{y}'_{zb32}$	3172.31
$\bar{y}_{zb33}$	33153.25	$\bar{y}'_{zb33}$	19661.72	$\bar{y}'_{zb33}$	16192.56
$\bar{y}_{zb34}$	28610.46	$\bar{y}'_{zb34}$	18875.01	$\bar{y}'_{zb34}$	12790.44
$\bar{y}_{zb35}$	36702.23	$\bar{y}'_{zb35}$	20212.81	$\bar{y}'_{zb35}$	19538.01
$\bar{y}_{reg}$	41210.94	$\bar{y}'_{reg}$	20794.12	$\bar{y}'_{reg}$	16919.97
y_{N1}	41236.83	y'_{N1}	20808.04	y'_{N1}	55716.02
y_{N2}	41236.26	y'_{N2}	20808.47	y'_{N2}	60743.40
y_{N3}	41236.53	y'_{N3}	20808.11	y'_{N3}	59049.46
y_{N4}	41236.71	y'_{N4}	20808.28	y'_{N4}	54175.60
y_{N5}	41236.16	y'_{N5}	20808.76	y'_{N5}	56327.54
y_{N6}	41236.03	y'_{N6}	20808.22	y'_{N6}	59162.35
y_{N7}	41236.55	y'_{N7}	20808.41	y'_{N7}	59026.51

Chapter 5

An extension of robust regression techniques for two auxiliary variables

5.1 Background

For future development, each community needs careful planning to manage its affairs efficiently. Successful preparation or planning requires many types of data that are reasonably accurate. Everything is changing rapidly in this modem environment, requiring the regular collection of up-to-date information. It is possible to collect data in two ways, which are a complete survey of the enumeration and a sample survey. Since data collection is subject to time and cost constraints, regular data collection by full enumeration is typically not feasible. The only solution then is sample surveys. Through surveying part of a population as a sample, more effort can be made to gather more accurate data through hiring by better-trained workers, organization, monitoring, etc. compared to full details\enumeration, see, Chand (1975).

Abid et al. (2018) have suggested a new family of estimators with utilizing variables in simple random sampling. Their study shows by involving the non-conventional location measures for mean estimation provides more efficient results than the conventional location measures. It is worthy to note that their estimators are based on generalized versions of regression-type estimators with non-traditional regressions weight. These estimators provide better results in presence of extreme values. This study, proposes a new family of estimations by using robust regressions tools. We have extended the idea by replacing the ordinary least regression coefficient with different functions of robust regression functions available in the literature. These functions are highly robust in presence of extreme values. The general form of the MSE of the proposed class of estimator is also derived. The real-life data sets related to polio, taxation, and agriculture have been considered for measuring the efficiency of proposed estimators over existing ones. We have drawn the scatter plots and box plots for all these data sets. All these figures have shown that the data sets have the issue of extreme observations. Hence suitable for existing and proposed estimators. The theoretical outcomes are being supported by real-life data sets.

In survey sampling, it is regular to make utilization of auxiliary\supplementary information to acquire enhanced designs and more effective estimators. This information might be utilized at the planning phase of the study, in the estimation methodology, or at both stages. The huge amount of sampling literature portrays a variety of techniques for using supplementary information (see, e.g., Searls, 1964; Sarndal et al., 1992; Oral, 2011; Koyuncu, 2012; Shahzad et al., 2019 and Irfan et al., 2018, 2019). At the estimation stage, in many sampling situations of the survey, estimators of ratio and regression are commonly used when using supplementary information. These estimators are nearly identical, if relationship between the two variables (study and auxiliary) is a straight line that moves through the neighborhood of origin and the study variable's variance is proportional to the auxiliary variable(s). In practice, traditional ratio estimators are less efficient than regression estimators if this criterion is not met. In order to address this issue in the survey sampling literature, substantial research has been carried out to improve ratio estimators by providing different modified\adjusted ratio type estimators. In addition, if the study variable and the auxiliary variable(s) have a positive connection, the ratio estimator is used quite effectively. There are numerous practical instances (medical, biological, economical, and industrial sectors) when a positive correlation between two or more variables (one is the study variable, and the other are auxiliary variables) exists. Some real-life examples of positive correlation between the study variable and two auxiliary variables are:

(a) The sale of a particular commodity rises with the increase in the region's population and average per capita income.

(b) The productivity of the employee improves with both his previous experience and his educational or intelligence level.

(c) The human body's immunity increases from the risk of certain diseases by following healthy diets and paying attention to fitness, etc.

However, when the information related to two or more supplementary variables available, a vast amount of literature can be found in Olkin (1958); Raj (1965); Rao and Mudholkar (1967); Abu-Dayyeh et al. (2003); Lu and Yan (2014) and Abid et al. (2018). This chapter is based on the class of estimators developed by Abid et al. (2018). We have made an attempt to extend Abid et al. (2018) work and suggest a new and improved class of population mean estimators of a study variable.

The rest of the chapter is organized as follows. In Section 5.2 we provide a detail description of existing estimators, attributed to Abid et al. (2018). In next Section 5.3, we propose a new class of estimators and show some conceivable estimators having a place with the class. We show expressions of large sample properties, through mean square with minimum mean square error. In Section 5.4, to reveal insight into the productivity of the proposed class, various numerical illustrations are done with competitive estimators on the premises of real-life data sets. Finally, the conclusion is given in Section 5.5.

5.2 Abid et al. (2018) family of estimators

Suppose $\Omega = \{1, 2, \dots, N\}$ is a set of N bivariate units where $(x_i > 0,\ y_i > 0)$ representing the objective variable (Y) and the supplementary variable (X), respectively. Now we choose a sample of size n from the population by using SRSWOR and let $\gamma = \left(\frac{1}{n} - \frac{1}{N}\right)$. Additionally, let $\bar{Y}$ and $\bar{X}$ be the means of population for Y and X, respectively. The quantity of interest is the unknown population mean $\bar{Y}$. On the other hand, it is assumed that mean of the supplementary variable $\bar{X}$ is known. Under the $SRSWOR$ configuration, Abid et al. (2018) proposed a family of estimators for mean estimation. Let we introduce this class in generalize form as follows,

$$\hat{y}_{a(i)} = \omega_1 \bar{y}\left(\frac{\bar{X}_1 \alpha_{1(i)} + \beta_{1(i)}}{\bar{x}_1 \alpha_{1(i)} + \beta_{1(i)}}\right) + \omega_2 \bar{y}\left(\frac{\bar{X}_2 \alpha_{2(i)} + \beta_{2(i)}}{\bar{x}_2 \alpha_{2(i)} + \beta_{2(i)}}\right) \ for\ = 1,2,3,\dots,16 \qquad (5.1)$$

where, $\left(\alpha_{1(i)}, \beta_{1(i)}, \alpha_{2(i)}, \beta_{2(i)}\right)$ representing known non-conventional and conventional measures of location of X such as mid-range(MR), Hodges-Lehmann(HL), tri-mean $(TM)$and decile-mean (DM), coefficient of variationC_x, coefficient of kurtosis $\beta_2(x)$ and correlation ρ_{yx}. The sample average of the variable of interest Y denoted by $\bar{y}$. Further, $\bar{X}_1$ and $\bar{X}_2$ be the population means of first auxiliary variable and second auxiliary variable, respectively. Two weights, ω_1 and ω_2 attached for minimizing the mean square error (MSE) of $\hat{y}_{a(i)}$. All members of Abid et al. (2018) family provided see **Table (5.1)**. The MSE of $\hat{y}_{a(i)}$ is given by,

$$MSE\left(\hat{y}_{a(i)}\right) = \gamma \bar{Y}^2 (C_y^2 + \omega_1^2 \theta_1^2 C_{x1}^2 + \omega_2^2 \theta_2^2 C_{x2}^2 - 2\omega_1 \theta_1 \rho_{yx_1} C_y C_{x1} - 2\omega_2 \theta_2 \rho_{yx_2} C_y C_{x_2} + 2\omega_1 \omega_2 \theta_1 \theta_2 \rho_{x_1 x_2} C_{x_1} C_{x_2}) \qquad (5.2)$$

where, optimum value of ω_1 is:

$$\omega_1^* = \frac{\theta_2^2 C^2_{x_2} + \theta_1 \rho_{yx_1} C_y C_{x_1} - \theta_1\theta_2 \rho_{x_1x_2} C_{x_1} C_{x_2} - \theta_2 \rho_{yx_2} C_y C_{x_2}}{\theta_1^2 C^2_{x_1} + \theta_2^2 C^2_{x_2} - 2\theta_1\theta_2 \rho_{x_1x_2} C_{x_1} C_{x_2}}. \quad (5.3)$$

By using the condition of weights, $\omega_1^* + \omega_2^* = 1$.

Table 5.1: Family Members of Abid et al. (2018).

$\hat{y}_{a(l)}$	$\alpha_1(l)$	$\beta_{1(l)}$	$\alpha_2(l)$	$\beta_{2(l)}$
$\hat{y}_{a(1)}$	1	$M.R_1$	1	$M.R_2$
$\hat{y}_{a(2)}$	1	$T.M_1$	1	$T.M_2$
$\hat{y}_{a(3)}$	1	$H.L_1$	1	$H.L_2$
$\hat{y}_{a(4)}$	1	$D.M_1$	1	$D.M_2$
$\hat{y}_{a(5)}$	$\beta_2(x_1)$	$M.R_1$	$\beta_2(x_2)$	$M.R_2$
$\hat{y}_{a(6)}$	$\beta_2(x_1)$	$T.M_1$	$\beta_2(x_2)$	$T.M_2$
$\hat{y}_{a(7)}$	$\beta_2(x_1)$	$H.L_1$	$\beta_2(x_2)$	$H.L_2$
$\hat{y}_{a(8)}$	$\beta_2(x_1)$	$D.M_1$	$\beta_2(x_2)$	$D.M_2$
$\hat{y}_{a(9)}$	C_{x_1}	$M.R_1$	C_{x_2}	$M.R_2$
$\hat{y}_{a(10)}$	C_{x_1}	$T.M_1$	C_{x_2}	$T.M_2$
$\hat{y}_{a(11)}$	C_{x_1}	$H.L_1$	C_{x_2}	$H.L_2$
$\hat{y}_{a(12)}$	C_{x_1}	$D.M_1$	C_{x_2}	$D.M_2$
$\hat{y}_{a(13)}$	ρ_{yx_1}	$M.R_1$	ρ_{yx_2}	$M.R_2$
$\hat{y}_{a(14)}$	ρ_{yx_1}	$T.M_1$	ρ_{yx_2}	$T.M_2$
$\hat{y}_{a(15)}$	ρ_{yx_1}	$H.L_1$	ρ_{yx_2}	$H.L_2$
$\hat{y}_{a(16)}$	ρ_{yx_1}	$D.M_1$	ρ_{yx_2}	$D.M_2$

5.3 Proposed families of estimators

In regression analysis, the ordinary least squares (OLS) are the furthermost common traditional parametric method usually used for estimating model parameters due to its simplicity of computation and nice property. Agreeing with the theorem of Gauss-Markov, the OLS estimators are the best linear unbiased estimator. Even so, the OLS estimators are easily influenced by the existence of unusual observations (outliers) and will yield inaccurate estimates. The breakdown point of the OLS estimator is zero which indicates that it affected by only one unusual observation (single outlier). So robust methods as alternative method put forward which are not or less affected by such observations. For more details about robust regression and their tools, see e.g., Huber (1964, 1973); Hampel et al. (1986); Yohai (1987); Rousseeuw and Leroy (1987); Birkes and Dodge (1993); Bassett and Saleh (1994); Al-Noor and Mohammad (2013) and Ali et al. (2021). With robust regression, numerous ratio estimators are introduced, see, e.g., Shahzad et al. (2019); Zaman and Bulut (2019a, 2019b); Shahzad and Hanif (2019); Bulut and Zaman (2019) and Zaman (2019), using single supplementary variable. So, taking motivation from their work, we propose the following family of robust regression estimators with two auxiliary variables, through using four robust estimators that are: the least absolute deviations (LAD) regression, Huber-M (HbM) Hampel-M (HpM), and Tukey-M (TkY) as given by,

$$\hat{y}_{p_1} = \omega_1\left[\bar{y} + b_{yx_1(lad)}(\bar{X}_1 - \bar{x}_1)\right] + \omega_2\left[\bar{y} + b_{yx_2(lad)}(\bar{X}_2 - \bar{x}_2)\right], \tag{5.4}$$

$$\hat{y}_{p_2} = \omega_1\left[\bar{y} + b_{yx_1(hbm)}(\bar{X}_1 - \bar{x}_1)\right] + \omega_2\left[\bar{y} + b_{yx_2(hbm)}(\bar{X}_2 - \bar{x}_2)\right], \tag{5.5}$$

$$\hat{y}_{p_3} = \omega_1\left[\bar{y} + b_{yx_1(hpm)}(\bar{X}_1 - \bar{x}_1)\right] + \omega_2\left[\bar{y} + b_{yx_2(hpm)}(\bar{X}_2 - \bar{x}_2)\right], \tag{5.6}$$

$$\hat{y}_{p_4} = \omega_1\left[\bar{y} + b_{yx_1(tky)}(\bar{X}_1 - \bar{x}_1)\right] + \omega_2\left[\bar{y} + b_{yx_2(tky)}(\bar{X}_2 - \bar{x}_2)\right]. \tag{5.7}$$

In its general form, the proposed family of estimators can be written as:

$$\hat{y}_{p_i} = \omega_1\left[\bar{y} + b_{yx_1(i)}(\bar{X}_1 - \bar{x}_1)\right] + \omega_2\left[\bar{y} + b_{yx_2(i)}(\bar{X}_2 - \bar{x}_2)\right].\ for\ i = 1,2,..4 \tag{5.8}$$

The LAD regression is known to be the first step to achieving robustness towards the influence of outliers in Y-direction. It proposed and improved respectively by Boscovich (1757); Edgeworth (1887) and Al-Noor and Mohammad (2013). The aim of this technique is to provide a robust estimator that minimizes the total of absolute residuals. A good method for determining the optimal line among all lines that passing through a particular data point, say (x_0, y_0), described by Birkes and Dodge (1993).

The M-estimators technique was proposed and extended by Huber (1964, 1973) to achieving robustness towards the influence of outliers in Y-direction. This technique is based on the idea of putting a non-negative symmetric function of the residuals, say $\phi(r_i)$, instead of the squared residual in OLS (see, Huber 1964, 1973, 1981; Hampel 1971; and Tukey 1977), among several authors, designed formulae for the objective function as follows:

Huber-M estimator (HbM) considered the objective function, via $v = 4.685$ or 6, as,

$$\phi(r_i)=\begin{cases}\frac{1}{2}{r_i}^2 & for\ |r_i| \le v,\\ v\left(|r_i| - \frac{1}{2}v\right) & for\ |r_i| > v.\end{cases} \tag{5.9}$$

Hampel-M estimator (HpM) considered the objective function, via. $\mathrm{g} = 1.7, \mathrm{h} = 3.4$ and $\mathrm{v} = 8.5$, as,

$$\phi(r_i) = \begin{cases}\frac{1}{2}{r_i}^2, & for\ 0 < |r_i| \le g\\ g\left(|r_i| - \frac{1}{2}{r_i}^2\right), & for\ g < |r_i| \le h\\ -\frac{g}{2(v-h)}(v - r_i)^2 + \frac{g}{h}(h + v - g), & for\ h < |r_i| \le v\\ \frac{g}{h}(h + v - g). & for\ v < |r_i|.\end{cases} \tag{5.10}$$

Tukey-M estimator (TkY) considered the objective function, via $\mathrm{v} = 4.685$ or 6, as,

$$\phi(r_i) = \begin{cases}\frac{v^2}{6}\left[1 - \left\{1 - \left(\frac{r_i}{v}\right)^2\right\}^3\right], & for\ |r_i| \le v\\ \frac{v^2}{6}. & for\ |r_i| \le v\end{cases} \tag{5.11}$$

To get MSE, let us express $\bar{y} = (1 + \eta_y)\bar{Y}, \quad \bar{x}_1 = (1 + \eta_{x1})\bar{X}_1$ and

$\bar{x}_2 = (1+\eta_{x_2})\bar{X}_2$. Utilizing these notations η_i $(i = y, x_1, x_2)$, we can write:

$E(\eta_y) = E(\eta_{x1}) = E(\eta_{x2}) = 0, E(\eta_y^2) = \gamma C_y^2,\ E(\eta_{x1}^2) = \gamma C_{x1}^2 (\eta_{x2}^2) = \gamma C_{x2}^2$,

$E(\eta_y \eta_{x2}) = \gamma C_{yx1}^2$, $E(\eta_y \eta_{x2}) = \gamma C_{yx2}^2$ and $E(\eta_{x1}\eta_{x2}) = \gamma C_{x_1 x_2}$. Now expending $\hat{y}_{p_l}$ in terms of η_y, η_{x1} and η_{x2} as:

$$\hat{y}_{p_l} = [\omega_1 \bar{Y}(1+\eta_y) - b_{y.x1(l)}\bar{X}_1\eta_{x1}] + [\omega_2 \bar{Y}(1+\eta_y) - b_{y.x2(l)}\bar{X}_2\eta_{x2}]. \tag{5.12}$$

By taking expectation after sequring the eq. (5.9), up to the Order n^{-1}, we get:

$$MSE(\hat{y}_{p_l}) = \bar{Y}^2 + \omega_1^2\delta_A + \omega_2^2\delta_B + 2\omega_1\omega_2\delta_C - 2\omega_1\delta_D - \omega_2\delta_E, \tag{5.13}$$

where,

$$\delta_A = [\bar{Y}^2 + \gamma\{S_y^2 + B_{y.x1(l)}(B_{y.x1(l)}S_{x1} - 2\rho S_y)S_{x1}\}],$$

$$\delta_B = [\bar{Y}^2 + \gamma\{S_y^2 + B_{y.x2(l)}(B_{y.x2(l)}S_{x2} - 2\rho S_y)S_{x2}\}],$$

$$\delta_C = [\bar{Y}^2 + \gamma\{S_y^2 + B_{y.x2(l)}\rho_{y.x2(l)}S_yS_{x2} - B_{y.x1(l)}S_{x_1}\rho_{y.x1(l)}S_yS_{x_1}$$
$$+ B_{y.x1(l)}B_{y.x2(l)}\rho_{x1.x2}S_{x1}S_{x2}\}],$$

$$\delta_D = \delta_E = \bar{Y}^2.$$

By partially differentiating equation (5.10) w.r.t. ω_1 and ω_2, we obtained the optimum values as given by

$$\omega_1^{opt} = \left[\frac{\delta_B\delta_D - \delta_C\delta_E}{\delta_A\delta_B - \delta_C^2}\right] \quad \text{and} \quad \omega_2^{opt} = \left[\frac{\delta_A\delta_E - \delta_C\delta_D}{\delta_A\delta_B - \delta_C^2}\right].$$

Substitution of ω_1^{opt} *and* ω_2^{opt} in equation (5.10) provides the minimum *MSE* of $\hat{y}_{p_l}$ as,

$$MSE_{min}(\hat{\bar{y}}_{p_i}) = \left[\bar{Y}^2 - \frac{\delta_B\delta_D^2 - 2\delta_C\delta_D\delta_E + \delta_A\delta_E^2}{\delta_A\delta_B - \delta_C^2}\right]. \quad (5.14)$$

5.4 Numerical illustrations

5.4.1 Real life applications

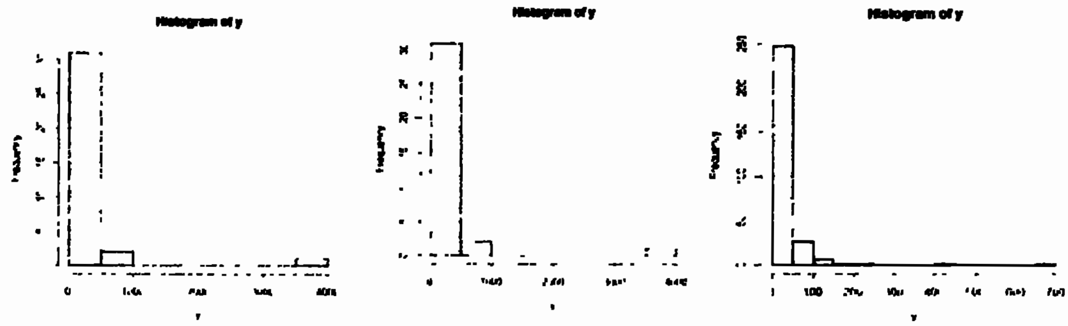
For performance evaluation between proposed and existing estimator, we used three real-life data sets. The necessary information about the data, which is used for this study, is given below:

First Population (Pop-1): We consider this data from Cochran (1977). Where, Y is taken as "Number of placebo children", X_1 is taken as "Number of paralytic polio cases in the inoculated group" and X_2 is taken as "Number of paralytic polio cases in the not inoculated group". **Fig. (5.1a)** depicts non-normality. Box-plots with Scatter-plots are in **Figures. (5.2(a), 5.3(a)** and **5.4(a))**, locating the outliers, separately and combine in Y, X_1 *and* X_2 respectively.

Second Population (Pop-2): We consider the data set of Sarndal et al. (1992). Where, Y is taken as "P85 i.e. 1985 population in thousands", X_1 is taken as "RMT85 i.e. revenues from 1985 municipal taxation (in millions of kronor)" and X_2 is taken as "SS82 i.e. number of Social-Democratic seats in municipal council". **Fig. (5.1b)** displays non-normality. Box-plots with Scatter-plots are in **Figures. (5.2(b), 5.3(b)** and **5.4(b))**, pointing out the outliers, individually and combine in Y, X_1 *and* X_2 respectively.

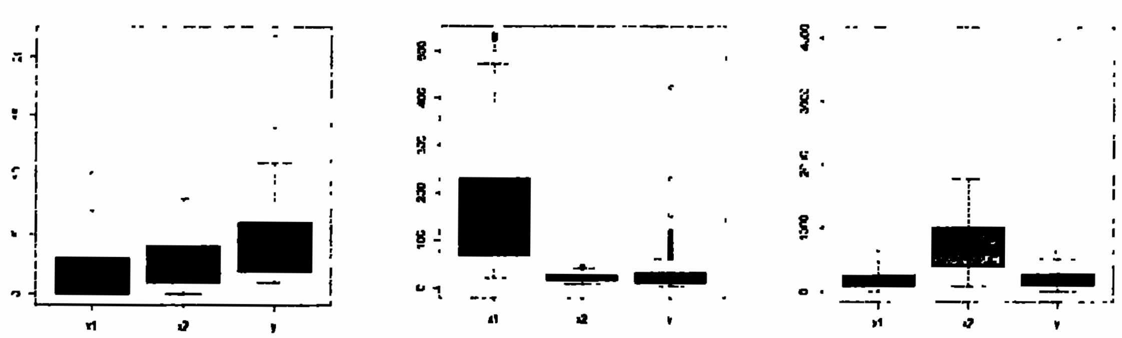
Third Population (Pop-3): We consider the data set of Sukhatme and Sukhatme (1970). Where, Y is taken as "area (acres) under wheat in 1937", X_1 is taken as "area (acres) under wheat in 1936", and X_2 is taken as "total cultivated area (acres) in 1931". **Fig. (5.1c)** displays non-normality. Box-plots and Scatter-plots in **Figures. (5.2(c),**

5.3(c) and 5.4(c)), pointing out the presence of outliers, individually and combine in Y, X_1 and X_2 respectively.



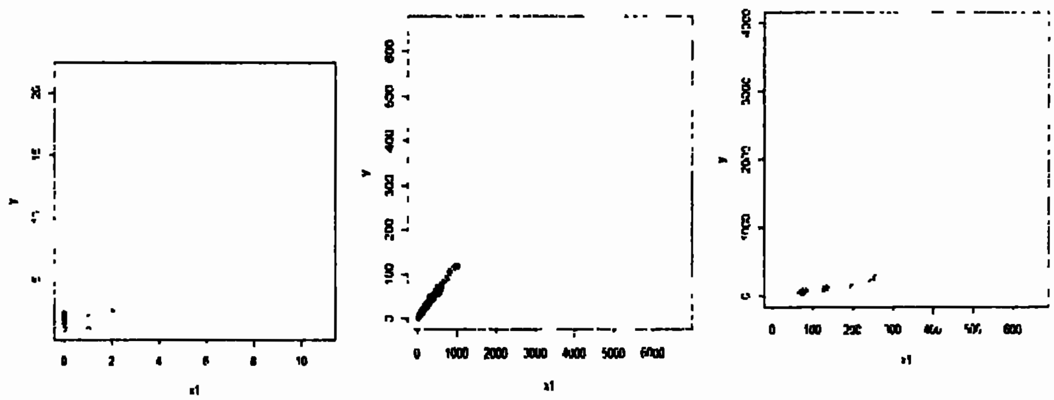
(a) (b) (c)

Fig. 5.1: Histogram Pop-1, Pop-2 and Pop-3.



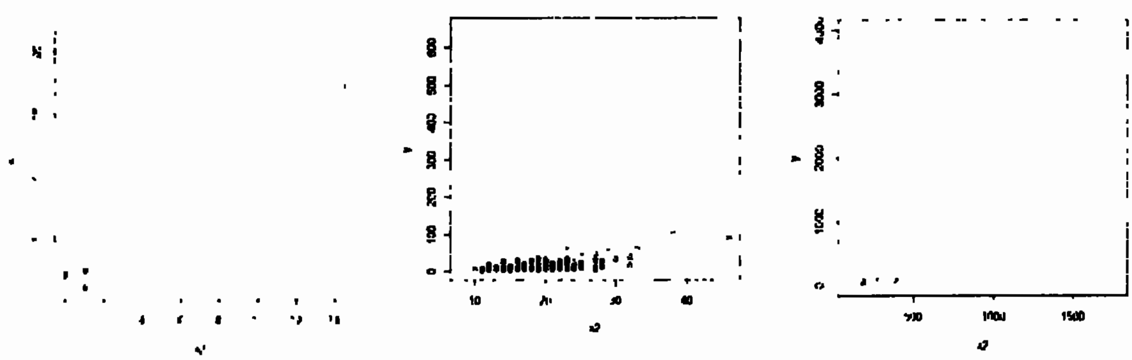
(a) (b) (c)

Fig. 5.2: Box plot Pop-1, Pop-2 and Pop-3.



(a) (b) (c)

Fig. 5.3: Plot (x_1, y) Pop-1, Pop-2 and Pop-3.



(a) (b) (c)

Fig. 5.4: Plot (x_2, y) Pop-1, Pop-2 and Pop-3.

All the referenced populations, see, **Figures. 5.1(a, b, and c)**, are a non-normal behaviors with the presence of outliers, so these are suitable for non-traditional measures as shown by Abid et al. (2018), and for the proposed class containing robust regression tools. All the left characteristics for these three population are given in **Table (5.2)**.

Table 5.2: Characteristics of Populations.

Char.	Pop-1	Pop-2	Pop-3
N	34	284	34
n	10	35	10
$\bar{Y}$	4.923529	29.36268	307.2941
$\bar{X}_1$	2.588235	245.088	218.4118
$\bar{X}_2$	2.911765	22.18662	765.3529
C_1	1.023331	1.75586	2.176777
C_{x_1}	1.233278	2.433136	0.7678148
C_{x_2}	1.148006	0.3267727	0.6169129
ρ_{yx_1}	0.7328235	5 0.9606978	0.4143947
ρ_{yx_2}	0.6426412	0.4748835	0.3906281
$\rho_{x_1x_2}$	0.6837759	0.4007188	0.8307546
$\beta_2(x_1)$	1.756713	87.74826	0.5274551
$\beta_2(x_2)$	2.950253	0.4405829	0.1002026
$M.R_1$	5.5	3370.5	334
$M.R_2$	7	27	933
$T.M_1$	1.5	131.25	162.25
$T.M_2$	2.25	21.5	705.25
$H.L_1$	2	144.5	190
$H.L_2$	2	21.5	718.5
$D.M_1$	2.133333	164.3333	206.4222
$D.M_2$	2.488889	21.76667	749.3333
$B_{yx_1}(lad)$	1.357143	0.1122449	0.9305556
$B_{yx_2}(lad)$	0.86	1.5	0.2821589
$B_{y.x_1}(hbm)$	1.241429	0.1110843	0.964638
$B_{yx_2}(hbm)$	0.8374021	1.836543	0.2892683
$B_{yx_1}(hpm)$	1.301028	0.1207181	0.9888889
$B_{yx_2}(hpm)$	0.848955	1.814633	0.2803501
$B_{yx_1}(tky)$	1.509991	0.121492	0.9697517
$B_{yx_2}(tky)$	0.5862744	1.177127	0.2773134

To indicate the domination of the proposed estimators over the existing estimators, we have found the percentage relative efficiency (PRE) for the three considered populations. The results (PRE) of proposed class w.r.t. the numerical findings of Abid et al. (2018) in **Tables (5.3 - 5.5)**. It is clearly show, the relative efficiency of the proposed estimators higher than Abid et al. (2018). With the first population, see, **Table**

(5.3), the proposed estimators $\hat{y}_{p_i}$; $i = 1,\ldots,4$ achieve more efficient and appears $\hat{y}_{p_1}$ to be the best. Also, by examining the PRE values w.r.t. $\hat{y}_{(a)_i}$ we find that:

$$PRE(\hat{y}_{P1}) > PRE(\hat{y}_{p_3}) > PRE(\hat{y}_{p_4}) > PRE(\hat{y}_{p_2}).$$

The proposed estimator $\hat{y}_{P1}$ achieve high efficiency with all $\hat{y}_{a(i)}$, and the five largest value of the efficiency were associated, respectively, with $\hat{y}_{a(13)}, \hat{y}_{a(1)}, \hat{y}_{a(9)}, \hat{y}_{a(16)}$, and $\hat{y}_{a(5)}$. In addition, these results of the highest efficiency are stable with other proposed estimators$(\hat{y}_{p_2}, \hat{y}_{p_3}, \hat{y}_{p_4})$.

Table 5.3: The PRE of $\hat{y}_{p_i}$, $\hat{y}_{(i)}$ and $\hat{y}_{a(i)}$ in Pop-1.

Est.	$\hat{y}_{P1}$	$\hat{y}_{p2}$	$\hat{y}_{p3}$	$\hat{y}_{p4}$
$\hat{y}_{a(1)}$	140.0776	138.1485	139.2744	138.4545
$\hat{y}_{a(2)}$	104.1546	102.7203	103.5574	102.9478
$\hat{y}_{a(3)}$	107.2748	105.7974	106.6597	106.0318
$\hat{y}_{a(4)}$	108.2625	106.7716	107.6418	107.0081
$\hat{y}_{a(5)}$	115.1965	113.6101	114.5360	113.8617
$\hat{y}_{a(6)}$	103.6845	102.2567	103.0901	102.4832
$\hat{y}_{a(7)}$	102.6522	101.2385	102.0636	101.4627
$\hat{y}_{a(8)}$	102.3520	100.9425	101.7652	101.1661
$\hat{y}_{a(9)}$	131.5258	129.7145	130.7717	130.0018
$\hat{y}_{a(10)}$	102.7707	101.3554	102.1815	101.5799
$\hat{y}_{a(11)}$	103.8510	102.4208	103.2556	102.6477
$\hat{y}_{a(12)}$	105.1386	103.6907	104.5358	103.9204
$\hat{y}_{a(13)}$	150.6142	148.5400	149.7506	148.8691
$\hat{y}_{a(14)}$	109.4513	107.9440	108.8237	108.1831
$\hat{y}_{a(15)}$	114.0503	112.4797	113.3964	112.7288
$\hat{y}_{a(16)}$	116.8215	115.2127	116.1517	115.4679

With the second population, see, **Table (5.4)**, the proposed estimators $\hat{y}_{p_i}$; $i = 1,\ldots,4$ achieve more efficient and $\hat{y}_{p_3}$ is the best among all. Also, by examining the PRE values w.r.t. $\hat{y}_{a(i)}$ we find that:

$$PRE(\hat{y}_{P1}) > PRE(\hat{y}_{p_3}) > PRE(\hat{y}_{p_4}) > PRE(\hat{y}_{p_2}).$$

The proposed estimator $\hat{y}_{P3}$ achieve high efficiency with all $\hat{y}_{a(i)}$, and the five highest values of efficiency were associated respectively with $\hat{y}_{a(1)}$, $\hat{y}_{a(13)}$, $\hat{y}_{a(9)}$, $\hat{y}_{a(3)}$ and $\hat{y}_{a(4)}$. In addition, these results of the highest efficiency are stable with other proposed estimators$(\hat{y}_{p_1}, \hat{y}_{p_2}, \hat{y}_{p_4})$.

Table 5.4: The PRE of $\hat{y}_{p_i}$, $\hat{y}_{(i)}$ and $\hat{y}_{a(i)}$ in Pop-2.

Est.	$\hat{y}_{p_1}$	$\hat{y}_{p2}$	$\hat{y}_{p_3}$	$\hat{y}_{p4}$
$\hat{y}_{a(1)}$	897.3651	908.5861	917.8452	893.1003
$\hat{y}_{a(2)}$	113.2658	114.6821	115.8508	112.7275
$\hat{y}_{a(3)}$	115.9237	117.3733	118.5694	115.3728
$\hat{y}_{a(4)}$	115.8495	117.2982	118.4935	115.2989
$\hat{y}_{a(5)}$	109.2186	110.5843	111.7112	108.6995
$\hat{y}_{a(6)}$	107.4210	108.7642	109.8726	106.9104
$\hat{y}_{a(7)}$	107.4264	108.7697	109.8781	106.9158
$\hat{y}_{a(8)}$	107.4681	108.8119	109.9208	106.9573
$\hat{y}_{a(9)}$	177.8697	180.0939	181.9291	177.0244
$\hat{y}_{a(10)}$	109.8424	111.2159	112.3493	109.3204
$\hat{y}_{a(11)}$	110.0201	111.3958	112.5310	109.4972
$\hat{y}_{a(12)}$	110.3018	111.6811	112.8192	109.7776
$\hat{y}_{a(13)}$	575.6524	582.8505	588.7902	572.9165
$\hat{y}_{a(14)}$	112.9870	114.3998	115.5656	112.4500
$\hat{y}_{a(15)}$	113.6587	115.0799	116.2527	113.1185
$\hat{y}_{a(16)}$	114.6674	116.1012	117.2843	114.1224

With the third population, see, **Table (5.5)**, the proposed estimators $\hat{y}_{p_i}$; $i = 1, \dots, 4$ achieve more efficient and $\hat{y}_{p_3}$ is the best among all. Also, by examining PRE values w.r.t. $\hat{y}_{a(i)}$ we find that:

$$PRE(\hat{y}_{P3}) > PRE(\hat{y}_{p_2}) > PRE(\hat{y}_{p_1}) > PRE(\hat{y}_{p_4}).$$

The proposed estimators $\hat{y}_{P3}$ achieve high efficiency with all $\hat{y}_{a(i)}$and the five highest values of efficiency were associated respectively with $\hat{\bar{y}}_{a(13)}, \hat{\bar{y}}_{a(1)}, \hat{\bar{y}}_{a(9)}, \hat{\bar{y}}_{a(16)}$, and $\hat{\bar{y}}_{a(15)}$. In addition, these results of the highest efficiency are stable with other proposed estimators$(\hat{\bar{y}}_{p_1}, \hat{\bar{y}}_{p_2}, \hat{\bar{y}}_{p_4})$.

Overall, this numerical illustration can support to shed light on the assets of the robust regression tools in mean estimation. In fact, it offers a clear indication that more alternatives estimators than the reviewed estimators can be valuable if a positive correlation exist between the study and two auxiliary variables and small or moderate samples are drawn from the population in presence of outliers.

Table 5.5: The PRE of $\hat{\bar{y}}_{p_i}, \bar{y}_{(i)}$ and $\bar{y}_{a(i)}$ in Pop-3.

Est.	$\hat{\bar{y}}_{p_1}$	$\hat{\bar{y}}_{p2}$	$\hat{\bar{y}}_{p_3}$	$\hat{\bar{y}}_{p4}$
$\bar{y}_{a(1)}$	135.0121	135.3746	135.7115	135.5167
$\bar{y}_{a(2)}$	129.5213	129.8691	130.1923	130.0054
$\bar{y}_{a(3)}$	130.8204	131.1716	131.4980	131.3093
$\bar{y}_{a(4)}$	131.1736	131.5258	131.8531	131.6639
$\bar{y}_{a(5)}$	123.1811	123.5118	123.8192	123.6414
$\bar{y}_{a(6)}$	123.3469	123.6780	123.9858	123.8079
$\bar{y}_{a(7)}$	123.2933	123.6243	123.9320	123.7541
$\bar{y}_{a(8)}$	123.2793	123.6103	123.9179	123.7400
$\bar{y}_{a(9)}$	135.3704	135.7339	136.0717	135.8764
$\bar{y}_{a(10)}$	129.3495	129.6968	130.0196	129.8329
$\bar{y}_{a(11)}$	130.5756	130.9261	131.2520	131.0636
$\bar{y}_{a(12)}$	131.1128	131.4649	131.7920	131.6029
$\bar{y}_{a(13)}$	140.4524	140.8295	141.1800	140.9773
$\bar{y}_{a(14)}$	132.1696	132.5244	132.8542	132.6635
$\bar{y}_{a(15)}$	134.1787	134.5389	134.8737	134.6801
$\bar{y}_{a(16)}$	134.9917	135.3541	135.6910	135.4962

5.4.2 Simulation study, Pop-4

We present the efficiencies of the estimators $t_{p1} - t_{p4}$ upon the estimators $t_{\alpha(1)} - t_{\alpha(16)}$, by conducting the simulation study. Here, S is used of the comparisons two multi-variate normal distributions (ND) for (Y, X_1, X_2) with means $(\bar{Y}, \bar{X}_1, \bar{X}_2) = (4.9, 4.9, 4.9)$ and covariance matrices given respectively by

$$\Sigma = \begin{bmatrix} 9.9 & 2.9 & 2.8 \\ 2.9 & 1.9 & 1.0 \\ 2.8 & 1.0 & 1.9 \end{bmatrix}, \qquad \rho_{yx_1} = 0.651, \qquad \rho_{yx_2} = 0.669.$$

For the utilization of robust tool, and add noise in Y, Ali et al. (2021). From these population of 8000, SRSWOR with size $n = (250, 300)$ are selected for the K^{th} sample, the estimators (t_{p_i}, t_{α_i}) are computed. In this way, for each (t_{p_i}, t_{α_i}), the MSE is determined as $(MSE(\theta) = {\sum_{k=1}^{k}(\hat{\theta}^{(k)} - \bar{Y})^2}/_{K}$, Where $(\hat{\theta}^{(k)}$ is denoting (t_{p_i}, t_{α_i}) estimators. For comparison purposes, PRE is computed. The PRE results are provided **Tables (5.6 - 5.8).**

Table 5.6: Simulation results $n = 100$

Est.	$\hat{\bar{y}}_{p1}$	$\hat{\bar{y}}_{p2}$	$\hat{\bar{y}}_{p3}$	$\hat{\bar{y}}_{p4}$
$\hat{\bar{y}}_{\alpha(1)}$	169.6223	169.0531	169.7314	169.0095
$\hat{\bar{y}}_{\alpha(2)}$	168.0350	167.4711	168.1430	167.4279
$\hat{\bar{y}}_{\alpha(3)}$	171.5989	171.0230	171.7092	170.9788
$\hat{\bar{y}}_{\alpha(4)}$	168.0623	117.2982	168.1704	167.4551
$\hat{\bar{y}}_{\alpha(5)}$	200.0640	167.4984	200.1926	199.3412
$\hat{\bar{y}}_{\alpha(6)}$	199.2537	199.3927	199.3818	198.5338
$\hat{\bar{y}}_{\alpha(7)}$	199.2537	198.5851	199.3818	198.5338
$\hat{\bar{y}}_{\alpha(8)}$	199.5578	198.8881	199.6861	198.8367
$\hat{\bar{y}}_{\alpha(9)}$	199.0180	198.3501	199.1459	198.2989
$\hat{\bar{y}}_{\alpha(10)}$	197.7178	197.0543	197.8449	197.0034
$\hat{\bar{y}}_{\alpha(11)}$	200.3690	199.6967	200.4978	199.6451
$\hat{\bar{y}}_{\alpha(12)}$	197.7679	197.1043	197.8950	197.0533
$\hat{\bar{y}}_{\alpha(13)}$	179.0312	178.4305	179.1463	178.3844

$\hat{y}_{a(14)}$	177.4091	176.8138	177.5232	176.7681
$\hat{y}_{a(15)}$	180.9575	180.3503	181.0739	180.3037
$\hat{y}_{a(16)}$	177.4460	176.8505	177.5600	176.8048

Table 5.7: Simulation results for $n = 150$.

Est.	$\hat{\hat{y}}_{p_1}$	$\hat{\hat{y}}_{p2}$	$\hat{\hat{y}}_{p_3}$	$\hat{\hat{y}}_{p4}$
$\hat{y}_{a(1)}$	167.4462	167.0432	167.6674	167.0070
$\hat{y}_{a(2)}$	165.8972	165.4979	166.1164	165.4621
$\hat{y}_{a(3)}$	169.4416	169.0338	169.6654	168.9972
$\hat{y}_{a(4)}$	165.9139	165.5145	166.1330	165.4787
$\hat{y}_{a(5)}$	199.4046	198.9247	199.6680	198.8816
$\hat{y}_{a(6)}$	198.6101	198.1320	198.8724	198.0891
$\hat{y}_{a(7)}$	198.6101	198.1320	198.8724	198.0891
$\hat{y}_{a(8)}$	198.8770	198.3983	199.1397	198.3553
$\hat{y}_{a(9)}$	196.3419	195.8693	196.6013	195.8269
$\hat{y}_{a(10)}$	195.0813	194.6118	195.3390	194.5696
$\hat{y}_{a(11)}$	197.7850	197.3089	198.0463	197.2662
$\hat{y}_{a(12)}$	195.1085	194.6389	195.3663	194.5968
$\hat{y}_{a(13)}$	176.7075	176.2821	176.9409	176.2440
$\hat{y}_{a(14)}$	175.1244	174.7029	175.3558	174.6651
$\hat{y}_{a(15)}$	178.6755	178.2454	178.9115	178.2068
$\hat{y}_{a(16)}$	175.1469	174.7253	175.3782	174.6875

Table 5.8: Simulation results for $n = 200$.

Est.	$\hat{\hat{y}}_{p_1}$	$\hat{\hat{y}}_{p2}$	$\hat{\hat{y}}_{p_3}$	$\hat{\hat{y}}_{p4}$
$\hat{y}_{a(1)}$	172.0687	171.7234	172.4609	171.6853
$\hat{y}_{a(2)}$	170.4591	170.1170	170.8476	170.0793
$\hat{y}_{a(3)}$	174.2228	173.8731	174.6199	173.8346
$\hat{y}_{a(4)}$	170.4636	170.1214	170.8521	170.0837
$\hat{y}_{a(5)}$	206.6926	206.2777	207.1636	206.2320
$\hat{y}_{a(6)}$	205.8665	205.4533	206.3357	205.4078
$\hat{y}_{a(7)}$	205.8665	205.4535	206.3357	205.4078
$\hat{y}_{a(8)}$	206.1327	205.7189	206.6024	205.6733
$\hat{y}_{a(9)}$	202.2088	201.8029	202.6696	201.7582

$\hat{y}_{a(10)}$	200.9173	200.5140	201.3752	200.4696
$\hat{y}_{a(11)}$	203.8269	203.4178	204.2914	203.3727
$\hat{y}_{a(12)}$	200.9219	200.5186	201.3798	200.4742
$\hat{y}_{a(13)}$	181.7720	181.4071	182.1862	181.3669
$\hat{y}_{a(14)}$	180.1319	179.7703	180.5424	179.7305
$\hat{y}_{a(15)}$	183.9101	183.5409	184.3292	183.5002
$\hat{y}_{a(16)}$	180.1391	179.7775	180.5496	179.7377

5.5 Conclusion of the chapter

Usually, in survey sampling, it uses additional information to obtain enhanced designs and further accurate estimators. In this paper, based on information related to two auxiliary variables, a new class of robust regression estimators for population mean has been proposed getting motivation from ratio type estimators of Zaman (2019), Zaman and Bulut (2019a) and Bulut and Zaman (2019). The proposed estimators are an extension of Abid et al. (2018) work and rely on robust regression tools. Three real-life data sets attributed to Sukhatme and Sukhatme (1970); Cochran (1977) and Sarndal et al. (1992) pointing out the presence of outliers, have been considered in the numerical illustration. It is observed that the values of the PRE's of the proposed estimators are higher than those for the existing estimators attributed to Abid et al. (2018) for all three real-life data sets. This means that the proposed estimators are more efficient than the ones under comparison. In addition, robustness to outliers is an additional feature of the proposed estimators. Consequently, we recommended using the proposed estimators over the existing estimators attributed to Abid et al. (2018), especially in the presence of unusual data observations.

Chapter 6

Mean estimation under stratified sampling with some transformations

6.1 Background

Estimation of population mean in literature is done by different sampling techniques. In modern surveys, Stratified sampling is used to improve the precision in estimation. Stratified sampling can be done by combined and separate ratio estimators. Our current research is based on the modification of the combination of ratio and product estimators, for estimating the population mean under Stratified random sampling scheme. We used auxiliary information to improve precision of estimates and get more efficient results. Some known parameters of auxiliary variable X such as coefficient of variation, C_x, coefficient of kurtosis, $\beta_{2(x)}$, etc. has used for purposes of the research. Different existing ratio and product type estimators in Stratified Sampling have been considered for comparison with the proposed estimators. The numerical results of the real-life data set support the theoretical findings.

In this chapter, we have proposed some estimators of finite population mean by using the transformations of coefficient of variation, C_x, coefficient of kurtosis, $\beta_{2(x)}$ and some real numbers in Stratified random sampling without replacement scheme. The bias and MSE of the proposed estimator are also obtained up to first order approximation. It is observed that the proposed estimators are efficient than the traditional mean, ratio, Bahl and tuteja (1991); Koyuncu and Kadilar (2009) and Singh and Solanki (2013), estimators. We have applied a real data set using Stratified random sampling technique for measuring the efficiency of the estimators considered here.

Use of auxiliary information improves precision of estimates and efficiency of estimators in estimation process. Some researchers used different transformations for some known parameters of auxiliary variable X such as coefficient of variation, standard deviation, correlation coefficient, coefficient of kurtosis, skewness, etc. For traditional mean estimators see, Bahl and tuteja (1991), traditional regression, Koyuncu and Kadilar (2009); Singh and Solanki (2013) and Solanki and Singh (2014) estimators are given below.

Suppose a finite population $\zeta = \left[u_1^*, u_2^*, u_3^*, \ldots, u_N^*\right]$ of size N. Let Y be taken as study variable and X be taken as auxiliary variable having values y_i and x_i in unit u_i^* ($i = 1, 2, \ldots, N$). A n_h random sample drawn without replacement from N_h population in stratum h, ($h = 1, 2, \ldots, L$). Where $\overline{X}, \overline{Y}$ are the population means of x and y respectively, Assuming population of size N, is divided into L strata containing N_h units, where ($h = 1, 2, \ldots, L$) such as $\sum_{h=1}^{L} N_h = N$ and $\sum_{h=1}^{L} n_h = n$,

with,

$$\bar{x}_{st} = \sum_{h=1}^{L} W_h \bar{x}_h \quad \text{and} \quad \bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h .$$

$$\lambda_0 = \frac{\overline{y_{st}} - \overline{Y}}{\overline{Y}}, \qquad E(\lambda_0) = 0, \qquad E\left(\lambda_0^2\right) = \sum_{h=1}^{L} W_h^2 f_h' \frac{S_{yh}^2}{\overline{Y}^2} = \eta_0,$$

$$\lambda_1 = \frac{\overline{x_{st}} - \overline{X}}{\overline{X}}, \qquad E(\lambda_1) = 0, \qquad E\left(\lambda_1^2\right) = \sum_{h=1}^{L} W_h^2 f_h' \frac{S_{xh}^2}{\overline{X}^2} = \eta_1,$$

$$E(\lambda_0\lambda_1)=\sum_{h=1}^{L}W_h^2 f_h' \frac{S_{yxh}}{\overline{X}\overline{Y}}=\eta_{01},$$

$$f_h=\frac{n_h}{N_h}, \qquad f_h'=\left(\frac{1-f_h}{n_h}\right).$$

Some other notations are used in estimators as given below:

$$\Omega_1=\sum_{h=1}^{L}W_h S_{xh}, \quad \Omega_2=\sum_{h=1}^{L}W_h C_{xh}, \quad \Omega_3=\sum_{h=1}^{L}W_h \beta_{1h}(x)$$

$$\Omega_4=\sum_{h=1}^{L}W_h \beta_{2h}(x), \quad \Omega_5=\sum_{h=1}^{L}W_h \rho_h, \quad \Omega_6=\sum_{h=1}^{L}W_h \Delta_h(x).$$

Mean estimators under stratified random sampling

$$\bar{y}_{st}=\sum_{h=1}^{L}W_h \bar{y}_h$$

where,

$$\bar{y}_h=\frac{1}{n_h}\sum_{i=1}^{n_h}y_{hi}.$$

Variance of unbiased sample mean is

$$\operatorname{var}\left(\bar{y}_{st}\right)=\overline{Y}^2\eta_1. \tag{6.1}$$

Ratio estimator under stratified random sampling

$$G_{R(st)}=\bar{y}_{st}\left(\frac{\overline{X}}{\bar{x}_{st}}\right).$$

The expressions of bias and MSE of Ratio estimators are

$$bias\left(G_{R(st)}\right)=\overline{Y}\left(\eta_0-\eta_{01}\right)$$

$$MSE\left(G_{R(st)}\right)=\bar{Y}^2\left(\eta_0+\eta_1-2\eta_{01}\right). \quad (6.2)$$

Bahl and Tuteja (1991) combined ratio exponential estimator under stratified random sampling as:

$$G_{Bt(st)}=\bar{y}_{st}\exp\left(\frac{\bar{X}-\bar{x}_{st}}{\bar{X}+\bar{x}_{st}}\right). \quad (6.3)$$

The expressions of bias and MSE of Bahl and Tuteja (1991) exponential estimator are

$$bias\left(G_{Bt(st)}\right)=\bar{Y}\left(\frac{3}{8}\eta_0-\frac{1}{2}\eta_{01}\right). \quad (6.4)$$

$$MSE\left(G_{Bt(st)}\right)=\bar{Y}^2\left(\eta_0+\frac{1}{4}\eta_1-\eta_{01}\right). \quad (6.5)$$

Traditional regression estimator under stratified sampling as:

$$G_{\mathrm{Reg}(st)}=\bar{y}_{st}+b_{st}\left(\bar{X}-\bar{x}_{st}\right). \quad (6.6)$$

The expression of MSE of regression estimator are:

$$MSE\left(G_{\mathrm{Reg}(st)}\right)=\bar{Y}^2\eta_1\left(1-\rho_{st}^2\right). \quad (6.7)$$

where ,

$$\rho_{st}=\frac{\eta_{01}}{\sqrt{\eta_0}\sqrt{\eta_1}}$$

Family of ratio estimators of Koyuncu and Kadilar (2009) in stratified random sampling

$$K_i=\phi\bar{y}_{st}\left[\frac{a_{st}\bar{X}+b_{st}}{\alpha\left(a_{st}\bar{x}_{st}+b_{st}\right)+(1-\alpha)\left(a_{st}\bar{X}+b_{st}\right)}\right]^g. \quad \text{for, } i=1,2,3,\dots,9 \quad (6.8)$$

For the family of estimators in table, MSE and bias can be expressed as:

$$bias(\mathrm{K}) = \phi\bar{Y}\left[\frac{g(g+1)}{2}\alpha^2 v^2 \eta_1 - g\alpha v \eta_{01}\right] + \bar{Y}(\phi - 1) \cdot \tag{6.9}$$

$$\mathrm{MSE(K)} = \bar{Y}^2 \begin{Bmatrix} \phi^2\eta_0 + \left(\phi^2\left(2g^2+g\right) - \phi\left(g^2+g\right)\right)\alpha^2 v^2 \eta_1 \\ -2g\alpha v\left(2\phi^2 - \phi\right)\eta_{01} + (\phi-1)^2 \end{Bmatrix}, \tag{6.10}$$

where ,

$$A = \left(g^2 + g\right)\alpha^2 v^2 \eta_1 - 2g\alpha v \eta_{01} + 2,$$

$$B = \eta_0 + \left(2g^2 + g\right)\alpha^2 v^2 \eta_1 - 4g\alpha v \eta_{01} + 1.$$

Table 6.1: Family of ratio estimators of Koyuncu and Kadilar (2009) in stratified random sampling.

ratio estimators $g=1$, $\alpha=1$	*ast*	*bst*
K_1	1	0
K_2	1	Ω_2
K_3	Ω_4	Ω_2
K_4	Ω_2	Ω_4
K_5	1	Ω_1
K_6	Ω_3	Ω_1
K_7	Ω_4	Ω_1
K_8	1	Ω_5
K_9	1	Ω_4

Minimum MSE is

$$MSE_{\min}(K) = \bar{Y}^2\left[1 - \frac{A^2}{4B}\right]. \tag{6.11}$$

Family of estimators of Singh and Solanki (2013) in stratified random sampling

$$T_{SSr} = \left[\phi_1 \bar{y}_{st}\left\{\frac{\alpha\left(a_{st}\bar{x}_{st} + b_{st}\right) + (1-\alpha)\left(a_{st}\bar{X} + b_{st}\right)}{\left(a_{st}\bar{X} + b_{st}\right)}\right\}^{\delta} + \phi_2 \bar{y}_{st}\left\{\frac{\left(a_{st}\bar{X} + b_{st}\right)}{\alpha\left(a_{st}\bar{x}_{st} + b_{st}\right) + (1-\alpha)\left(a_{st}\bar{X} + b_{st}\right)}\right\}^{g}\right]. \tag{6.12}$$

for $r = 1, 2, 3, \ldots, 17$ and $(g=1, \delta=0, \alpha=1)$

For the family of estimators in table, MSE and bias can be expressed as:

$$bias(T_{SS})=\bar{Y}\left[\phi_1\left\{1+\alpha\delta\upsilon\eta_{01}+\frac{\delta(\delta-1)}{2}\alpha^2\upsilon^2\eta_1\right\}+\phi_2\left\{1-\alpha\delta\upsilon\eta_{01}+\frac{g(g+1)}{2}\alpha^2\upsilon^2\eta_1\right\}-1\right], \tag{6.13}$$

where,

$$\upsilon=\frac{a_{st}\bar{X}}{a_{st}\bar{X}+b_{st}}.$$

$$MSE(T_{SS})=\bar{Y}^2\left[1+\phi_1^2C+\phi_2^2B+2\phi_1\phi_2D-2\phi_1E-2\phi_2A\right], \tag{6.14}$$

where,

$$A_{SS}=\left[1-\alpha g\upsilon\eta_{01}+\frac{g(g+1)}{2}\alpha^2\upsilon^2\eta_1\right],$$

$$B_{SS}=\left[1+\eta_0-4\alpha g\upsilon\eta_{01}+g(2g+1)\alpha^2\upsilon^2\eta_1\right],$$

$$C_{SS}=\left[1+\eta_0+4\alpha g\upsilon\eta_{01}+g(2g-1)\alpha^2\upsilon^2\eta_1\right],$$

$$D_{SS}=\left[1+\eta_0+2\alpha(\delta-g)\upsilon\eta_{01}+\left(\frac{\alpha^2\upsilon^2}{2}\right)(\delta-g)(\delta-g-1)\eta_1\right],$$

$$E_{SS}=\left[1-\alpha\delta\upsilon\eta_{01}+\frac{\delta(\delta-1)}{2}\alpha^2\upsilon^2\eta_1\right].$$

Differentiating MSE partially, with respect to $\emptyset_1$and $\emptyset_2$ and equating to zero, we get the following optimum values of $\emptyset_1$and $\emptyset_2$.

$$\phi_{1(opt)}=\frac{(B_{SS}E_{SS}-A_{SS}D_{SS})}{\left(B_{SS}C_{SS}-D_{ss}^2\right)}, \qquad \phi_{2(opt)}=\frac{(A_{SS}C_{SS}-D_{SS}E_{SS})}{\left(B_{SS}C_{SS}-D_{st}^2\right)}.$$

Table 6.2: Family of estimators of Singh and Solanki (2013) in stratified sampling.

Estimators	ast	bst	Estimators	ast	Bst
T_{ss1}	1	0	T_{ss10}	1	Ω_6
T_{ss2}	1	Ω_2	T_{ss11}	Ω_6	Ω_2
T_{ss3}	Ω_4	Ω_2	T_{ss12}	Ω_2	Ω_6
T_{ss4}	Ω_2	Ω_4	T_{ss13}	Ω_6	Ω_1
T_{ss5}	1	Ω_1	T_{ss14}	Ω_5	Ω_6
T_{ss6}	Ω_3	Ω_1	T_{ss15}	Ω_6	Ω_5
T_{ss7}	Ω_4	Ω_1	T_{ss16}	Ω_5	Ω_1
T_{ss8}	1	Ω_5	T_{ss17}	Ω_2	Ω_1
T_{ss9}	1	Ω_4			

Minimum MSE is

$$MSE_{\min}(T_{SS})=\bar{Y}^2\left[1-\frac{\left(B_{SS}E_{ss}^2-2A_{SS}D_{SS}E_{SS}+A_{ss}^2C_{SS}\right)}{\left(B_{SS}C_{SS}-D_{ss}^2\right)}\right]. \tag{6.15}$$

Family of estimators of Solanki and Singh (2014) in stratified random sampling

$$T_{SK_q}=\phi_1\bar{y}_{st}\left[\frac{\bar{X}^*}{\alpha\bar{x}_{st}^*+(1-\alpha)\bar{X}^*}\right]^g+\phi_2\bar{y}_{st}\exp\left[\frac{\delta\left(\bar{X}^*-\bar{x}_{st}^*\right)}{\left(\bar{X}^*+\bar{x}_{st}^*\right)}\right]. \tag{6.16}$$

where, $\bar{X}^*=\left(a_{st}\bar{X}+b_{st}\right)$, $\bar{x}_{st}^*=\left(a_{st}\bar{x}_{st}+b_{st}\right)$ $\quad for\ q=1,2,3,\dots,17$

and $(g=0,\ \delta=1)$

For the family of estimators in table, MSE and bias can be expressed as:

$$bias(T_{SK})=\bar{Y}\left[\phi_1\left\{1+\left(\frac{g\alpha\upsilon}{2}\right)\eta_1\left(\alpha\upsilon(g+1)-2\frac{\eta_{01}}{\eta_1}\right)\right\}+\phi_2\left\{1+\left(\frac{\delta\upsilon}{8}\right)\eta_1\left(\upsilon(\delta+2)-4\frac{\eta_{01}}{\eta_1}\right)\right\}-1\right]. \tag{6.17}$$

$$MSE(T_{SK})=\bar{Y}^2\left[1+\phi_1^2A_{SK}+\phi_2^2C_{SK}+2\phi_1\phi_2D_{SK}-2\phi_1B_{SK}-2\phi_2E_{SK}\right]. \tag{6.18}$$

Table 6.3: Family of estimators of Singh and Solanki (2014) in SRS.

Estimators	ast	bst	Estimators	ast	bst
T_{sk1}	1	0	T_{sk7}	1	Ω_5
T_{sk2}	1	Ω_2	T_{sk8}	1	Ω_4
T_{sk3}	Ω_4	Ω_2	T_{sk9}	Ω_2	Ω_5
T_{sk4}	Ω_2	Ω_4	T_{sk10}	Ω_5	Ω_2
T_{sk5}	1	Ω_1	T_{sk11}	Ω_4	Ω_5
T_{sk6}	Ω_4	Ω_1	T_{sk12}	Ω_5	Ω_4

where,

$$A_{SK}=\left[1+\eta_0+\alpha^2\upsilon^2\left(2g^2+g\right)\eta_1-4\alpha g\upsilon\eta_{01}\right],$$

$$B_{SK}=\left[1+\frac{\alpha^2\upsilon^2\left(g^2+g\right)}{2}\eta_1-\alpha g\upsilon\eta_{01}\right],$$

$$C_{SK}=\left[1+\eta_0+\frac{\upsilon^2\left(\delta^2+\delta\right)}{2}\eta_1-2\delta\upsilon\eta_{01}\right],$$

$$D_{SK}=\left[1+\eta_0+\left(\frac{A^*\upsilon^2}{8}\right)\eta_1-2\upsilon\left(2\alpha g+\delta\right)\eta_{01}\right],$$

$$E_{SK}=\left[1-\left\{\frac{\left(\delta^2+2\delta\right)}{8}\upsilon^2\right\}\eta_1+\left(\frac{\delta\upsilon}{2}\right)\eta_{01}\right],$$

where,

$$A^*=\left[\left(2ag+\delta\right)^2+2\left(2a^2g+\delta\right)\right].$$

Differentiating MSE partially, with respect to $\emptyset_1$ and $\emptyset_2$ and equating to zero, we get the following optimum values of $\emptyset_1$ and $\emptyset_2$.

$$\phi_{1(opt)}=\frac{\left(B_{SK}C_{SK}-D_{SK}E_{SK}\right)}{\left(A_{SK}C_{SK}-D_{SK}^2\right)},\qquad \phi_{2(opt)}=\frac{\left(A_{SK}E_{SK}-B_{SK}D_{SK}\right)}{\left(B_{SK}C_{SK}-D_{SK}^2\right)}.$$

Minimum MSE is,

$$MSE_{\min}(T_{sk})=\bar{Y}^2\left[1-\frac{\left(B_{sk}^2C_{sk}-2B_{sk}D_{sk}E_{sk}+A_{sk}E_{sk}^2\right)}{\left(A_{sk}C_{sk}-D_{sk}^2\right)^2}\right]. \tag{6.19}$$

6.2 A Proposed class of estimators for the estimation of population mean

In this section, some improved estimators for finite population mean in stratified random sampling are proposed, which are based on first order approximation. Formulation of the proposed estimators are explained step-by-step as follows:

6.2.1 First proposed estimator

For the formulation of our first proposition, follows Rao (1966). The average of ratio product estimator is:

$$\bar{y}_{Rao} = \phi_1 \bar{y} + \phi_2 \left(\bar{X} - \bar{x}\right). \tag{6.20}$$

The average of ratio and product estimators is

$$\bar{y}_{w} = \frac{\bar{y}}{2}\left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}}\right). \tag{6.21}$$

By replacing $\bar{y}$ with $\bar{y}_W$ in equation (6.15)

$$\bar{y}_M = \phi_1 \bar{y}_w + \phi_2 \left(\bar{X} - \bar{x}\right). \tag{6.22}$$

By replacing $\bar{y}_M$ in Bhal and Tuteja (1991), we suggest the following estimator:

$$U_{P_1} = \bar{y}_w \exp\left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}}\right). \tag{6.23}$$

$$U_{P_1} = \left[\phi_1 \frac{\bar{y}_{st}}{2}\left(\frac{\bar{X}}{\bar{x}_{st}} + \frac{\bar{x}_{st}}{\bar{X}}\right) + \phi_2\left(\bar{X} - \bar{x}_{st}\right)\right] \exp\left[\frac{a\left(\bar{X} - \bar{x}_{st}\right)}{a\left(\bar{X} - \bar{x}_{st}\right) + 2b}\right]. \tag{6.24}$$

We used transformation of $a = \sum_{h=1}^{L} W_h C_{xh}$, $b = \sum_{h=1}^{L} W_h \beta_{2h}(x)$.

$$bias\left(U_{P_1}\right) = \phi_1 \bar{Y}\left(1 - \frac{\vartheta \eta_{01}}{2} + \alpha_1 \eta_1\right) + \phi_2 \bar{X} \frac{\vartheta^2 \eta_1}{2} - \bar{Y}, \tag{6.25}$$

where,

$$\alpha_1=\frac{1}{2}+\frac{3\vartheta^2}{8} \text{ and } \quad \vartheta=\frac{a\bar{X}}{a\bar{X}+b},$$

$$MSE\left(U_{P_1}\right)=\bar{Y}^2+\phi_1^2U_{A1(i)}+\phi_2^2U_{B1(i)}+2\phi_1\phi_2U_{C1(i)}-2\phi_1U_{D1(i)}-2\phi_2U_{E1(i)}. \tag{6.26}$$

$$MSE_{min}\left(U_{P_1}\right)=\bar{Y}^2-\frac{U_{A1(i)}U_{E1(i)}^2+U_{B1(i)}U_{D1(i)}^2-2U_{C1(i)}U_{D1(i)}U_{E1(i)}}{U_{A1(i)}U_{B1(i)}-U_{C1(i)}^2}. \tag{6.27}$$

The expressions of above proof are explained in appendix A.

6.2.2 Second proposed estimator

For the formulation of second proposed estimator our motivation from

The average of exponential ratio and product estimator is:

$$\bar{y}_{ew}=\frac{\bar{y}}{2}\left\{\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right)+\exp\left(\frac{\bar{x}-\bar{X}}{\bar{X}+\bar{x}}\right)\right\}. \tag{6.28}$$

By replacing $\bar{y}_{ew}$ with $\bar{y}_w$, in U_{p_1}, we propose the following estimator:

$$U_{P_2}=\bar{y}_{ew}\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right). \tag{6.29}$$

$$U_{P_2}=\left[\phi_1\frac{\bar{y}_{st}}{2}\left\{\exp\left(\frac{\bar{X}-\bar{x}_{st}}{\bar{X}+\bar{x}_{st}}\right)+\exp\left(\frac{\bar{x}_{st}-\bar{X}}{\bar{X}+\bar{x}_{st}}\right)\right\}+\phi_2\left(\bar{X}-\bar{x}_{st}\right)\right]\exp\left[\frac{a\left(\bar{X}-\bar{x}_{st}\right)}{a\left(\bar{X}-\bar{x}_{st}\right)+2b}\right]. \tag{6.30}$$

We used transformation of $a=\sum_{h=1}^{L}W_hC_{xh}$, $b=\sum_{h=1}^{L}W_h\beta_{2h}(x)$.

$$bias\left(U_{P_2}\right)=\phi_1\bar{Y}\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_2\eta_1\right)-\phi_2\bar{X}\frac{\vartheta^2\eta_1}{2}-\bar{Y}, \tag{6.31}$$

where,

$\alpha_2 = \frac{3}{8} + \frac{3\vartheta^2}{8}$ and $\vartheta = \frac{a\bar{X}}{a\bar{X}+b}$.

$$MSE\left(U_{P_2}\right) = \bar{Y}^2 + \phi_1^2 U_{A2(t)} + \phi_2^2 U_{B2(t)} + 2\phi_1\phi_2 U_{C2(t)} - 2\phi_1 U_{D2(t)} - 2\phi_2 U_{E2(t)} . \quad (6.32)$$

$$MSE_{\min}\left(U_{P_2}\right) = \bar{Y}^2 - \frac{U_{A2(t)}U_{E2(t)}^2 + U_{B2(t)}U_{D2(t)}^2 - 2U_{C2(t)}U_{D2(t)}U_{E2(t)}}{U_{A2(t)}U_{B2(t)} - U_{C2(t)}^2} . \quad (6.33)$$

The expressions of above proof are explained in appendix A.

6.2.3 Third proposed estimator

For the formulation of third proposed estimator our motivation from, by replacing $\bar{y}$ with $\bar{y}_w$, in U_{p_2}, we propose the following estimator:

$$U_{P_3} = \left[\phi_1 \frac{\bar{y}_{st}}{4}\left(\frac{\bar{X}}{\bar{x}_{st}} + \frac{\bar{x}_{st}}{\bar{X}}\right)\left\{\exp\left(\frac{\bar{X}-\bar{x}_{st}}{\bar{X}+\bar{x}_{st}}\right) + \exp\left(\frac{\bar{x}_{st}-\bar{X}}{\bar{X}+\bar{x}_{st}}\right)\right\} + \phi_2\left(\bar{X}-\bar{x}_{st}\right)\right]\exp\left[\frac{a\left(\bar{X}-\bar{x}_{st}\right)}{a\left(\bar{X}-\bar{x}_{st}\right)+2b}\right] . \quad (6.34)$$

We used transformation of $a = \sum_{h=1}^{L} W_h C_{xh}$, $b = \sum_{h=1}^{L} W_h \beta_{2h}(x)$.

$$bias\left(U_{P_3}\right) = \phi_1 \bar{Y}\left(1 - \frac{\vartheta\eta_{01}}{2} + \alpha_3\eta_1\right) + \phi_2 \bar{X}\frac{\vartheta^2\eta_1}{2} - \bar{Y} , \quad (6.35)$$

where,

$\alpha_3 = \frac{7}{8} + \frac{3\vartheta^2}{8}$ and $\vartheta = \frac{a\bar{X}}{a\bar{X}+b}$.

$$MSE\left(U_{P_3}\right) = \bar{Y}^2 + \phi_1^2 U_{A3(t)} + \phi_2^2 U_{B3(t)} + 2\phi_1\phi_2 U_{C3(t)} - 2\phi_1 U_{D3(t)} - 2\phi_2 U_{E3(t)} , \quad (6.36)$$

$$MSE_{\min}\left(U_{P_3}\right) = \bar{Y}^2 - \frac{U_{A3(t)}U_{E3(t)}^2 + U_{B3(t)}U_{D3(t)}^2 - 2U_{C3(t)}U_{D3(t)}U_{E3(t)}}{U_{A3(t)}U_{B3(t)} - U_{C3(t)}^2} . \quad (6.37)$$

The expressions of above proof are explained in appendix A.

6.2.4 Fourth proposed estimator

For the formulation of fourth proposed estimator our motivation from Usman and Hanif (2018) and U_{p_1}, we propose the following estimator:

$$U_{p_4}=\left[\phi_1\frac{\overline{y_{st}}}{2}\left(\frac{\overline{X}}{\overline{x_{st}}}+\frac{\overline{x_{st}}}{\overline{X}}\right)+(1-2\phi_2)\left(\overline{X}-\overline{x_{st}}\right)\right]\exp\left[\frac{a\left(\overline{X}-\overline{x_{st}}\right)}{a\left(\overline{X}-\overline{x_{st}}\right)+2b}\right]. \tag{6.38}$$

$$for\ \ a\ =\ 0\ \ and\, b\ =1$$

$$bias\left(U_{p_4}\right)=\phi_1\overline{Y}\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_4\eta_1\right)+(1-2\phi_2)\overline{X}\frac{\vartheta^2\eta_1}{2}-\overline{Y}, \tag{6.39}$$

where,

$$\alpha_4=\frac{1}{2}+\frac{3\vartheta^2}{8}\qquad \text{and} \qquad \vartheta=\frac{a\overline{X}}{a\overline{X}+b}.$$

$$MSE\left(U_{p_4}\right)=\left(\overline{Y}^2+\overline{X}^2\eta_1-\overline{X}\overline{Y}\vartheta\eta_1\right)+\phi_1^2U_{A4(t)}+\phi_2^2U_{B4(t)}+2\phi_1\phi_2U_{C4(t)}-2\phi_1U_{D4(t)}-2\phi_2U_{E4(t)}, \tag{6.40}$$

$$MSE_{\min}\left(U_{p_4}\right)=\left(\overline{Y}^2+\overline{X}^2\eta_1-\overline{X}\overline{Y}\vartheta\eta_1\right)-\frac{U_{A4(t)}U_{E4(t)}^{\ 2}+U_{B4(t)}U_{D4(t)}^{\ 2}-2U_{C4(t)}U_{D4(t)}U_{E4(t)}}{U_{A4(t)}U_{B4(t)}-U_{C4(t)}^{\ 2}}. \tag{6.41}$$

The expressions of above proof are explained in appendix A.

6.2.5 Fifth proposed estimator

For the formulation of fifth proposed estimator our motivation from Usman and Hanif (2018) and U_{p_2}, we propose the following estimator:

$$U_{p_5}=\left[\phi_1\frac{\overline{y_{st}}}{2}\left\{\exp\left(\frac{\overline{X}-\overline{x_{st}}}{\overline{X}+\overline{x_{st}}}\right)+\exp\left(\frac{\overline{x_{st}}-\overline{X}}{\overline{X}+\overline{x_{st}}}\right)\right\}+(1-2\phi_2)\left(\overline{X}-\overline{x_{st}}\right)\right]\exp\left[\frac{a\left(\overline{X}-\overline{x_{st}}\right)}{a\left(\overline{X}-\overline{x_{st}}\right)+2b}\right]. \tag{6.42}$$

for, $a\ =0\ and\ b=1$

$$bias(U_{P_5}) = \phi_1\bar{Y}\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_5\eta_1\right)+(1-2\phi_2)\bar{X}\frac{\vartheta^2\eta_1}{2}-\bar{Y}, \tag{6.43}$$

where,

$$\alpha_5 = \frac{3}{8}+\frac{3\vartheta^2}{8} \quad \text{and} \quad \vartheta=\frac{a\bar{X}}{a\bar{X}+b}.$$

$$MSE(U_{P_5}) = \left(\bar{Y}^2+\bar{X}^2\eta_1-\bar{X}\bar{Y}\vartheta\eta_1\right)+\phi_1^2U_{A5(i)}+\phi_2^2U_{B5(i)}+2\phi_1\phi_2U_{C5(i)}-2\phi_1U_{D5(i)}-2\phi_2U_{E5(i)}, \tag{6.44}$$

$$MSE_{\min}(U_{P_5}) = \left(\bar{Y}^2+\bar{X}^2\eta_1-\bar{X}\bar{Y}\vartheta\eta_1\right)-\frac{U_{A5(i)}U_{E5(i)}^2+U_{B5(i)}U_{D5(i)}^2-2U_{C5(i)}U_{D5(i)}U_{E5(i)}}{U_{A5(i)}U_{B5(i)}-U_{C5(i)}^2}. \tag{6.45}$$

The expressions of above proof are explained in appendix A.

6.2.6 Sixth proposed estimator

For the formulation of sixth proposed estimator our motivation from Usman and Hanif (2018) and U_{p_3}, we propose the following estimator:

$$U_{P_6} = \left[\phi_1\frac{\bar{y}_{st}}{4}\left(\frac{\bar{X}}{\bar{x}_{st}}+\frac{\bar{x}_{st}}{\bar{X}}\right)\left\{\exp\left(\frac{\bar{X}-\bar{x}_{st}}{\bar{X}+\bar{x}_{st}}\right)+\exp\left(\frac{\bar{x}_{st}-\bar{X}}{\bar{X}+\bar{x}_{st}}\right)\right\}+(1-2\phi_2)(\bar{X}-\bar{x}_{st})\right]\exp\left[\frac{a(\bar{X}-\bar{x}_{st})}{a(\bar{X}-\bar{x}_{st})+2b}\right]. \tag{6.46}$$

$$for \quad a=0 \; and \; b=1$$

$$bias(U_{P_6}) = \phi_1\bar{Y}\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_6\eta_1\right)+(1-2\phi_2)\bar{X}\frac{\vartheta^2\eta_1}{2}-\bar{Y}, \tag{6.47}$$

where,

$$\alpha_6 = \frac{7}{8}+\frac{3\vartheta^2}{8} \quad \text{and} \quad \vartheta=\frac{a\bar{X}}{a\bar{X}+b}.$$

$$MSE(U_{P_6}) = \left(\bar{Y}^2+\bar{X}^2\eta_1-\bar{X}\bar{Y}\vartheta\eta_1\right)+\phi_1^2U_{A6(i)}+\phi_2^2U_{B6(i)}+2\phi_1\phi_2U_{C6(i)}-2\phi_1U_{D6(i)}-2\phi_2U_{E6(i)}, \tag{6.48}$$

$$MSE_{\min}\left(U_{P_6}\right)=\left(\bar{Y}^2+\bar{X}^2\eta_1-\bar{X}\bar{Y}\vartheta\eta_1\right)-\frac{U_{A6(i)}U_{E6(i)}{}^2+U_{B6(i)}U_{D6(i)}{}^2-2U_{C6(i)}U_{D6(i)}U_{E6(i)}}{U_{A6(i)}U_{B6(i)}-U_{C6(i)}{}^2}. \quad (6.49)$$

The expressions of above proof are explained in appendix A.

6.2.7 Seventh proposed estimator

For the formulation of seventh proposed estimator our motivation from Bhal and Tuteja (1991).

By adding $\bar{y}_w$ in $\hat{\bar{y}}_{Rao}$, we develop the following class of estimator:

$$\bar{y}_{WR}=\bar{y}_w+\bar{y}_{Rao}. \quad (6.50)$$

$$\bar{y}_{WR}=\left[\frac{\bar{y}}{2}\left(\frac{\bar{X}}{\bar{x}}+\frac{\bar{x}}{\bar{X}}\right)+\phi_1\bar{y}+\phi_2\left(\bar{X}-\bar{x}\right)\right]. \quad (6.51)$$

By replacing $\bar{y}_w$ in $\hat{\bar{y}}_{WR}$, in Bhal and Tuteja, we proposed the following estimator:

$$U_{P_7}=\bar{y}_{WR}\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right). \quad (6.52)$$

$$U_{P_7}=\left[\frac{\bar{y}_{st}}{2}\left(\frac{\bar{X}}{\bar{x}_{st}}+\frac{\bar{x}_{st}}{\bar{X}}\right)+\phi_1\bar{y}_{st}+\phi_2\left(\bar{X}-\bar{x}_{st}\right)\right]\exp\left[\frac{a\left(\bar{X}-\bar{x}_{st}\right)}{a\left(\bar{X}-\bar{x}_{st}\right)+2b}\right]. \quad (6.53)$$

We used transformation of $a=\sum_{h=1}^{L}W_hC_{xh}$, $b=\sum_{h=1}^{L}W_h\beta_{2h}(x)$.

$$bias\left(U_{P_7}\right)=\bar{Y}\left(-\frac{\vartheta\eta_1}{2}+\frac{3\vartheta^2\eta_1}{8}-\frac{\vartheta\eta_{01}}{2}+\frac{\eta_1}{2}\right)+\phi_1\bar{Y}\left(1-\frac{\vartheta\eta_{10}}{2}+\frac{\vartheta\eta_1}{2}+\frac{3\vartheta^2\eta_1}{8}\right)+\phi_2\bar{X}\frac{\vartheta\eta_1}{2}, \quad (6.54)$$

where,

$$\alpha_7=\frac{1}{2}+\frac{3\vartheta^2}{8} \quad \text{and} \quad \vartheta=\frac{a\bar{X}}{a\bar{X}+b}.$$

$$MSE\left(U_{P_7}\right)=\bar{Y}^2\left(\eta_0+\frac{\vartheta^2\eta_1}{4}-\vartheta\eta_{01}\right)+\phi_1^2U_{A7(i)}+\phi_2^2U_{B7(i)}+2\phi_1\phi_2U_{C7(i)}-2\phi_1U_{D7(i)}-2\phi_2U_{E7(i)}, \tag{6.55}$$

$$MSE_{\min}\left(U_{P_7}\right)=\bar{Y}^2\left(\eta_0+\frac{\vartheta^2\eta_1}{4}-\vartheta\eta_{01}\right)-\frac{U_{A7(i)}U_{E7(i)}^{\;2}+U_{B7(i)}U_{D7(i)}^{\;2}-2U_{C7(i)}U_{D7(i)}U_{E7(i)}}{U_{A7(i)}U_{B7(i)}-U_{C7(i)}^{\;2}}. \tag{6.56}$$

The expressions of above proof are explained in appendix A.

6.2.8 Eighth proposed estimator

For the formulation of eighth proposed estimator our motivation comes from, By replacing $\bar{y}_w$ with $\bar{y}_{ew}$, in U_{p_7}, we propose the following estimator:

$$U_{P_8}=\left[\frac{\bar{y}_{st}}{2}\left\{\exp\left(\frac{\bar{X}-\bar{x}_{st}}{\bar{X}+\bar{x}_{st}}\right)+\exp\left(\frac{\bar{x}_{st}-\bar{X}}{\bar{X}+\bar{x}_{st}}\right)\right\}+\phi_1\bar{y}_{st}+\phi_2\left(\bar{X}-\bar{x}_{st}\right)\right]\exp\left[\frac{a\left(\bar{X}-\bar{x}_{st}\right)}{a\left(\bar{X}-\bar{x}_{st}\right)+2b}\right]. \tag{6.57}$$

We used transformation of $a=\sum_{h=1}^{L}W_hC_{xh}$, $b=\sum_{h=1}^{L}W_h\beta_{2h}(x)$.

$$bias\left(U_{P_8}\right)=\bar{Y}\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_8\eta_1\right)+\phi_1\bar{Y}\left(1+\frac{3\vartheta^2\eta_1}{8}-\frac{\vartheta\eta_{01}}{2}\right)+\phi_2\bar{X}\frac{\vartheta^2\eta_1}{2}-\bar{Y}, \tag{6.58}$$

where,

$$\alpha_8=\frac{3}{8}+\frac{3\vartheta^2}{8} \quad \text{and} \quad \vartheta=\frac{a\bar{X}}{a\bar{X}+b}.$$

$$MSE\left(U_{P_8}\right)=\bar{Y}^2\left(\eta_0+\frac{\vartheta^2\eta_1}{4}-\vartheta\eta_{01}\right)+\phi_1^2U_{A8(i)}+\phi_2^2U_{B8(i)}+2\phi_1\phi_2U_{C8(i)}-2\phi_1U_{D8(i)}-2\phi_2U_{E8(i)}, \tag{6.59}$$

$$MSE_{\min}\left(U_{P_8}\right)=\bar{Y}^2\left(\eta_0+\frac{\vartheta^2\eta_1}{4}-\vartheta\eta_{01}\right)-\frac{U_{A8(i)}U_{E8(i)}^{\;2}+U_{B8(i)}U_{D8(i)}^{\;2}-2U_{C8(i)}U_{D8(i)}U_{E8(i)}}{U_{A8(i)}U_{B8(i)}-U_{C8(i)}^{\;2}}. \tag{6.60}$$

The expressions of above proof are explained in appendix A.

6.2.9 Ninth proposed estimator

For the formulation of ninth proposed estimator our motivation from, By repeating substitutions in U_{p_7}, and U_{p_8}, we propose the following estimator:

$$U_{p_9}=\left[\frac{\overline{y}_{st}}{4}\left(\frac{\overline{X}}{\overline{x}_{st}}+\frac{\overline{x}_{st}}{\overline{X}}\right)\left\{\exp\left(\frac{\overline{X}-\overline{x}_{st}}{\overline{X}+\overline{x}_{st}}\right)+\exp\left(\frac{\overline{x}_{st}-\overline{X}}{\overline{X}+\overline{x}_{st}}\right)\right\}+\phi_1\overline{y}_{st}+\phi_2\left(\overline{X}-\overline{x}_{st}\right)\right]\exp\left[\frac{a\left(\overline{X}-\overline{x}_{st}\right)}{a\left(\overline{X}-\overline{x}_{st}\right)+2b}\right]. \quad (6.61)$$

We used transformation of $a=\sum_{h=1}^{L}W_hC_{xh}$, $b=\sum_{h=1}^{L}W_h\beta_{2h}(x)$.

$$bias\left(U_{p_9}\right)=\overline{Y}\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_9\eta_1\right)+\phi_1\overline{Y}\left(1+\frac{3\vartheta^2\eta_1}{8}-\frac{\vartheta\eta_{10}}{2}\right)+\phi_2\overline{X}\frac{\vartheta^2\eta_1}{2}-\overline{Y}, \quad (6.62)$$

where,

$$\alpha_9=\frac{7}{8}+\frac{3\vartheta^2}{8} \quad \text{and} \quad \vartheta=\frac{a\overline{X}}{a\overline{X}+b}.$$

$$MSE\left(U_{p_9}\right)=\overline{Y}^2\left(\eta_0+\frac{\vartheta^2\eta_1}{4}-\vartheta\eta_{01}\right)+\phi_1^2U_{A9(i)}+\phi_2^2U_{B9(i)}+2\phi_1\phi_2U_{C9(i)}-2\phi_1U_{D9(i)}-2\phi_2U_{E9(i)}, \quad (6.63)$$

$$MSE_{\min}\left(U_{p_9}\right)=\overline{Y}^2\left(\eta_0+\frac{\vartheta^2\eta_1}{4}-\vartheta\eta_{01}\right)-\frac{U_{A9(i)}U_{E9(i)}^{\;2}+U_{B9(i)}U_{D9(i)}^{\;2}-2U_{C9(i)}U_{D9(i)}U_{E9(i)}}{U_{A9(i)}U_{B9(i)}-U_{C9(i)}^{\;2}}. \quad (6.64)$$

The expressions of above proof are explained in appendix A.

6.3 Numerical investigation

For assessing the merits of the proposed class of estimators over existing ones, we investigate numerical results by utilizing the following real data set

Source of data: Murthy (1967), p. 228

The total sample size is $n = 45$ and strata are as under:

Strata 1 = X is less than 100, Strata 2 = X is between 100 to 200, Strata 3 = X is between 200 to 500, Strata 4 = X is greater than 500.

X = Data on number of workers,

Y = Output for 80 factories in a region

Table 6.4: Descriptive Data.

Stratum	1st	2nd	3rd	4th	Total
N_h	25.00	23.00	16.00	16.00	80.00
n_h	14.00	13.00	9.00	9.00	45
$\bar{X}_h$	71.00	140.69	362.93	749.50	284.75
$\bar{Y}_h$	3156.64	4766.22	6334.19	7795.31	5182.64
$\beta_{2(x)h}$	1.75	2.19	1.61	1.90	3.53
C_{xh}	0.20	0.19	0.25	0.23	0.94
S_{xh}	14.61	28.03	91.38	174.46	270.49
S_{yh}	740.01	515.69	501.39	653.09	1835.66
S_{xy}	8830.78	11900.60	43903.70	111718.00	454033.30
ρ_h	0.81	0.8231	0.95	0.98	0.91
λ_h	0.03	0.03	0.04	0.04	
W_h	0.31	0.28	0.20	0.20	

Table 6.5: MSE values of Unbiased, ratio, exponential and regression estimators under Stratified Random Sampling.

Estimators	MSE
$\bar{y}_{st}$	6093.681
$G_{R(st)}$	5550.412
$G_{B(st)}$	2676.871
$G_{Reg(st)}$	2671.006

Table 6.6: MSE values of Koyuncu and Kadilar (2009) estimators under Stratified Random Sampling.

Estimators	MSE	Estimators	MSE
K_1	5549.451	K_6	5550.044
K_2	5550.173	K_7	5549.493
K_3	5549.815	K_8	5552.181
K_4	5584.171	K_9	5556.219
K_5	5549.533		

Table 6.7: MSE values of Singh and Solanki (2013) estimators under Stratified Random Sampling.

Estimators	MSE	Estimators	MSE
T_{ss1}	2669.749	T_{ss10}	2669.758
T_{ss2}	2669.751	T_{ss11}	2669.752
T_{ss3}	2669.75	T_{ss12}	2669.787
T_{ss4}	2669.835	T_{ss13}	2670.285
T_{ss5}	2670.225	T_{ss14}	2669.759
T_{ss6}	2670.895	T_{ss15}	2669.76
T_{ss7}	2670.028	T_{ss16}	2670.27
T_{ss8}	2669.758	T_{ss17}	2670.784
T_{ss9}	2669.769		

Table 6.8: MSE values of Singh and Solanki (2014) estimators under Stratified Random Sampling.

Estimators	MSE	Estimators	MSE
T_{sk1}	2670.359	T_{sk7}	2670.369
T_{sk2}	2670.36	T_{sk8}	2670.359
T_{sk3}	2670.359	T_{sk9}	2670.405
T_{sk4}	2670.362	T_{sk10}	2670.359
T_{sk5}	2670.613	T_{sk11}	2670.429
T_{sk6}	2670.961	T_{sk12}	2670.359

Table 6.9: MSE values of proposed estimators under stratified random sampling.

Estimators	MSE	Estimators	MSE
U_{p_1}	2666.951	U_{p_6}	2669.2001
U_{p_2}	2668.081	U_{p_7}	2667.267
U_{p_3}	2662.461	U_{p_8}	2668.348
U_{p_4}	2669.4532	U_{p_9}	2662.918
U_{p_5}	2669.3021		

6.4 Summary of the chapter

Ratio estimators are biased estimators, which are used in under stratified random sampling. Many researchers ducted case studies to improve ratio estimators using various transformations. Many modifications have been made to impose the population mean in stratified random sampling.

Our present study is also focused on modifying the combination of combined ratio and product estimators for the estimation of population mean using stratified random sampling. For this purpose, different new estimators were proposed and compared with the existing estimators in Stratified random sampling.

Our numerical results of the suggested and existing estimators based on the MSE. All results indicate that the MSE of the suggested modified estimators are lower than the MSE of existing estimators. We conclude, therefore, that the suggested modified estimators are the better and more efficient estimators as compared to the existing ones.

Chapter 7

Conclusions with future suggestions

7.1 Conclusion

In survey sampling, the location parameters estimation is one of the major concern by statisticians. For this or even more precise results, the auxiliary variables play a vital role. When outliers occur in the data set, traditional ratio and regression type estimators are not providing suitable results. In such situations robust techniques are suitable for efficient estimation of population mean parameters.

In this study, we have used up to two auxiliary variables for the estimation of population parameter *i.e.* mean in several sampling designs, e.g. simple random sampling, stratified random sampling and systematic random sampling, etc. To measure the performance, we have utilized both type of data sets, *i.e.* real and simulated. In this regard, we have suggested some robust regression based estimators for the estimation of mean of sensitive variable Z when mean of the auxiliary variable is known in SRS. Different robust regression methods are used such as LAD, H-M, H-MM, LTS, LMS, Hample-M, and Tukey-M. For all the proposed estimators, the MSE have been estimated up to first degree of approximation. All the proposed estimators are compared with their existing estimators. Theoretically as well as numerically, the results reveal that the proposed estimators are more efficient than their competitors.

In similar situations, but with different sampling design, such as, systematic random sampling, some ratio type robust estimators have been suggested in this study as well. We have also provided a new class of estimator based on robust regression with utilizing C_x, $\beta_2(x)$ and ρ as known auxiliary information. These estimations are also compared with their competitor taken from literature and found efficient results.

Following, the most recent studies, such as, Zaman (2019) and Zaman and Bulut (2019a), we proposed generalized robust-regression-type estimators to estimate the population mean in simple random sample design. As usual, the mean squared errors are also derived up to first degree of approximation. This study is also extended for different study scheme, *i.e.* two stage random sampling. Based on the numerical findings, our estimators are for away in the sense of imposed performance than their competitors.

We have extended our work and suggested a family of robust type regression estimator by utilizing two auxiliary variables in SRS. The general expressions of mean square errors are also provided. We compared our findings with the finding of Zaman (2019), Zaman and Bulut (2019a) and the results show that the proposed estimators are better. We also introduced some new estimators using transformed auxiliary variable to estimate the population mean. We obtained the numerical results using real life data sets.

7.2 Recommendation

Outliers badly affect the statistical analysis. So, robust regression based means estimators are very fruit in presence of mean estimation. Also these regression estimators require less auxiliary information as compare to reviewed estimators. Hence it is recommended to use the proposed estimators.

7.3 Future direction

Someone can extend this study in the deviation of multivariate case and in different random sampling designs, such as cluster and multiphase random sampling.

References

Abid, M., Abba N., Nazir, H. Z., & Lin, Z. (2016). Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters. *Revista Clombiana de Estadistica*, 39(1), 63-79.

Abid, M., Abbas, N., & Riaz, M. (2016a). Improved modified ratio estimators of population mean based on deciles. *Chiang Mai Journal of Science*, 43(1), 1311-1323.

Abid, M., Nazir, H. Z., Riaz, M., Lin, Z., & Tahir, H. M. (2018). Improved ratio estimators using some robust measures. *Hacettepe Journal of Mathematics and Statistics*, 47(5), 1375-1393.

Abid, M., Sherwani, R. A. K., Abbas, N., & Raza, M. A. (2016). Some improved modified ratio estimators based on decile mean of an auxiliary variable. *Pakistan Journal of Statistics and Operation Research. 12*(4), 787-797.

Abu-Dayyeh, W. A., Ahmed, M. S., Ahmed, R. A., & Muttlak, H. A. (2003). Some estimators of a finite population mean using auxiliary information. *Applied Mathematics and Computation*, 139(2), 287-298.

Ali, N., Ahmad, I., Hanif, M., & Shahzad, U. (2021). Robust-regression-type estimators for improving mean estimation of sensitive variables by using auxiliary information. *Communications in Statistics - Theory and Methods*, 50(4), 979-992.

Al-Noor N. H., & Mohammad, A. A. (2013). Model of robust regression with parametric and nonparametric methods. *Mathematical Theory and Modeling*, 3, 27-39.

Al-Omari, A. I., Jabber, K., & Ibrahim, A. (2008). Modified ratio-type estimators of the mean using extreme ranked set sampling. *Journal of Mathematics and Statistics*, 4(3), 150-155.

Al-Omari, A. I., Ibrahim, K., & Jemain, A. A. (2009). New ratio estimators of the mean using simple random sampling and ranked set sampling methods. *Revista Investigacion Operational*, 30(2), 97-108.

Al-Omari, A. I., & Jabber, K. (2010). Improvement in estimating the population mean in double extreme ranked set sampling. *International Mathematical Forum*, 5(26), 1265-1275.

Al-Omari, A. I. (2012). Ratio estimation of population mean using auxiliary information in simple random sampling and median ranked set sampling. *Statistics and Probability Letters*, 82(11), 1883-1990.

Al-Omari, A. I., & Bouza, C. N. (2015). Ratio estimators of the population mean with missing values using ranked set sampling. *Environmetrics*, 26(2), 67-76.

Al-Omari, A. I., & Al-Nasser, A. D. (2018). Ratio estimation using multistage median ranked set sampling approach. *Journal of Statistical Theory and Practice*, 12(3), 512-529.

Alma, O. G. (2011). Comparison of robust regression methods in linear regression. Int. J. Contemp. *Math-Science*, 6(9), 409-421.

Bahl, S., & Tuteja, R. K. (1991). Ratio and product type exponential estimators. *Journal of information and optimization sciences*, 12(1), 159-164.

Bar-Lev, S. K., Bobovitch, E., & Boukai, B. (2004). A note on randomized response models for quantitative data. *Metrika*, 60, 255–260.

Bassett, Jr. G. W., & Saleh, A. K. (1994). L1Estimation of the median of a survey population. *Journal of Nonparametric Statistics*, 3(3-4), 277-283.

Birkes, D., & Dodge, Y. (1993). Alternative Methods of Regression, John Wiley & Sons Publication, New York.

Boscovich, R. 1. (1757). De literraria expeditions per pontificiam synopsis amplioris operis. ac habentur pluraeius ex exempla riaetiam sensorum impressa. Bononiensi Scientiarum et Artium Inst. Algae Acad. Commen. 4, 353-396.

Bouza, C. N., Al-Omari, A. I., Santiago, A., & Sautto, J. M. (2017). Ratio type estimation using the knowledge of the auxiliary variable for ranking and estimating. *International Journal of Statistics and Probability*, 6(2), 21-30.

Bulut, H., & Zaman, T. (2019). An improved class of robust ratio estimators by using the minimum covariance determinant estimation. *Communications in Statistics-Simulation and Computation.* DOI:10.1080/03610918.2019.1697818.

Chand, L. (1975). Some ratio-type estimators based on two or more auxiliary variables, Ph.D., Iowa State University.

Cochran, W. G. (1977). Sampling techniques. New York, NY: John Wiley and Sons.

Edgeworth, F. Y. (1887). On observations relating to several quantities. *Hermthena,* 6; 279-285.

Eichhorn, B. H., & Hayre, L. S. (1983). Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning Inference*, 7, 307-316.

Fan, W., & Yan, Z. (2010). Factors affecting response rate of the Web survey: A systematic review. *Computers in Human Behavior*, 26(2), 132-139.

Fox, J., & Weisberge, S. (2002). Robust regression. An R and S-Plus companion to applied regression, 91

Gervini, D., & Yohai, V. J. (2002). A class of Robust and fully efficient regression estimators. *The Annals of Statistics*, 30(2), 583-616.

Greenberg, B. G., Kubler, R. R. Jr., Abernathy, J. R., & Horvitz, D. G. (1971). Applications of RR technique in obtaining quantitative data. *Journal of the American Statistical Association,* 66, 243-250.

Gupta, S., Gupta, B., & Sing, S. (2002). Estimation of sensitivity level of personal interview Survey questions. *Journal of Statistical Planning and Inference*, 100, 239-247.

Gupta, S., & Shabbir, J. (2006). An alternative to Warner's randomized response model. *Journal of modern applied statistical methods*, 5(2), 328-331.

Gupta, S., Shabbir, J., & Sehra, S. (2010). Mean and sensitivity estimation in optional randomized response model. *Journal of statistical planning and inference,* 140(10), 2870-2874.

Hampel, F. R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 42 (6), 1887-1896.

Hampel, F.R., Ronchetti, E. M., Rousseeuw, P.J., & Stahel, W. A. (1986). Robust Statistics: The approach based on influence functions, John Wiley & Sons Publication, New York.

Hanif, M., & Shahzad, U. (2019). Estimation of population variance using kernel matrix, *Journal of Statistics and Management Systems*, 22 (3), 563-586.

Huber, P. J. (1964). Robust estimation of a location parameter, *The Annals of Mathematical Statistics*, 35, 73–101.

Huber, P. J. (1973). Robust regression: asymptotic, conjectures and Monte Carlo. *The Annals of Statistics*, 1(5), 799-821.

Huber, P. J. (1981) Robust Statistics, John Wiley and Sons Publication, New York.

Huber, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in statistics* (pp. 492-518). *Springer, New York, NY.*

Huang, J., Yan, L., Cheng, Z., Wu, H., Du, L., Wang, J.,..., & Zang, Y. (2010). A randomized trial comparing radiofrequency ablation and surgical resection for HCC conforming to the Milan criteria. *Annals of surgery*, 252(6), 903-912.

Irfan, M., Javed, M., Abid, M., & Lin, Z. (2018). Improved ratio type estimators of population mean based on median of a study variable and an auxiliary variable. *Hacettepe Journal of Mathematics and Statistics*, 47(3), 659-673.

Irfan, M., Javed, M., & Lin, Z. (2019). Improved Estimation of Population Mean Through Known Conventional and Non-Conventional Measures of Auxiliary

Variable. *Iranian Journal of Science and Technology, Transactions A: Science*, 43(4), 1851-1862.

Jemain, A. A., Al-Omari, A. I., & Ibrahim, K. (2008). Modified ratio estimator for the population mean using double median ranked set sampling. *Pakistan Journal of Statistics*, 24(3): 217-226.

Kadilar, C., Candan, M., & Cingi, H. (2007). Ratio estimators using robust regression. *Hacettepe Journal of Mathematics and Statistics*, *36*(2), 181-188.

King, G., & Wand, J. (2007). Comparing incomparable survey responses. Evaluating and selecting anchoring vignottes. *Political Ananlysis*, 15(1), 46-66.

Koyuncu, N., & Kadilar, C. (2009). Ratio and product estimators in stratified random sampling. *Journal of statistical planning and inference*, *139*(8), 2552-2558.

Koyuncu, N. (2012). Efficient estimators of population mean using auxiliary attributes. *Applied Mathematics and Computation*, *218*(22), 10900-10905.

Koyuncu, N., Gupta, S., & Sousa, R. (2014). Exponential type estimators of the mean of a sensitive variable in the presence of non- sensitive auxiliary information. *Communications in statistics-simulation and computation*, 43(7), 1538-1594.

Kumar, M., Singh, R., Singh, A. K., & Smarandache, F. (2011). Some ratio type estimators under measurement errors. *World applied sciences Journal*, 14(2), 272-276.

Kumar, G., Bharti, P., Yadav, S. K., & Kumar, S. (2019). Improvement for estimating population mean in simple random sampling scheme. *Journal of Reliability and Statistical Studies*, 41-49.

Leroy, A. M., & Rousseeuw, P. J. (1987). Robust regression and outlier detection. *Wiley Series in Probability and Mathematical Statistics, New York.*

Li, G., Peng, H., & Zhu, L. (2011). Nonconcave penalized M-estimation with a diverging number of parameters. *Statistica Sinica*, 391-419.

Lohr, S. (1999). Sampling: Design and Analysis. Duxbury Press.

Lu, J., & Yan, Z. (2014). A class of ratio estimators of a finite population mean using two auxiliary variables. *PLOS ONE*, 2(9), 1-6.

Marona, R., & Zamar, R. (2002). Robust estimates of location and dispersion for high dimensional data set. *Technometrics*, 44, 307-317.

Marona, R. A., martin, D. R., & Yohai, V. (2006). Robust Statistics: Theory and Methods. New York: John Wiley & Sons.

Millar, M. M., & Dillman, D. A. (2011). Improving response to web and mixed-mode surveys. *Public opinion Q*, 75(2), 249-269.

Mohebbi, M., Nourijelyani, K., & Zeraati, H. (2007). A simulation study on robust alternatives of least squares regression. *Journal of Applied Science*, 7(22), 3469-3476.

Murthy, M. N. (1967). Sampling Theory and Methods (Statistical Publishing Society, Calcutta, India.

Nadia, H., & Mohammad, A. A. (2013). Model of robust regression with parametric and nonparametric methods. *Mathematical Theory and Modeling*, 3, 27–39.

Naz, F., Abid, M., Nawaz, T., & Pang, T. (2019). Enhancing the efficiency of the ratio-type estimators of population variance with a blend of information on robust location measures. *Scientia Iranica.* 10.24200/sci.2019.5633.1385.

Neyman, J. (1938). Contribution to the theory of sampling human populations. *Journal of the American Statistical Association,* 33(201), 101-116

Olkin, I. (1958). Multivariate ratio estimation for finite populations, *Biometrika,* 45, 154 -165.

Oral, E., & Oral, E. (2011). A robust alternative to the ratio estimator under non-normality. *Statistics & probability letters, 81*(8), 930-936.

Pollock, K. H., & Beck, Y. (1976). A comparison of three randomized response models for quantitative data. *Journal of the American Statistical Association,* 71, 884–886.

Pradhan, B. K. (2005). A chain regression estimator in two phase sampling using multi-auxiliary information. *Bulletin of the Malaysian mathematical sciences society,* 28 (1), 81-86.

Raj, D. (1965). On a method of using multi-auxiliary information in sample surveys, *Journal of the American Statistical Association,* 60(309), 270-277.

Rao, P. S. R. S., & Mudholkar, G. S. (1967). Generalized multivariate estimator for the mean of finite populations. *Journal of the American Statistical Association,* 62(319), 1009-1012.

Renaud, O., & Victoria-Faser, M. P. (2010). A Robust Coefficient of determination for regression. *Journal of Statistical Planning and Inference,* 140(7), 1852-1862.

Reynolds, W. M. (1982). Development of reliable and valid short from of the Marlowe crown sdb scale. *Journal of clinical psychology*, 38, 119-125.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical association*, 79(388), 871-880.

Rousseeuw, P.J. & Yohai, V. (1984). Robust ion by means of S-estimators, in Robust and non-linear time series analysis, (pp. 256–272), *Springer Verilog, New York, N. Y.*

Rousseeuw, P. J., & Leroy, A. M. (2005). Robust Regression and Outlier Detection. *John Wiley and Sons Publication, New York.*

Sarndal, C. E., Swensson, B., & Wretman, J. (1992). Model Assisted Survey Sampling, *Springer.*

Searls, D. T. (1964). The utilization of a known coefficient of variation in the estimation procedure, *Journal of the American Statistical Association*, 59(308), 1225-1226.

Sen, P. K. (1968). Estimate of the regression coefficient based on Kendall's Tau; *Journal of the American Statistical Association*, 63, 1379- 1389.

Shahzad, U. (2016). On the Estimation Of Population Mean Under Systematic Sampling Using Auxiliary Attributes. *Oriental Journal Physical Sciences*, 1 & 2, 17–22.

Shahzad, U., Hanif, M., Koyuncu, N., & Garcia Luengo, A.V. (2017). A new family of estimators for mean estimation alongside the sensitivity issue. *Journal of Reliability and Statistical Studies*, 10, 43–63.

Shahzad, U., Hanif, M., & Koyuncu, N. (2018). A new estimator for mean under stratified random sampling. *Mathematical Sciences, 12*(3), 163-169.

Shahzad, U., Perri P.F., & Hanif, M. (2018). A new class of ratio-type estimators for improving mean estimation of non-sensitive and sensitive variables by using supplementary information. *Communications in Statistics- Simulation and Computation.* DOI: 10.1080/03610918.2018.1458133.

Shahzad, U., & Hanif, M. (2019). Some imputation based new estimators of population mean under non-response, *Journal of Statistics and Management System.,* 22(8), 1381-1399.

Shahzad, U., Perri, P. F., & Hanif, M. (2019). A new class of ratio-type estimators for improving mean estimation of non-sensitive and sensitive variables by using supplementary information. *Communications in Statistics-Simulation and Computation, 48*(9), 2566-2585.

Shahzad, U., Al-Noor, N. H., Afshan, N, Alilah, D. A., Hanif, M., & Anas, M. M. (2021). Minimum Covariance determinant-based quantile robust regression type estimators for mean parameter. *Mathematical Problem in Engineering.*

Siegel, A. F. (1982). Robust regression using repeated medians. *Biometrika,* 69, 242-244

Singh, D., & Chaudhary, F. S. (1986). Theory and analysis of sample survey designs. New Age International Publisher, New Delhi, India.

Singh, H. P., & Solanki, R. S. (2013). Efficient ratio and product estimators in stratified random sampling. *Communications in Statistics-Theory and Methods, 42*(6), 1008-1023.

Singh, S. (2003). Advanced Sampling Theory with Applications. How Michael 'Selected' Amy. Kluwer Academic Publishers, Dordrecht.

Solanki, R. S., & Singh, H. P. (2014). An Efficient Class of Estimators for the Population Mean Using Auxiliary Information in Stratified Random Sampling. *Communications in Statistics-Theory and Methods*, *43*(16), 3380-3401.

Subzar, M., Bouza, C. N., & Al-Omari, A. I. (2019). Utilization Of Different Robust Regression Techniques For Estimation Of Finite Population Mean In SRSWOR in Case Of Presence Of Outliers Through Ratio Method Of Estimation. *Revista Investigation Operational*, *40*(5), 600-609.

Subzar, M., Ibrahim Al- Omari, A., & Alanzi, A. R. A. (2020). The robust regression methods for estimating of finite population mean based on SRSWOR in case of outliers. *Computers, Materials & Continua*, 65(1), 1380-2020.

Sukhatme, B. V. (1962). Some ratio-type estimators in two-phase sampling. *Journal of the Ameri-can Statistical Association*, 57, 299, 628-632.

Sukhatme, P. V., & Sukhatme, B. V. (1970). Sampling theory of surveys with applications, Iowa State University press, Ames, USA. And polynomial regression analysis. Indagations Mathematica 12, 85-91.

Theil, H. (1950). A rank invariant method of linear

Tukey, J. W. (1977). Exploratory data analysis. MA: Addison-Wesley.

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

Warner, S. L. (1971). The linear randomized response model. *Journal of the American Statistical Association,* 66, 884–888.

Wrighr, P. G. (1928). The tariff on Animal and vegetable Oils. New York: Macmillan.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15, 642-656.

Zaman, T., & Bulut, H. (2018). Modified ratio estimators using robust regression methods. *Communications in Statistics - Theory and Methods*, DOI: 10.1080/03610926.2018.1441419.

Zaman, T. (2019). Improvement of modified ratio estimators using robust regression methods. *Applied Mathematics and Computation*, 348, 627-631.

Zaman, T., & Bulut, H. (2019a). Modified ratio estimators using robust regression methods. *Communications in Statistics-Theory and Methods*, 48(8), 2039-2048.

Zaman, T., & Bulut, H. (2019b). Modified regression estimators using robust regression methods and covariance matrices in stratified random sampling. *Communications in Statistics-Theory and Methods*, DOI: 10.1080/03610926.2019.1588324.

Zaman, T., & Kadilar, C. (2019). New class of exponential estimators for finite population mean in two-phase sampling. *Communications in Statistics - Theory and Methods,* DOI: 10.1080/03610926.2019.1643480.

Zaman, T., Dunder, E., Audu, A., Alilah, D. a, Shahzad, U., & Hanif, M. (2021). Robust regression- ratio type estimators of the mean utilizing two auxiliary variables. A simulation study. *Mathematical problems in Engineering.*

Appendix A

The explanation of the expressions used in above proof of proposed estimators.

1st Estimator

Substituting values of $\overline{y_{st}}$ and $\overline{x_{st}}$ in equation (6.18).

$$=\left[\phi_1\bar{Y}\left(\frac{1}{2}+\frac{\lambda_0}{2}\right)\left((1+\lambda_1)^{-1}+(1+\lambda_1)\right)+\phi_2\left(-\bar{X}\lambda_1\right)\right]\left(1-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^2\lambda_1^2}{8}\right), \quad \text{(A1)}$$

$$U_{P_1}-\bar{Y}=\phi_1\bar{Y}\left(1+\lambda_0-\frac{\vartheta\lambda_1}{2}-\frac{\vartheta\lambda_0\lambda_1}{2}+\alpha_1\lambda_1^2\right)-\phi_2\bar{X}\left(\lambda_1-\frac{\vartheta^2\lambda_1^2}{2}\right)-\bar{Y}. \quad \text{(A2)}$$

Applying expectation

$$bias\left(U_{P_1}\right)=\phi_1\bar{Y}\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_1\eta_1\right)+\phi_2\bar{X}\frac{\vartheta^2\eta_1}{2}-\bar{Y}. \quad \text{(A3)}$$

$$\begin{aligned}\left(U_{P_1}-\bar{Y}\right)^2=&\bar{Y}^2+\phi_1^2\bar{Y}^2\left(1+\lambda_0^2+\frac{\vartheta^2\lambda_1^2}{4}-2\vartheta\lambda_0\lambda_1+2\alpha_1\lambda_1^2\right)+\phi_2^2\bar{X}^2\lambda_1^2\\&+2\phi_1\phi_2\bar{Y}\bar{X}\left(\vartheta\lambda_1^2-\lambda_0\lambda_1\right)-2\phi_1\bar{Y}^2\left(1-\frac{\vartheta\lambda_0\lambda_1}{2}+\alpha_1\lambda_1^2\right)\\&-2\phi_2\frac{\bar{Y}\bar{X}\vartheta\lambda_1^2}{2}\end{aligned} \quad \text{(A4)}$$

Applying expectation

$$MSE\left(U_{P_1}\right)=\bar{Y}^2+\phi_1^2U_{A1(t)}+\phi_2^2U_{B1(t)}+2\phi_1\phi_2U_{C1(t)}-2\phi_1U_{D1(t)}-2\phi_2U_{E1(t)}. \quad \text{(A5)}$$

Where,

$$U_{A1(t)}=\bar{Y}^2\left(1+\eta_0+\frac{\vartheta^2\eta_1}{4}-2\vartheta\eta_{01}+2\alpha_1\eta_1\right),$$

$$U_{B1(t)}=\bar{X}^2\eta_1,$$

$$U_{C1(i)}=\overline{YX}\left(\vartheta\eta_1-\eta_{01}\right),$$

$$U_{D1(i)}=\overline{Y}^2\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_1\eta_1\right),$$

$$U_{E1(i)}=\frac{\overline{YX}\vartheta\eta_1}{2}.$$

Differentiating MSE partially, with respect to $\emptyset_1$ and $\emptyset_2$ and equating to zero, we get the following optimum values of $\emptyset_1$ and $\emptyset_2$

$$\phi_{1(opt)}=\frac{U_{B1(i)}U_{D1(i)}-U_{C1(i)}U_{E1(i)}}{U_{A1(i)}U_{B1(i)}-U_{C1(i)}^{\ 2}},\ \phi_{2(opt)}=\frac{U_{A1(i)}U_{E1(i)}-U_{C1(i)}U_{D1(i)}}{U_{A1(i)}U_{B1(i)}-U_{C1(i)}^{\ 2}}\ . \qquad \textit{(say)}$$

The minimum MSE of U_{p_1}

$$MSE_{\min}\left(U_{p_1}\right)=\overline{Y}^2-\frac{U_{A1(i)}U_{E1(i)}^{\ 2}+U_{B1(i)}U_{D1(i)}^{\ 2}-2U_{C1(i)}U_{D1(i)}U_{E1(i)}}{U_{A1(i)}U_{B1(i)}-U_{C1(i)}^{\ 2}}\ . \tag{A6}$$

2nd Estimator

Substituting values of $\overline{y_{st}}$ and $\overline{x_{st}}$ in equation (6.24)

$$=\left[\phi_1\overline{Y}\left(\frac{1}{2}+\frac{\lambda_0}{2}\right)\left\{\exp(-\lambda_1)(2+\lambda_1)^{-1}+\exp(\lambda_1)(2+\lambda_1)^{-1}\right\}+\phi_2\left(-\overline{X}\lambda_1\right)\right]$$
$$\left(1-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^2\lambda_1^{\ 2}}{8}\right) \tag{A7}$$

$$U_{P_2}-\overline{Y}=\phi_1\overline{Y}\left(1+\lambda_0-\frac{\vartheta\lambda_1}{2}-\frac{\vartheta\lambda_0\lambda_1}{2}+\alpha_2\lambda_1^{\ 2}\right)+\phi_2\overline{X}\left(\lambda_1-\frac{\vartheta^2\lambda_1^{\ 2}}{2}\right)-\overline{Y}\cdot \tag{A8}$$

Applying expectation

$$bias\left(U_{P_2}\right)=\phi_1\overline{Y}\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_2\eta_1\right)-\phi_2\overline{X}\frac{\vartheta^2\eta_1}{2}-\overline{Y}\cdot \tag{A9}$$

$$\left(U_{P_2}-\bar{Y}\right)^2=\bar{Y}^2+\phi_1^2\bar{Y}^2\left(1+\lambda_0^2+\frac{\vartheta^2\lambda_1^2}{4}-2\vartheta\lambda_0\lambda_1+2\alpha_2\lambda_1^2\right)+\phi_2^2\bar{X}^2\lambda_1^2$$
$$+2\phi_1\phi_2\bar{Y}\bar{X}\left(\vartheta\lambda_1^2-\lambda_0\lambda_1\right)-2\phi_1\bar{Y}^2\left(1-\frac{\vartheta\lambda_0\lambda_1}{2}+\alpha_2\lambda_1^2\right) \tag{A10}$$
$$-2\phi_2\frac{\bar{Y}\bar{X}\vartheta\lambda_1^2}{2}$$

Applying expectation

$$MSE\left(U_{P_2}\right)=\bar{Y}^2+\phi_1^2U_{A2(i)}+\phi_2^2U_{B2(i)}+2\phi_1\phi_2U_{C2(i)}-2\phi_1U_{D2(i)}-2\phi_2U_{E2(i)}. \tag{A11}$$

where,

$$U_{A2(i)}=\bar{Y}^2\left(1+\eta_0+\frac{\vartheta^2\eta_1}{4}-2\vartheta\eta_{01}+2\alpha_1\eta_1\right),$$

$$U_{B2(i)}=\bar{X}^2\eta_1,$$

$$U_{C2(i)}=\bar{Y}\bar{X}\left(\vartheta\eta_1-\eta_{01}\right),$$

$$U_{D2(i)}=\bar{Y}^2\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_1\eta_1\right),$$

$$U_{E2(i)}=\frac{\bar{Y}\bar{X}\vartheta\eta_1}{2}.$$

Differentiating MSE partially, with respect to $\emptyset_1$ and $\emptyset_2$ and equating to zero, we get the following optimum values of $\emptyset_1$ and $\emptyset_2$

$$\phi_{1(opt)}=\frac{U_{B2(i)}U_{D2(i)}-U_{C2(i)}U_{E2(i)}}{U_{A2(i)}U_{B2(i)}-U_{C2(i)}^2},\ \phi_{2(opt)}=\frac{U_{A2(i)}U_{E2(i)}-U_{C2(i)}U_{D2(i)}}{U_{A2(i)}U_{B2(i)}-U_{C2(i)}^2}. \qquad \textit{(say)}$$

The minimum MSE of U_{p_2}

$$MSE_{\min}\left(U_{P_2}\right)=\bar{Y}^2-\frac{U_{A2(i)}U_{E2(i)}^2+U_{B2(i)}U_{D2(i)}^2-2U_{C2(i)}U_{D2(i)}U_{E2(i)}}{U_{A2(i)}U_{B2(i)}-U_{C2(i)}^2}. \tag{A12}$$

3rd Estimator

Substituting values of $\overline{y_{st}}$ and $\overline{x_{st}}$ in equation (6.29)

$$=\phi_1\overline{Y}\left(1+\lambda_0+\frac{7\lambda_1^2}{8}\right)\left(1-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^2\lambda_1^2}{8}\right)-\phi_2\overline{X}\lambda_1\left(1-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^2\lambda_1^2}{8}\right). \tag{A13}$$

$$U_{P_3}-\overline{Y}=\phi_1\overline{Y}\left(1+\lambda_0-\frac{\vartheta\lambda_1}{2}-\frac{\vartheta\lambda_0\lambda_1}{2}+\alpha_3\lambda_1^2\right)-\phi_2\overline{X}\left(\lambda_1-\frac{\vartheta^2\lambda_1^2}{2}\right)-\overline{Y}. \tag{A14}$$

Applying expectation

$$bias\left(U_{P_3}\right)=\phi_1\overline{Y}\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_3\eta_1\right)+\phi_2\overline{X}\frac{\vartheta^2\eta_1}{2}-\overline{Y}. \tag{A15}$$

$$\begin{aligned}\left(U_{P_3}-\overline{Y}\right)^2=\overline{Y}^2+\phi_1^2\overline{Y}^2&\left(1+\lambda_0^2+\frac{\vartheta^2\lambda_1^2}{4}-2\vartheta\lambda_0\lambda_1+2\alpha_3\lambda_1^2\right)+\phi_2^2\overline{X}^2\lambda_1^2\\&+2\phi_1\phi_2\overline{Y}\overline{X}\left(\vartheta\lambda_1^2-\lambda_0\lambda_1\right)-2\phi_1\overline{Y}^2\left(1-\frac{\vartheta\lambda_0\lambda_1}{2}+\alpha_3\lambda_1^2\right)\\&-2\phi_2\frac{\overline{Y}\overline{X}\vartheta\lambda_1^2}{2}\end{aligned} \tag{A16}$$

Applying expectation,

$$MSE\left(U_{P_3}\right)=\overline{Y}^2+\phi_1^2U_{A3(t)}+\phi_2^2U_{B3(t)}+2\phi_1\phi_2U_{C3(t)}-2\phi_1U_{D3(t)}-2\phi_2U_{E3(t)}. \tag{A17}$$

where,

$$U_{A3(t)}=\overline{Y}^2\left(1+\eta_0+\frac{\vartheta^2\eta_1}{4}-2\vartheta\eta_{01}+2\alpha_1\eta_1\right),$$

$$U_{B3(t)}=\overline{X}\eta_1,$$

$$U_{C3(t)}=\overline{Y}\overline{X}\left(\vartheta\eta_1-\eta_{01}\right),$$

$$U_{D3(t)}=\overline{Y}^2\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_1\eta_1\right),$$

$$U_{E3(i)}=\frac{\overline{Y}\overline{X}\vartheta\eta_1}{2}.$$

Differentiating MSE partially, with respect to $\emptyset_1$ and $\emptyset_2$ and equating to zero, we get the following optimum values of $\emptyset_1$ and $\emptyset_2$

$$\phi_{1(opt)}=\frac{U_{B3(i)}U_{D3(i)}-U_{C3(i)}U_{E3(i)}}{U_{A3(i)}U_{B3(i)}-U_{C3(i)}^{2}},\ \phi_{2(opt)}=\frac{U_{A3(i)}U_{E3(i)}-U_{C3(i)}U_{D3(i)}}{U_{A3(i)}U_{B3(i)}-U_{C3(i)}^{2}}. \qquad \textit{(say)}$$

The minimum MSE of U_{p_3}

$$MSE_{\min}\left(U_{P_3}\right)=\overline{Y}^2-\frac{U_{A3(i)}U_{E3(i)}^{2}+U_{B3(i)}U_{D3(i)}^{2}-2U_{C3(i)}U_{D3(i)}U_{E3(i)}}{U_{A3(i)}U_{B3(i)}-U_{C3(i)}^{2}}. \tag{A18}$$

4th Estimator

Substituting values of $\overline{y_{st}}$ and $\overline{x_{st}}$ in equation (6.33)

$$=\left[\phi_1\overline{Y}\left(\frac{1}{2}+\frac{\lambda_0}{2}\right)\left((1+\lambda_1)^{-1}+(1+\lambda_1)\right)+(1-2\phi_2)\left(-\overline{X}\lambda_1\right)\right]\left(1-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^2\lambda_1^2}{8}\right). \tag{A19}$$

$$U_{P_4}-\overline{Y}=\phi_1\overline{Y}\left(1+\lambda_0-\frac{\vartheta\lambda_1}{2}-\frac{\vartheta\lambda_0\lambda_1}{2}+\alpha_4\lambda_1^2\right)-(1-2\phi_2)\overline{X}\left(\lambda_1-\frac{\vartheta^2\lambda_1^2}{2}\right)-\overline{Y}. \tag{A20}$$

Applying expectation

$$bias\left(U_{P_4}\right)=\phi_1\overline{Y}\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_4\eta_1\right)+(1-2\phi_2)\overline{X}\frac{\vartheta^2\eta_1}{2}-\overline{Y}. \tag{A21}$$

$$\begin{aligned}\left(U_{P_4}-\overline{Y}\right)^2=&\left(\overline{Y}^2+\overline{X}^2\lambda_1^2-\overline{X}\overline{Y}\vartheta\lambda_1^2\right)+\phi_1^2\overline{Y}^2\left(1+\lambda_0^2+\frac{\vartheta^2\lambda_1^2}{4}-2\vartheta\lambda_0\lambda_1+2\alpha_4\lambda_1^2\right)\\&+\phi_2^2 4\overline{X}^2\lambda_1^2+2\phi_1\phi_2\overline{Y}\overline{X}\left(2\lambda_0\lambda_1-2\vartheta\lambda_1^2\right)-2\phi_1\overline{Y}\left(\overline{Y}-\frac{\overline{Y}\vartheta\lambda_0\lambda_1}{2}+\overline{Y}\alpha_4\lambda_1^2+\overline{X}\lambda_0\lambda_1-\overline{X}\vartheta\lambda_1^2\right)\\&-2\phi_2\overline{Y}\overline{X}\vartheta\lambda_1^2\end{aligned} \tag{A22}$$

Applying expectation

$$MSE\left(U_{P_4}\right)=\left(\overline{Y}^2+\overline{X}^2\eta_1-\overline{X}\overline{Y}\vartheta\eta_1\right)+\phi_1^2U_{A4(i)}+\phi_2^2U_{B4(i)}+2\phi_1\phi_2U_{C4(i)}\,.$$
$$-2\phi_1U_{D4(i)}-2\phi_2U_{E4(i)} \tag{A23}$$

where,

$$U_{A4(i)}=\overline{Y}^2\left(1+\eta_0+\frac{\vartheta^2\eta_1}{4}-2\vartheta\eta_{01}+2\alpha_4\eta_1\right),$$

$$U_{B4(i)}=4\overline{X}^2\eta_1,$$

$$U_{C4(i)}=\overline{Y}\overline{X}\left(2\eta_{01}-2\vartheta\eta_1\right),$$

$$U_{D4(i)}=\overline{Y}\left(\overline{Y}-\frac{\overline{Y}\vartheta\eta_{01}}{2}+\overline{Y}\alpha_4\eta_1+\overline{X}\eta_{01}-\overline{X}\vartheta\eta_1\right),$$

$$U_{E4(i)}=\overline{Y}\overline{X}\vartheta\eta_1.$$

Differentiating MSE partially, with respect to $\emptyset_1$ and $\emptyset_2$ and equating to zero, we get the following optimum values of $\emptyset_1$ and $\emptyset_2$

$$\phi_{1(opt)}=\frac{U_{B4(i)}U_{D4(i)}-U_{C4(i)}U_{E4(i)}}{U_{A4(i)}U_{B4(i)}-U_{C4(i)}^2},\ \phi_{2(opt)}=\frac{U_{A4(i)}U_{E4(i)}-U_{C4(i)}U_{D4(i)}}{U_{A4(i)}U_{B4(i)}-U_{C4(i)}^2}. \qquad \textit{(say)}$$

the minimum MSE of U_{p_4}

$$MSE_{min}\left(U_{P_4}\right)=\left(\overline{Y}^2+\overline{X}^2\eta_1-\overline{X}\overline{Y}\vartheta\eta_1\right)-\frac{U_{A4(i)}U_{E4(i)}^2+U_{B4(i)}U_{D4(i)}^2-2U_{C4(i)}U_{D4(i)}U_{E4(i)}}{U_{A4(i)}U_{B4(i)}-U_{C4(i)}^2}. \tag{A24}$$

5th Estimator

Substituting values of $\overline{y_{st}}$ and $\overline{x_{st}}$ in equation (6.37)

$$=\phi_1\overline{Y}\left(1+\lambda_0+\frac{3\lambda_1^2}{8}\right)\left(1-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^2\lambda_1^2}{8}\right)-(1-2\phi_2)\overline{X}\lambda_1\left(1-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^2\lambda_1^2}{8}\right), \tag{A25}$$

$$U_{p_5}-\bar{Y}=\phi_1\bar{Y}\left(1+\lambda_0-\frac{\vartheta\lambda_1}{2}-\frac{\vartheta\lambda_0\lambda_1}{2}+\alpha_2\lambda_1^2\right)-(1-2\phi_2)\bar{X}\left(\lambda_1-\frac{\vartheta^2\lambda_1^2}{2}\right)-\bar{Y}\cdot \tag{A26}$$

Applying expectation

$$bias\left(U_{p_5}\right)=\phi_1\bar{Y}\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_2\eta_1\right)+(1-2\phi_2)\bar{X}\frac{\vartheta^2\eta_1}{2}-\bar{Y}\cdot \tag{A27}$$

$$\begin{aligned}\left(U_{p_5}-\bar{Y}\right)^2&=\left(\bar{Y}^2+\bar{X}^2\lambda_1^2-\bar{X}\bar{Y}\vartheta\lambda_1^2\right)+\phi_1^2\bar{Y}^2\left(1+\lambda_0^2+\frac{\vartheta^2\lambda_1^2}{4}-2\vartheta\lambda_0\lambda_1+2\alpha_5\lambda_1^2\right)\\&+\phi_2^2 4\bar{X}^2\lambda_1^2+2\phi_1\phi_2\bar{Y}\bar{X}\left(2\lambda_0\lambda_1-2\vartheta\lambda_1^2\right)-2\phi_1\bar{Y}\left(\bar{Y}-\frac{\bar{Y}\vartheta\lambda_0\lambda_1}{2}+\bar{Y}\alpha_5\lambda_1^2+\bar{X}\lambda_0\lambda_1-\bar{X}\vartheta\lambda_1^2\right)\\&-2\phi_2\bar{X}\left(\bar{Y}\vartheta\lambda_1^2-2\bar{X}\lambda_1^2\right)\end{aligned} \tag{A27}$$

Applying expectation

$$\begin{aligned}MSE\left(U_{p_5}\right)&=\left(\bar{Y}^2+\bar{X}^2\eta_1-\bar{X}\bar{Y}\vartheta\eta_1\right)+\phi_1^2U_{A5(i)}+\phi_2^2U_{B5(i)}+2\phi_1\phi_2U_{C5(i)}.\\&-2\phi_1U_{D5(i)}-2\phi_2U_{E5(i)}\end{aligned} \tag{A28}$$

where,

$$U_{A5(i)}=\bar{Y}^2\left(1+\eta_0+\frac{\vartheta^2\eta_1}{4}-2\vartheta\eta_{01}+2\alpha_5\eta_1\right),$$

$$U_{B5(i)}=4\bar{X}^2\eta_1,$$

$$U_{C5(i)}=\bar{Y}\bar{X}\left(2\eta_{01}-2\vartheta\eta_1\right)$$

$$U_{D5(i)}=\bar{Y}\left(\bar{Y}-\frac{\bar{Y}\vartheta\eta_{01}}{2}+\bar{Y}\alpha_5\eta_1+\bar{X}\eta_{01}-\bar{X}\vartheta\eta_1\right),$$

$$U_{E5(i)}=\bar{X}\left(\bar{Y}\vartheta\eta_1-2\bar{X}\eta_1\right).$$

Differentiating MSE partially, with respect to ϕ_1 and ϕ_2 and equating to zero, we get the following optimum values of ϕ_1 and ϕ_2

$$\phi_{1(opt)} = \frac{U_{B5(i)}U_{D5(i)} - U_{C5(i)}U_{E5(i)}}{U_{A5(i)}U_{B5(i)} - U_{C5(i)}^{2}}, \quad \phi_{2(opt)} = \frac{U_{A5(i)}U_{E5(i)} - U_{C5(i)}U_{D5(i)}}{U_{A5(i)}U_{B5(i)} - U_{C5(i)}^{2}}. \qquad \textit{(say)}$$

The minimum MSE of U_{p_5}

$$MSE_{min}\left(U_{p_5}\right) = \left(\overline{Y}^2 + \overline{X}^2\eta_1 - \overline{X}\overline{Y}\vartheta\eta_1\right) - \frac{U_{A5(i)}U_{E5(i)}^{2} + U_{B5(i)}U_{D5(i)}^{2} - 2U_{C5(i)}U_{D5(i)}U_{E5(i)}}{U_{A5(i)}U_{B5(i)} - U_{C5(i)}^{2}}. \tag{A29}$$

6th Estimator

Substituting values of $\overline{y_{st}}$ and $\overline{x_{st}}$ in equation (6.40)

$$= \left[\phi_1\overline{Y}\left(\frac{1}{4} + \frac{\lambda_0}{4}\right)(2 + \lambda_1^2)\left\{\exp\left(\frac{-\lambda_1}{2}\right)\left(1 - \frac{\lambda_1}{2} + \frac{\lambda_1^2}{4}\right) + \exp\left(\frac{\lambda_1}{2}\right)\left(1 - \frac{\lambda_1}{2} + \frac{\lambda_1^2}{4}\right)\right\} + (1 - 2\phi_2)(-\overline{X}\lambda_1)\right] \left(1 - \frac{\vartheta\lambda_1}{2} + \frac{3\vartheta^2\lambda_1^2}{8}\right) \tag{A30}$$

$$U_{p_6} - \overline{Y} = \phi_1\overline{Y}\left(1 + \lambda_0 - \frac{\vartheta\lambda_1}{2} - \frac{\vartheta\lambda_0\lambda_1}{2} + \alpha_6\lambda_1^2\right) - (1 - 2\phi_2)\overline{X}\left(\lambda_1 - \frac{\vartheta^2\lambda_1^2}{2}\right) - \overline{Y} \cdot \tag{A31}$$

Taking expectation

$$bias\left(U_{p_6}\right) = \phi_1\overline{Y}\left(1 - \frac{\vartheta\eta_{01}}{2} + \alpha_6\eta_1\right) + (1 - 2\phi_2)\overline{X}\frac{\vartheta^2\eta_1}{2} - \overline{Y}. \tag{A32}$$

$$\begin{aligned} \left(U_{p_6} - \overline{Y}\right)^2 &= \left(\overline{Y}^2 + \overline{X}^2\lambda_1^2 - \overline{X}\overline{Y}\vartheta\lambda_1^2\right) + \phi_1^2\overline{Y}^2\left(1 + \lambda_0^2 + \frac{\vartheta^2\lambda_1^2}{4} - 2\vartheta\lambda_0\lambda_1 + 2\alpha_6\lambda_1^2\right) \\ &+ \phi_2^2 4.\overline{X}^2\lambda_1^2 + 2\phi_1\phi_2\overline{Y}.\overline{X}\left(2\lambda_0\lambda_1 - 2\vartheta\lambda_1^2\right) - 2\phi_1\overline{Y}\left(\overline{Y} - \frac{\overline{Y}\vartheta\lambda_0\lambda_1}{2} + \overline{Y}\alpha_6\lambda_1^2 + \overline{X}\lambda_0\lambda_1 - \overline{X}\vartheta\lambda_1^2\right) \\ &- 2\phi_2\overline{X}\left(\overline{Y}\vartheta\lambda_1^2 - 2\overline{X}\lambda_1^2\right) \end{aligned} \tag{A33}$$

Taking expectation

$$\begin{aligned} MSE\left(U_{p_6}\right) &= \left(\overline{Y}^2 + \overline{X}^2\eta_1 - \overline{X}\overline{Y}\vartheta\eta_1\right) + \phi_1^2U_{A6(i)} + \phi_2^2U_{B6(i)} + 2\phi_1\phi_2U_{C6(i)} \\ &- 2\phi_1U_{D6(i)} - 2\phi_2U_{E6(i)} \end{aligned} \tag{A34}$$

where,

$$U_{A6(i)}=\bar{Y}^2\left(1+\eta_0+\frac{\vartheta^2\eta_1}{4}-2\vartheta\eta_{01}+2\alpha_6\eta_1\right),$$

$$U_{B6(i)}=4\bar{X}^2\eta_1,$$

$$U_{C6(i)}=\bar{Y}\bar{X}\left(2\eta_{01}-2\vartheta\eta_1\right)$$

$$U_{D6(i)}=\bar{Y}\left(\bar{Y}-\frac{\bar{Y}\vartheta\eta_{01}}{2}+\bar{Y}\alpha_6\eta_1+\bar{X}\eta_{01}-\bar{X}\vartheta\eta_1\right),$$

$$U_{E6(i)}=\bar{X}\left(\bar{Y}\vartheta\eta_1-2\bar{X}\eta_1\right).$$

Differentiating MSE partially, with respect to ϕ_1 and ϕ_2 and equating to zero, we get the following optimum values of ϕ_1 and ϕ_2

$$\phi_{1(opt)}=\frac{U_{B6(i)}U_{D6(i)}-U_{C6(i)}U_{E6(i)}}{U_{A6(i)}U_{B6(i)}-U_{C6(i)}^{\ 2}},\ \phi_{2(opt)}=\frac{U_{A6(i)}U_{E6(i)}-U_{C6(i)}U_{D6(i)}}{U_{A6(i)}U_{B6(i)}-U_{C6(i)}^{\ 2}}. \qquad \textit{(say)}$$

The minimum MSE of U_{p_6}

$$MSE_{min}\left(U_{P_6}\right)=\left(\bar{Y}^2+\bar{X}^2\eta_1-\bar{X}\bar{Y}\vartheta\eta_1\right)-\frac{U_{A6(i)}U_{E6(i)}^{\ 2}+U_{B6(i)}U_{D6(i)}^{\ 2}-2U_{C6(i)}U_{D6(i)}U_{E6(i)}}{U_{A6(i)}U_{B6(i)}-U_{C6(i)}^{\ 2}}. \qquad (A35)$$

7th Estimator

Substituting values of $\overline{y_{st}}$ and $\overline{x_{st}}$ in equation (6.47)

$$=\left[\bar{Y}\left(\frac{1}{2}+\frac{\lambda_0}{2}\right)\left((1+\lambda_1)^{-1}+(1+\lambda_1)\right)+\phi_1\bar{Y}(1+\lambda_0)+\phi_2\left(-\bar{X}\lambda_1\right)\right]\left(1-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^2\lambda_1^2}{8}\right). \qquad (A36)$$

$$\begin{aligned}U_{P_7}-\bar{Y}=&\bar{Y}\left(\lambda_0-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^2\lambda_1^2}{8}-\frac{\vartheta\lambda_0\lambda_1}{2}+\frac{\lambda_1^2}{2}\right)+\phi_1\bar{Y}\left(1+\lambda_0-\frac{\vartheta\lambda_1}{2}-\frac{\vartheta\lambda_1\lambda_0}{2}+\frac{\vartheta\lambda_1^2}{2}+\frac{3\vartheta^2\lambda_1^2}{8}\right)\\&-\phi_2\bar{X}\lambda_1\left(1-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^2\lambda_1^2}{8}\right)\end{aligned} \qquad (A37)$$

Applying expectation

$$bias\left(U_{P_7}\right)=\overline{Y}\left(-\frac{\vartheta\eta_1}{2}+\frac{3\vartheta^2\eta_1}{8}-\frac{\vartheta\eta_{01}}{2}+\frac{\eta_1}{2}\right)+\phi_1\overline{Y}\left(1-\frac{\vartheta\eta_{10}}{2}+\frac{\vartheta\eta_1}{2}+\frac{3\vartheta^2\eta_1}{8}\right)+\phi_2\overline{X}\frac{\vartheta\eta_1}{2}, \tag{A38}$$

$$\begin{aligned}\left(U_{P_7}-\overline{Y}\right)^2=&\overline{Y}^2\left(\lambda_0^2+\frac{\vartheta^2\lambda_1^2}{4}-\vartheta\lambda_0\lambda_1\right)+\phi_1^2\overline{Y}^2\left(1+\lambda_0^2+\frac{\vartheta^2\lambda_1^2}{4}-2\vartheta\lambda_0\lambda_1+\frac{3}{4}\vartheta^2\lambda_1^2\right)+\phi_2^2\overline{X}^2\lambda_1^2\\&+2\phi_1\phi_2\overline{Y}\overline{X}\left(\vartheta\lambda_1^2-\lambda_0\lambda_1\right)-2\phi_1\overline{Y}^2\left(\lambda_0^2-\frac{\vartheta\lambda_0\lambda_1}{2}-\frac{\vartheta\lambda_0\lambda_1}{2}+\frac{\vartheta^2\lambda_1^2}{4}-\frac{\vartheta\lambda_0\lambda_1}{2}+\alpha_7\lambda_1^2\right)\\&-2\phi_2\overline{X}\overline{Y}\left(\lambda_0\lambda_1-\frac{\vartheta\lambda_1^2}{2}\right)\end{aligned} \tag{A}$$

Applying expectation

$$\begin{aligned}MSE\left(U_{P_7}\right)=&\overline{Y}^2\left(\eta_0+\frac{\vartheta^2\eta_1}{4}-\vartheta\eta_{01}\right)+\phi_1^2U_{A7(i)}+\phi_2^2U_{B7(i)}+2\phi_1\phi_2U_{C7(i)}\\&-2\phi_1U_{D7(i)}-2\phi_2U_{E7(i)}\end{aligned} \tag{A39}$$

where,

$$U_{A7(i)}=\overline{Y}^2\left(1+\eta_0+\frac{\vartheta^2\eta_1}{4}-2\vartheta\eta_{01}+\frac{3}{4}\vartheta^2\eta_1\right),$$

$$U_{B7(i)}=\overline{X}^2\eta_1,$$

$$U_{C7(i)}=\overline{Y}\overline{X}\left(\vartheta\eta_1-\eta_{01}\right),$$

$$U_{D7(i)}=\overline{Y}^2\left(\eta_0-\frac{\vartheta\eta_{01}}{2}-\frac{\vartheta\eta_{01}}{2}+\frac{\vartheta^2\eta_1}{4}-\frac{\vartheta\eta_{01}}{2}+\alpha_7\eta_1\right),$$

$$U_{E7(i)}=\overline{X}\overline{Y}\left(\eta_{01}-\frac{\vartheta\eta_1}{2}\right).$$

Differentiating MSE partially, with respect to $\phi_1$and ϕ_2 and equating to zero, we get the following optimum values of $\phi_1$and ϕ_2:

$$\phi_{1(opt)}=\frac{U_{B7(i)}U_{D7(i)}-U_{C7(i)}U_{E7(i)}}{U_{A7(i)}U_{B7(i)}-U_{C7(i)}^2},\ \phi_{2(opt)}=\frac{U_{A7(i)}U_{E7(i)}-U_{C7(i)}U_{D7(i)}}{U_{A7(i)}U_{B7(i)}-U_{C7(i)}^2}. \qquad (say)$$

The minimum MSE of U_{P_7}

$$MSE_{\min}\left(U_{P_7}\right)=\bar{Y}^2\left(\eta_0+\frac{\vartheta^2\eta_1}{4}-\vartheta\eta_{01}\right)-\frac{U_{A7(i)}U_{E7(i)}^{\ 2}+U_{B7(i)}U_{D7(i)}^{\ 2}-2U_{C7(i)}U_{D7(i)}U_{E7(i)}}{U_{A7(i)}U_{B7(i)}-U_{C7(i)}^{\ 2}}. \tag{A40}$$

8th Estimator

Substituting values of $\overline{y_{st}}$ and $\overline{x_{st}}$ in equation (6.51)

$$=\left[\phi_1\bar{Y}\left(\frac{1}{2}+\frac{\lambda_0}{2}\right)\left\{\exp(-\lambda_1)(2+\lambda_1)^{-1}+\exp(\lambda_1)(2+\lambda_1)^{-1}\right\}+\phi_1\bar{Y}(1+\lambda_0)+\phi_2\left(-\bar{X}\lambda_1\right)\right] \left(1-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^2\lambda_1^{\ 2}}{8}\right) \tag{A41}$$

$$U_{P_8}-\bar{Y}=\bar{Y}\left(1+\lambda_0-\frac{\vartheta\lambda_1}{2}-\frac{\vartheta\lambda_0\lambda_1}{2}+\alpha_8\lambda_1^{\ 2}\right)+\phi_1\bar{Y}\left(1-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^2\lambda_1^{\ 2}}{8}+\lambda_0-\frac{\vartheta\lambda_1\lambda_0}{2}\right) -\phi_2\bar{X}\left(\lambda_1-\frac{\vartheta^2\lambda_1^{\ 2}}{2}\right)-\bar{Y} \tag{A42}$$

Applying expectation

$$bias\left(U_{P_8}\right)=\bar{Y}\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_8\eta_1\right)+\phi_1\bar{Y}\left(1+\frac{3\vartheta^2\eta_1}{8}-\frac{\vartheta\eta_{01}}{2}\right)+\phi_2\bar{X}\frac{\vartheta^2\eta_1}{2}-\bar{Y}, \tag{A43}$$

$$\begin{aligned}\left(U_{P_8}-\bar{Y}\right)^2=&\bar{Y}^2\left(\lambda_0^{\ 2}+\frac{\vartheta^2\lambda_1^{\ 2}}{4}-\vartheta\lambda_0\lambda_1\right)+\phi_1^{\ 2}\bar{Y}^2\left(1+\lambda_0^{\ 2}+\frac{\vartheta^2\lambda_1^{\ 2}}{4}-2\vartheta\lambda_0\lambda_1+\frac{3}{4}\vartheta^2\lambda_1^{\ 2}\right)+\phi_2^{\ 2}\bar{X}^2\lambda_1^{\ 2}\\&+2\phi_1\phi_2\bar{Y}\bar{X}\left(\vartheta\lambda_1^{\ 2}-\lambda_0\lambda_1\right)-2\phi_1\bar{Y}^2\left(\lambda_0^{\ 2}-\frac{\vartheta\lambda_0\lambda_1}{2}-\frac{\vartheta\lambda_0\lambda_1}{2}+\frac{\vartheta^2\lambda_1^{\ 2}}{4}-\frac{\vartheta\lambda_0\lambda_1}{2}+\alpha_8\lambda_1^{\ 2}\right)\\&-2\phi_2\bar{X}\bar{Y}\left(\lambda_0\lambda_1-\frac{\vartheta\lambda_1^{\ 2}}{2}\right)\end{aligned} \tag{A44}$$

Applying expectation

$$MSE\left(U_{P_8}\right)=\bar{Y}^2\left(\eta_0+\frac{\vartheta^2\eta_1}{4}-\vartheta\eta_{01}\right)+\phi_1^{\ 2}U_{A8(i)}+\phi_2^{\ 2}U_{B8(i)}+2\phi_1\phi_2U_{C8(i)} -2\phi_1U_{D8(i)}-2\phi_2U_{E8(i)} \tag{A45}$$

where,

$$U_{A8(i)}=\bar{Y}^2\left(1+\eta_0+\frac{\vartheta^2\eta_1}{4}-2\vartheta\eta_{01}+\frac{3}{4}\vartheta^2\eta_1\right),$$

$$U_{B8(i)}=\overline{X}^{2}\eta_{1},$$

$$U_{C8(i)}=\overline{Y}\overline{X}\left(\vartheta\eta_{1}-\eta_{01}\right),$$

$$U_{D8(i)}=\overline{Y}^{2}\left(\eta_{0}-\frac{\vartheta\eta_{01}}{2}-\frac{\vartheta\eta_{01}}{2}+\frac{\vartheta^{2}\eta_{1}}{4}-\frac{\vartheta\eta_{01}}{2}+\alpha_{8}\eta_{1}\right),$$

$$U_{E8(i)}=\overline{X}\overline{Y}\left(\eta_{01}-\frac{\vartheta\eta_{1}}{2}\right).$$

Differentiating MSE partially, with respect to $\emptyset_1$and $\emptyset_2$ and equating to zero, we get the following optimum values of $\emptyset_1$and $\emptyset_2$:

$$\phi_{1(opt)}=\frac{U_{B8(i)}U_{D8(i)}-U_{C8(i)}U_{E8(i)}}{U_{A8(i)}U_{B8(i)}-U_{C8(i)}^{2}},\quad \phi_{2(opt)}=\frac{U_{A8(i)}U_{E8(i)}-U_{C8(i)}U_{D8(i)}}{U_{A8(i)}U_{B8(i)}-U_{C8(i)}^{2}}, \qquad (say)$$

The minimum MSE of U_{p_8}

$$MSE_{\min}\left(U_{p_8}\right)=\overline{Y}^{2}\left(\eta_{0}+\frac{\vartheta^{2}\eta_{1}}{4}-\vartheta\eta_{01}\right)-\frac{U_{A8(i)}U_{E8(i)}^{2}+U_{B8(i)}U_{D8(i)}^{2}-2U_{C8(i)}U_{D8(i)}U_{E8(i)}}{U_{A8(i)}U_{B8(i)}-U_{C8(i)}^{2}}. \qquad (A46)$$

9th Estimator

Substituting values of $\overline{y_{st}}$ and $\overline{x_{st}}$ in equation (6.55)

$$=\left[\overline{Y}\left(\frac{1}{4}+\frac{\lambda_0}{4}\right)(2+\lambda_1^{2})\left\{\exp\left(\frac{-\lambda_1}{2}\right)\left(1-\frac{\lambda_1}{2}+\frac{\lambda_1^{2}}{4}\right)+\exp\left(\frac{\lambda_1}{2}\right)\left(1-\frac{\lambda_1}{2}+\frac{\lambda_1^{2}}{4}\right)\right\}+\phi_1\overline{Y}(1+\lambda_0)+\phi_2\left(-\overline{X}\lambda_1\right)\right]$$
$$\left(1-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^{2}\lambda_1^{2}}{8}\right) \qquad (A47)$$

$$U_{p_9}-\overline{Y}=\overline{Y}\left(1+\lambda_0-\frac{\vartheta\lambda_1}{2}-\frac{\vartheta\lambda_0\lambda_1}{2}+\alpha_9\lambda_1^{2}\right)+\phi_1\overline{Y}\left(1-\frac{\vartheta\lambda_1}{2}+\frac{3\vartheta^{2}\lambda_1^{2}}{8}+\lambda_0-\frac{\vartheta\lambda_1\lambda_0}{2}\right)$$
$$-\phi_2\overline{X}\left(\lambda_1-\frac{\vartheta^{2}\lambda_1^{2}}{2}\right)-\overline{Y} \qquad (A48)$$

Applying expectation

$$bias\left(U_{p_9}\right)=\bar{Y}\left(1-\frac{\vartheta\eta_{01}}{2}+\alpha_9\eta_1\right)+\phi_1\bar{Y}\left(1+\frac{3\vartheta^2\eta_1}{8}-\frac{\vartheta\eta_{10}}{2}\right)+\phi_2\bar{X}\frac{\vartheta^2\eta_1}{2}-\bar{Y}, \tag{A49}$$

$$\begin{aligned}\left(U_{p_9}-\bar{Y}\right)^2&=\bar{Y}^2\left(\lambda_0^2+\frac{\vartheta^2\lambda_1^2}{4}-\vartheta\lambda_0\lambda_1\right)+\phi_1^2\bar{Y}^2\left(1+\lambda_0^2+\frac{\vartheta^2\lambda_1^2}{4}-2\vartheta\lambda_0\lambda_1+\frac{3}{4}\vartheta^2\lambda_1^2\right)+\phi_2^2\bar{X}^2\lambda_1^2\\&+2\phi_1\phi_2\bar{Y}\bar{X}\left(\vartheta\lambda_1^2-\lambda_0\lambda_1\right)-2\phi_1\bar{Y}^2\left(\lambda_0^2-\frac{\vartheta\lambda_0\lambda_1}{2}-\frac{\vartheta\lambda_0\lambda_1}{2}+\frac{\vartheta^2\lambda_1^2}{4}-\frac{\vartheta\lambda_0\lambda_1}{2}+\alpha_9\lambda_1^2\right)\\&-2\phi_2\bar{X}\bar{Y}\left(\lambda_0\lambda_1-\frac{\vartheta\lambda_1^2}{2}\right)\end{aligned} \tag{A50}$$

Applying expectation

$$\begin{aligned}MSE\left(U_{p_9}\right)&=\bar{Y}^2\left(\eta_0+\frac{\vartheta^2\eta_1}{4}-\vartheta\eta_{01}\right)+\phi_1^2U_{A9(i)}+\phi_2^2U_{B9(i)}+2\phi_1\phi_2U_{C9(i)}\\&-2\phi_1U_{D9(i)}-2\phi_2U_{E9(i)}\end{aligned} \tag{A51}$$

where

$$U_{A9(i)}=\bar{Y}^2\left(1+\eta_0+\frac{\vartheta^2\eta_1}{4}-2\vartheta\eta_{01}+\frac{3}{4}\vartheta^2\eta_1\right),$$

$$U_{B9(i)}=\bar{X}^2\eta_1,$$

$$U_{C9(i)}=\bar{Y}\bar{X}\left(\vartheta\eta_1-\eta_{01}\right),$$

$$U_{D9(i)}=\bar{Y}^2\left(\eta_0-\frac{\vartheta\eta_{01}}{2}-\frac{\vartheta\eta_{01}}{2}+\frac{\vartheta^2\eta_1}{4}-\frac{\vartheta\eta_{01}}{2}+\alpha_9\eta_1\right),$$

$$U_{E9(i)}=\bar{X}\bar{Y}\left(\eta_{01}-\frac{\vartheta\eta_1}{2}\right).$$

Differentiating MSE partially, with respect to ϕ_1 and ϕ_2 and equating to zero, we get the following optimum values of ϕ_1 and ϕ_2

$$\phi_{1(opt)}=\frac{U_{B9(i)}U_{D9(i)}-U_{C9(i)}U_{E9(i)}}{U_{A9(i)}U_{B9(i)}-U_{C9(i)}^2},\qquad \phi_{2(opt)}=\frac{U_{A9(i)}U_{E9(i)}-U_{C9(i)}U_{D9(i)}}{U_{A9(i)}U_{B9(i)}-U_{C9(i)}^2}. \qquad (say)$$

The minimum MSE of U_{p_9}

$$MSE_{\min}\left(U_{P_9}\right)=\bar{Y}^2\left(\eta_0+\frac{\vartheta^2\eta_1}{4}-\vartheta\eta_{01}\right)-\frac{U_{A9(i)}U_{E9(i)}{}^2+U_{B9(i)}U_{D9(i)}{}^2-2U_{C9(i)}U_{D9(i)}U_{E9(i)}}{U_{A9(i)}U_{B9(i)}-U_{C9(i)}{}^2} \quad . \tag{A52}$$