ZATEZALO, TANJA, Ph.D. Generalized Mixture Estimators for the Finite Population Mean. (2016)
Directed by Dr. Sat Narain Gupta and Dr. Javid Shabbir. 117 pp.

The first order approximation of the theoretical mean square error and assumption of bivariate normality are very often used for the ratio type estimators for the population mean and variance. We have examined the adequacy of the first order approximation and the robustness of various ratio type estimators. We observed that the first order approximation for ratio type mean estimators and ratio type variance estimators works well if the sampling fraction is small and that departure from the assumption of bivariate normality is not a problem for large samples. We have also proposed some generalized mixture estimators which are combinations of the commonly used estimators. We have also extended the proposed generalized mixture estimators to the case when the study variable is sensitive and a non sensitive auxiliary variable is available. We have shown that the proposed generalized mixture estimators are more efficient than other commonly used estimators. An extensive simulation study and numerical examples are also presented.

GENERALIZED MIXTURE ESTIMATORS FOR THE FINITE POPULATION MEAN

by

Tanja Zatezalo

A Dissertation Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Greensboro
2016

Approved by

____
Committee Co-Chair

____
Committee Co-Chair

To my husband, Milan

and our beautiful son, Stefan

APPROVAL PAGE

This dissertation written by Tanja Zatezalo has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Co-Chair ______________________
Sat Narain Gupta

Committee Co-Chair ______________________
Javid Shabbir

Committee Members ______________________
Scott James Richter

____
Haimeng Zhang

____
Shanmugathasan Suthaharan

____
Date of Acceptance by Committee

____
Date of Final Oral Examination

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisors Prof. Sat Narain Gupta and Prof. Javid Shabbir for their guidance, support and encouragement during this dissertation work. I would like especially to thank Dr. Sat Gupta for mentoring me through of all my PhD study. His continuous support, guidance, and dedication have helped me to grow as a researcher and finish this long journey. I would like to thank Prof. Javid Shabbir for collaboration in all of my research projects in this dissertation, his dedication and believing in me as a researcher.

Also, I would like to thank the rest of my thesis committee: Prof. Scott James Richter, Prof. Shanmugathasan Suthaharan, and Prof. Haimeng Zhang, for their insightful comments, encouragement, and input during this dissertation work. Thanks goes to our Director of Graduate Studies Prof. Greg Bell for his selfless guidance and all of his help.

I would like to thank my co-author Prof. Subhash Kumar Yadav from Department of Mathematics and Statistics (A Centre of Excellence) Dr. RML. Avadh University, Faizabad, India whose collaboration improved this work.

Last but not the least, I would like to thank my family for supporting and be motivation through all my hard work and long studies.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 General Discussion on Generalized Mixture Estimators

The purpose of a sample survey is to obtain information about the population based on a random sample. By "population" we mean a group of units defined according to the objective of the survey. Thus the population may consist of all the fields under a specified crop or all the agricultural holdings larger than a specified size, as in an agricultural survey; or all the households having four or more children, as in a socio-economic survey. Of course, the population may also refer to the human population of a country. The information we want to get maybe, for example, the total number of units, such as the number of farms that grow corn; or aggregate values of various characteristics, such as the total area under corn. We may also look for the mean of various characteristics per unit, such as the mean household size; or the proportion of units which have certain characteristics, or the proportion of households having income over a given level or having five or more children.

In survey research, there are situations in which the information is available for every unit in the population. If a variable's value is known for every unit of the population, then it is not a variable of direct interest. Instead it maybe employed to improve the sampling plan or to improve the estimation of another variable of interest. Such a variable is called an auxiliary variable. Ratio, product, and regression type estimators rely on the use of an auxiliary variable to estimate parameters of the study variable. Auxiliary variables have been used by various authors in various estimation situations.

Cochran (1940)[6] introduced the use of an auxiliary variable at the estimation stage and proposed the ratio estimator for the population mean. It is well known that the ratio type mean estimator ensures better efficiency than the sample mean estimator if the study variable and an auxiliary variable have strong positive correlations. For situations when they are negatively correlated, the product estimator was introduced by Robson (1957)[47]. The product estimator is also more efficient than the sample mean estimator.

The regression estimator is used when the regression line between the study variable and the auxiliary variable does not pass through the origin. It is a well known that the regression estimator is more efficient than the ratio estimator and the sample mean estimators for $\rho_{yx} \neq 0$. Modified ratio, product, and regression type estimators have been introduced by different authors.

Mohanty (1967)[33] used two auxiliary variables by combining the regression and ratio estimators. Srivastava (1971)[70] introduced a generalized estimator for the population mean using multiple auxiliary variables. Bahl & Tuteja (1991) [2] introduced exponential ratio and product type estimators which perform better than the ordinary ratio and product estimators, respectively. Also, in the same year, Rao (1991)[42] proposed a regression type estimator which performs better than ordinary regression estimator. Samiuddin and Hanif (2006)[49] combined the ratio and the regression estimators by using two auxiliary variables and improved Mohanty's(1967)[33] estimator.

Development continued in the form of exponential estimators for different situations such as the work by Singh and Vishwakarma (2007)[67] in double sampling and Sanaullah et al. (2014)[50] in stratified two-phase random sampling. Singh et al. (2008) [64] proposed a ratio-product type exponential estimator which is more efficient than ordinary exponential ratio and product type estimators of Bahl & Tuteja (1991).

Grover & Kaur (2011)[12] introduced a regression-exponential type estimator of the mean. Subramani (2013)[77] proposed a generalized modified ratio estimator for estimating the population mean using the known population parameters of an auxiliary variable such as coefficient of variation, coefficient of kurtosis, coefficient of skewness, the coefficient of correlation, and various quartiles.

Asghar's et al. (2014)[1] proposed the generalized exponential type estimator for the population variance. Following them, Shabbir and Gupta (2015)[53] proposed a new generalized exponential type estimator for the population variance which performs better then Asghar (2014) et al. estimator.

In this dissertation, some new generalized mixture estimators of the population mean of the study variable by combining the ratio, product, exponential, and regression estimators will be proposed. The main aim is to gain efficiency in comparisons to the existing generalized mixture estimators, and also use these estimators to estimate the population mean of the sensitive study variable when a non sensitive auxiliary variable is used.

1.2 Basic Ratio, Product, Regression and Exponential Estimators

Let $U = \{U_1, \ldots, U_N\}$ be a finite population of size N and let (y_i, x_i) be the values of the study variable Y and an auxiliary variable X on the ith unit U_i, $i = 1, \ldots N$.

Let a sample of size n be drawn from this population, using simple random sampling without replacement. The goal is to estimate the population mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$. Let $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$ be the population variance of the study variable Y. Let $\bar{X} = \sum_{i=1}^{N}\frac{x_i}{N}$ be the population mean and $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$ be the population variances of the auxiliary variable X. Let $\bar{y} = \sum_{i=1}^{n}\frac{y_i}{n}$ and $\bar{x} = \sum_{i=1}^{n}\frac{x_i}{n}$ be the sample means of the study variable and an auxiliary variable respectively. Let $S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{Y})(x_i - \bar{X})$ be the population covariance between the study variable and the auxiliary variable. We assume that the population mean $\bar{X}$ and the population variance S_x^2 of an auxiliary variable are known. Let ρ_{yx} be the correlation coefficient between the study variable and an auxiliary. Also, assume $C_x = \frac{S_x}{\bar{X}}$ and $C_y = \frac{S_y}{\bar{Y}}$ are the coefficients of variation of the study variable Y and an auxiliary variable X, and $C_{xy} = \frac{S_{xy}}{\bar{Y}\bar{X}}$ is the coefficient of covariance between Y and X.

It is well known that the variance of the sample mean, the unbiased estimator, is given by $Var(\bar{y}) = \lambda\bar{Y}^2 C_y^2$, where $\lambda = \frac{1-f}{n}$ and $f = \frac{n}{N}$ is the sampling fraction. We give below some other commonly known men estimators.

1.2.1 The Ratio Estimator

The ordinary ratio estimator for the population mean $\bar{Y}$ of the study variable is given by Cochran (1940)[6] as:

$$t_R = \bar{y}\frac{\bar{X}}{\bar{x}}. \tag{1.1}$$

The bias and the mean square error respectively of this estimator, up to the first order approximation, are given by:

$$Bias(t_R) = E\left(t_R - \bar{Y}\right) \approx \lambda \bar{Y}\left(C_x^2 - \rho_{yx} C_x C_y\right), \tag{1.2}$$

$$MSE(t_R) = E\left(t_R - \bar{Y}\right)^2 \approx \lambda \bar{Y}^2\left(C_x^2 - 2\rho_{yx} C_x C_y + C_y^2\right). \tag{1.3}$$

If the sample size n is sufficiently large, then up to the first order of approximation, the ratio estimator will be more efficient than the ordinary sample mean estimator if

$$\rho_{yx} > \frac{C_x}{2C_y}. \tag{1.4}$$

For situations where $C_x \approx C_y$, condition (1.4) becomes $\rho_{yx} > \frac{1}{2}$.

1.2.2 The Product Estimator

The product estimator is used when the study variable Y and the auxiliary variable X are negatively correlated. The estimator introduced by Robson (1957)[47], and revised by Murthy (1964) [34] is given by:

$$t_P = \bar{y}\frac{\bar{x}}{\bar{X}}. \tag{1.5}$$

The exact bias of the product estimator is given by:

$$Bias(t_P) = E(t_p - \bar{Y}) = \lambda \frac{S_{yx}}{\bar{X}}. \tag{1.6}$$

The mean square error, up to the first order of approximation, is given by:

$$MSE(t_P) = E(t_p - \bar{Y})^2 \approx \lambda \bar{Y}^2 \left(C_x^2 + 2\rho_{yx} C_x C_y + C_y^2 \right). \tag{1.7}$$

Up to the first order of approximation, the product estimator is more efficient than the ordinary sample mean if

$$\begin{aligned} & MSE(t_P) < Var(\bar{y}), \\ & \text{or if} \quad \rho_{yx} < -\frac{C_x}{2C_y}, \\ & \text{or if} \quad \rho_{yx} < -\frac{1}{2} \quad \text{when} \quad C_x \approx C_y. \end{aligned} \tag{1.8}$$

1.2.3 The Regression Estimator

The ratio type estimators often result in increased precision if the line of best fit of Y on X is linear and passes through the origin. If the line does not pass through the origin, it is better to use the regression estimator given by:

$$t_{Reg} = \bar{y} + \hat{\beta}_{yx} \left(\bar{X} - \bar{x} \right), \tag{1.9}$$

where $\hat{\beta}_{yx} = \frac{s_{xy}}{s_x^2}$ is the sample regression coefficient between Y and X. The bias of the regression estimator, up to the first order of approximation, is given by:

$$Bias(t_{Reg}) = E(t_{Reg} - \bar{Y}) \approx -\lambda \beta_{yx} \left\{ \frac{\mu_{12}}{\mu_{11}} - \frac{\mu_{03}}{\mu_{02}} \right\}, \tag{1.10}$$

where $\beta_{yx} = \frac{S_{xy}}{S_{x^2}}$ is the population regression coefficient between the study variable Y and the auxiliary variable X, and $\mu_{rs} = \frac{1}{N-1} \sum_{i=1}^{N} \left(y_i - \bar{Y} \right)^r (x_i - \bar{X})^s$. Also $s_x^2 =$

$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$ is the sample variance of X and $s_{xy}=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$ is the sample covariance between X and Y.

The mean square error, up to the first order of approximation, is given by:

$$MSE(t_{Reg}) = E\left(t_{Reg}-\bar{Y}\right)^2 \approx \lambda \bar{Y}^2 C_y^2\left(1-\rho_{yx}^2\right), \tag{1.11}$$

The conditions under which the regression estimator is more efficient than the ordinary sample mean and ratio estimator are given below:

(1) the regression estimator t_{Reg} is more efficient than the ordinary sample mean $\bar{y}$ if

$$MSE(t_{Reg}) < Var(\bar{y}), \quad \text{or if} \quad C_y^2 - C_y^2\left(1-\rho_{yx}^2\right) > 0, \quad \text{or if} \quad \rho_{xy} \neq 0, \quad \text{and}$$

(2) the regression estimator t_{Reg} is more efficient than the ratio estimator t_R if

$$MSE(t_{Reg}) < MSE(t_R), \text{ if } \ C_x^2 - 2\rho_{yx}C_xC_y + \rho_{yx}^2C_y^2 > 0, \text{ or if } \left(C_x-\rho_{yx}C_y\right)^2 > 0.$$

If the relationship between Y and X is linear, and passes through the origin, then the two estimators are equally efficient.

1.2.4 Bahl & Tuteja Exponential Estimators

The exponential type estimators are often used to improve efficiencies of the ratio and product type estimators and were introduced by Bahl and Tuteja (1991)[2] as:

$$t_{ER} = \bar{y}\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right), \quad \text{and} \tag{1.12}$$

$$t_{EP} = \bar{y}\exp\left(\frac{\bar{x}-\bar{X}}{\bar{X}+\bar{x}}\right). \tag{1.13}$$

The exponential part helps, since it captures the auxiliary variable effect for a longer duration. The bias of the exponential estimators, up to the first order of approximation, are given by:

$$Bias(t_{ER}) = \left(t_{ER} - \bar{Y}\right) \approx \lambda\bar{Y}\left(\frac{3}{8}C_x^2 - \frac{1}{2}\rho_{yx}C_yC_x\right), \quad \text{and} \tag{1.14}$$

$$Bias(t_{EP}) = \left(t_{EP} - \bar{Y}\right) \approx \lambda\bar{Y}\left(\frac{1}{2}\rho_{xy}C_yC_x - \frac{1}{8}C_x^2\right). \tag{1.15}$$

The mean square error of the exponential ratio and product type estimators, up to the first order of approximation, are given by:

$$MSE(t_{ER}) = \left(t_{ER} - \bar{Y}\right)^2 \approx \frac{1}{4}\lambda\bar{Y}^2\left(4C_y^2 + C_x^2 - 4\rho_{yx}C_yC_x\right), \quad \text{and} \tag{1.16}$$

$$MSE(t_{EP}) = \left(t_{EP} - \bar{Y}\right)^2 \approx \frac{1}{4}\lambda\bar{Y}^2\left(4C_y^2 + 4\rho_{yx}C_yC_x + C_x^2\right). \tag{1.17}$$

1.3 Randomized Response Methodology

In this dissertation, me also want to discuss mean estimators in situations where the study variable is sensitive and can not be observed directly. This is one of the most important issues in behavioral and social sciences. The respondents are sometimes asked a sensitive question such as their personal income, experiencing feelings of low self-worth and powerlessness, sexual orientation, number of sexual partners in last two years, number of miscarriages or abortions etc. To circumvent the social desirability bias (the tendency in people to present themselves in a socially acceptable light when they are confronted with a sensitive question) several methods have been developed. The randomized response technique (RRT) is one such method. In this dissertation our focus is on situations when the study sensitive variable Y can not be observed directly, but a highly correlated non-sensitive auxiliary variable X is observed directly. For example, the study variable may be the number of miscarriages or abortions, and the non sensitive auxiliary question may be the number of children for a woman. The Optional RRT models are models in which a respondent who considers a question sensitive provides a scrambled answer, and the rest provide a true answer to the sensitive question. This was discussed in Gupta et al. (2002,2006,2010)[14, 15, 18] and Kalucha et al.(2015)[27]. Our focus here is on non-optional RRT models and we will introduce a few of these models proposed by different authors.

1.3.1 Warner's (1971) model

Warner's (1971) [84] model is the quantitative additive version of Warner (1965)[83] Binary Randomized Response Technique and works as follows:

For a simple random sample with replacement, let Y the true response, and S be a scrambling variable with known mean $E(S) = \mu_s$ and known variance σ_s^2. The population mean μ_Y and the population variance σ_Y^2 of the study variable are unknown. Also, assume that the true response Y and scrambling variable S are independent. The reported response Z is the sum of the true response and the scrambling variable, and is given by:

$$Z = Y + S. \tag{1.18}$$

Since

$$E(Z) = \mu_Y + \mu_S, \tag{1.19}$$

it follows that an unbiased estimator of the sensitive variable mean is given by:

$$\hat{\mu}_Y = \bar{Z} - \mu_S. \tag{1.20}$$

The variance of this estimator is given by:

$$Var(\hat{\mu}_Y) = Var(\bar{Z}) = \frac{\sigma_Y^2}{n} + \frac{\sigma_s^2}{n}. \tag{1.21}$$

The second term in (1.21) is the "penalty" for randomizing. Also, note that an unbiased estimator for the variance is given by:

$$\hat{V}ar(\hat{\mu}_Y) = \frac{s_z^2}{n}, \quad \text{where} \quad s_z^2 \quad \text{is the sample variance of the reported responses.}$$

1.3.2 Sousa et al. (2010) Ratio Estimator

Many authors have estimated the mean of a sensitive variable when the primary variable is sensitive and there is no auxiliary variable available. Sousa et al. (2010)[69] proposed the ratio estimator of the mean for the sensitive variable Y which has a strong positive correlation with a non sensitive variable X. The model works as follows:

Let Y be the sensitive study variable, and X be the non-sensitive variable which is strongly (positively) correlated with Y. Let S be a scrambling variable independent of Y and X. Assume that the population mean $\bar{X}$ and the population variance S_x^2

of the auxiliary variable are known. Also, assume that the population mean and the population variance of the scrambling variable are known, and given as $\mu_S = 0$ and σ_s^2. The population mean $\bar{Y}$ and the population variance S_y^2 of the sensitive study variable are unknown. The respondent is asked to report a scrambled response for Y, but is asked to provided a true response for X. The reported response is the sum of the sensitive variable and the scrambling variable, and is given by:

$$Z = Y + S. \tag{1.22}$$

Note that $E(Z) = E(Y)$ since $\mu_s = 0$.

An estimator of the mean of the sensitive variable (Y) when the information on (X) is ignored, is the ordinary sample mean given by:

$$\hat{\mu}_Y = \bar{z}. \tag{1.23}$$

The mean square error of this unbiased estimator, when sampling is without replacement, is given by:

$$MSE(\hat{\mu}_Y) = \lambda S_z^2 = \lambda \left(S_y^2 + \sigma_s^2\right), \tag{1.24}$$

where $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2, \sigma_s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(s_i - \mu_S)^2$ and $S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$ are the population variances of the study variable, the scrambling variable and the auxiliary variable, respectively and $\lambda = \frac{1-f}{n}$ where $f = \frac{n}{N}$.

Sousa et al.(2010)[69] proposed the ratio estimator given by:

$$\hat{\mu}_R = \bar{z}\left(\frac{\bar{X}}{\bar{x}}\right). \tag{1.25}$$

The bias of this estimator, up to the second order of approximation is given as:

$$Bias^{(2)}(\hat{\mu}_R) = E(\hat{\mu}_R - \bar{Y}) \approx Bias^{(1)}(\hat{\mu}_R) + 3\lambda^3\bar{Y}\left(C_x^4 - \rho_{zx}C_zC_x^3\right), \tag{1.26}$$

where the bias up to the first order of approximation, is given by:

$$Bias^{(1)}(\hat{\mu}_R) \approx \lambda\bar{Y}\left(C_x^2 - \rho_{zx}C_zC_x\right). \tag{1.27}$$

The mean squared error, correct up to the second order of approximation, is given by:

$$MSE^{(2)}(\hat{\mu}_R) \approx MSE^{(1)}(\hat{\mu}_R) + 3\lambda^2\bar{Y}^2C_x^2\left[(1+2\rho_{zx}^2)C_z^2 + 3C_x^2 - 6\rho_{zx}C_zC_x\right] \tag{1.28}$$

where

$$MSE^{(1)}(\hat{\mu}_R) \approx \lambda^2\bar{Y}^2\left(C_z^2 + C_x^2 - 2\rho_{zx}C_zC_x^3\right). \tag{1.29}$$

is the corresponding mean square error up to the first of order of approximation. The difference between the two approximations for the mean square errors is given by:

$$3\lambda^2\bar{Y}^2C_x^2\left[(1+2\rho_{zx}^2)C_z^2 + 3C_x^2 - 6\rho_{zx}C_zC_x\right] \tag{1.30}$$

The difference (1.30) converges to zero as $n \to N$.

1.3.3 Gupta et al. (2012) Ordinary Regression Estimator Using RRT

Gupta et al.(2012)[19] proposed an ordinary regression estimator where the RRT estimator of the population mean $\bar{Y}$ of the sensitive study variable is improved by using a non-sensitive auxiliary variable X. The RRT regression estimator is given by:

$$\hat{\mu}_{Reg} = \bar{z} + \hat{\beta}_{zx}(\bar{X} - \bar{x}), \tag{1.31}$$

where $\hat{\beta}_{zx} = \frac{s_{zx}}{s_x^2}$ is the sample regression coefficient between Z and X, and $Z = Y + S$ is the scrambled response, where Y is the true response and S is a scrambeling variable.

The bias, up to the first order of approximation, is given by:

$$Bias(\hat{\mu}_{Reg}) \approx -\beta_{zx}\lambda \left\{ \frac{\mu_{12}}{\mu_{11}} - \frac{\mu_{03}}{\mu_{02}} \right\} \tag{1.32}$$

where $\beta_{zx} = \frac{S_{zx}}{S_x^2}$ is the population regression coefficient and $\mu_{rs} = \frac{1}{N-1} \sum_{i=1}^{N} \left(z_i - \bar{Z}\right) \left(x_i - \bar{X}\right)$. Also note that the following holds:

$$\beta_{zx} = \frac{S_{zx}}{S_x^2} = \frac{S_{yx}}{S_x^2} = \rho_{yx}\frac{S_y}{S_x} \quad \text{and} \quad \rho_{zx} = \frac{\rho_{yx}}{\sqrt{1 + \frac{\sigma_s^2}{S_y^2}}}, \tag{1.33}$$

where ρ_{yx} and ρ_{zx} are the coefficients of correlation between y and x, and z and x, respectively.

The mean square error, up to the first order of approximation, is given by:

$$MSE(\hat{\mu}_{Reg}) \approx \lambda \bar{Y}^2 C_z^2 \left(1 - \rho_{zx}^2\right) = \lambda S_y^2 \left[\left(1 + \frac{\sigma_s^2}{S_y^2}\right) - \rho_{yx}^2 \right]. \tag{1.34}$$

The conditions under which the RRT regression estimator is more efficient than the RRT ratio estimator and the RRT sample mean are given by:

(1) The RRT regression estimator $\hat{\mu}_{Reg}$ is more efficient than the RRT sample mean estimator if $\rho_{yx}^2 > 0$, and

(2) The RRT regression estimator $\hat{\mu}_{Reg}$ is more efficient than the RRT ratio estimator $\hat{\mu}_R$ if $(C_x - C_z\rho_{zx})^2 > 0$.

These conditions will always hold, indicating that up to the first order of approximation, the regression estimator performs better than the ordinary RRT sample mean and the RRT ratio estimator.

1.3.4 Gupta et al. (2012) Generalized Regression-Cum-Ratio Estimator

Many authors have used regression-cum-ratio estimators that combine the regression estimator and the ratio estimator. These include Ray and Singh (1981)[44], Perri (2004)[39], and Kadilar and Cingi (2004)[23]. Gupta et al. (2012)[19] proposed a similar hybrid estimator, as a generalized regression-cum-ratio estimator. The main idea was to see if further gains can be achieved by using a generalized regression-cum-ratio estimator, as compared to the Gupta et al (2012)[19] RRT regression estimator. The model works under the same conditions as the RRT regression estimator, and is given by:

$$\hat{\mu}_{GRR} = \left[k_1\bar{z} + k_2(\bar{X} - \bar{x})\right]\left(\frac{\bar{X}}{\bar{x}}\right), \tag{1.35}$$

where k_1 and k_2 are suitably chosen parameters. The bias, up to the first order of approximation, is given by:

$$Bias(\hat{\mu}_{GRR}) \approx (k_1 - 1)\bar{Z} + \lambda k_1 \bar{Z}\left(C_x^2 - \rho_{zx}C_zC_x\right) + \lambda k_2 \bar{X} C_x^2. \tag{1.36}$$

The minimum mean square error of the generalized regression-cum-ratio estimator, at the optimum values of k_1 and k_2 i.e.,

$$k_{1(opt)} = \frac{1 - \lambda C_x^2}{1 - \lambda\left[C_x^2 - C_z^2(1 - \rho_{zx}^2)\right]}, \quad \text{and} \quad k_{2(opt)} = \frac{\bar{Y}}{\bar{X}}\left[1 + k_{1(opt)}\left(\frac{\rho_{zx}C_z}{C_x} - 2\right)\right],$$

is given by:

$$MSE(\hat{\mu}_{GRR})_{min} \approx \bar{Y}^2 \frac{\lambda C_z^2 \left[1-\rho_{zx}^2\right]\left[1-\lambda C_x^2\right]}{\lambda C_z^2 \left[1-\rho_{zx}^2\right]+\left[1-\lambda C_x^2\right]}. \tag{1.37}$$

The conditions under which the generalized regression-cum-ratio estimator is more efficient than the ordinary RRT sample mean, RRT ratio estimator, and RRT regression estimator are given bellow:

(1)

$$MSE(\hat{\mu}_{GRR})_{min} < MSE(\hat{\mu}_Y) \quad \text{if} \quad \lambda\left(S_y^2+S_s^2\right)>0 \tag{1.38}$$

(2) $MSE(\hat{\mu}_{GRR})_{min} < MSE(\hat{\mu}_R)$ if

$$\left(\frac{C_x}{C_z}-\rho_{zx}\right)^2 + \frac{\lambda C_z^2\left(1-\rho_{zx}^2\right)}{\lambda C_z^2\left(1-\rho_{zx}^2\right)+\left(1-\lambda C_x^2\right)} > 0, \quad \text{and} \tag{1.39}$$

(3)

$$MSE(\hat{\mu}_{GRR})_{min} < MSE(\hat{\mu}_{Reg}) \quad \text{if} \quad \lambda C_z^2\left(1-\rho_{zx}^2\right)>0. \tag{1.40}$$

From these conditions, which always hold true, we can conclude that the generalized regression-cum-ratio estimator with optimal coefficients is always better than the ordinary RRT sample mean, RRT regression and RRT ratio estimators.

1.3.5 Koyuncu et al. (2014) Generalized Exponential Estimator

Many authors have studied exponential type estimators when the study variable Y is non sensitive. These include Bahl & Tuteja (1991)[2], Shabbir and Gupta

(2007)[51], Grover and Kaur (2011)[12] and Koyuncu (2012)[31]. Following Gupta et al. (2012)[19] and Bahl & Tuteja (1991)[2], Koyuncu et al. (2014)[32] proposed a generalized exponential type estimator of the mean $\bar{Y}$ of the sensitive study variable utilizing a non-sensitive auxiliary variable X. The model works under the same assumptions as the ordinary regression estimator and the generalized regression-cum-ratio estimator, and is given by:

$$\hat{\mu}_{GE} = \left[w_1\bar{z} + w_2(\bar{X} - \bar{x})\right] \exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right), \tag{1.41}$$

where w_1 and w_2 are the model parameters. The bias of this generalized exponential estimator, up to the first order of approximation, is given as:

$$Bias(\hat{\mu}_{GE}) \approx (w_1 - 1)\,\bar{Y} + \lambda w_1 \bar{Y}\left(\frac{3}{8}C_x^2 - \frac{1}{2}\rho_{zx}C_zC_x\right) + \frac{1}{2}w_2\lambda\bar{X}C_x^2. \tag{1.42}$$

The minimum mean square error of generalized exponential estimator, up to the first order of approximation, at the optimum values of w_1 and w_2 i.e.,

$$w_{1(opt)} = \frac{1 - \frac{1}{8}\lambda C_x^2}{1 + \lambda C_z^2\left(1 - \rho_{zx}^2\right)}, \quad \text{and}$$

$$w_{2(opt)} = \frac{\bar{Y}}{\bar{X}}\left[\frac{1}{2} - w_{1(opt)}\left(1 - \rho_{zx}\frac{C_z}{C_x}\right)\right],$$

is given by:

$$MSE_{min}(\hat{\mu}_{GE}) \approx \bar{Y}^2 \left[\left(1 - \frac{1}{4}\lambda C_x^2\right) - \frac{\left(1 - \frac{1}{8}\lambda C_x^2\right)^2}{1 + \lambda C_z^2 \left(1 - \rho_{zx}^2\right)} \right], \quad \text{or} \tag{1.43}$$

$$MSE_{min}(\hat{\mu}_{GE}) \approx \left\{ \frac{MSE(\hat{\mu}_{Reg})}{\left[1 + \frac{MSE(\hat{\mu}_{Reg})}{\bar{Y}^2}\right]} - \frac{\lambda C_x^2 \left[MSE(\hat{\mu}_{Reg}) + \lambda \frac{1}{16} C_x^2 \bar{Y}^2\right]}{4\left[1 + \frac{MSE(\hat{\mu}_{Reg})}{\bar{Y}^2}\right]} \right\}. \tag{1.44}$$

1.4 Motivation for this Work and Outline of the Dissertation

Many generalized mean estimators have been discussed in the earlier sections. The main motivation of this dissertation is to improve efficiency of some existing estimators by introducing some new generalized mixture estimators. The second motivation is to use the proposed generalized mixture estimators in the situations when the sensitive study variable cannot be observed directly and a non-sensitive auxiliary variable is available.

An outline of the work discussed in various chapters is given below.

Chapter 1 provides an introduction to the ordinary ratio, product, regression and exponential type estimators. Also some generalized estimators are discussed. We also provide an introduction to RRT models for the quantitative response. Modification of these RRT estimators using non-sensitive auxiliary variable is also introduced.

Chapter 2 focuses on the comparisons of the empirical mean square errors and the corresponding theoretical mean square errors for various ratio and product type mean estimators, as well as for some ratio type variance estimators. The purpose is to examine the adequacy of the first order approximation which is generally used in the calculation of mean square errors for rato estimators. Also we examine the robustness of these estimators since the main assumption for such situations is that Y and X have a bivariate normal distribution. A simulation study shows that the mean square error, up to the first order of approximation, generally works well whenever the sampling fraction $(\frac{n}{N})$ is small. Also, we have observed the the departure from the assumption of bivariate normality is not a problem when the sample size is large.

Chapter 3 introduces the generalized mixture estimators proposed by Zatezalo et al.(2016)[87], with mathematical derivations for the bias and mean square errors, up to the first order of approximation. The optimum values of the parameters involved, and the optimum mean square error, up to the first order of approximation, are derived. Corresponding theoretical and empirical comparisons with some commonly used generalized estimators are also presented.

Chapter 4 discusses the ordinary ratio, regression and some generalized mixture estimators when the study variable is sensitive in nature and a non-sensitive auxiliary variable is available. The mathematical derivations for the bias and the mean square error, up to the first order of approximation, are presented. Also, the minimum mean square errors for two special cases are derived. The efficiency comparisons with some existing RRT estimators of the sensitive variable in the presence of an auxiliary variable are also presented. Results of a numerical study are given at the end of this chapter.

Chapter 5 presents simulation results where we compare the estimators proposed in Chapter III and Chapter IV with different existing estimators.

Chapter 6 gives some concluding remarks and future research directions.

CHAPTER II

ADEQUACY OF THE FIRST ORDER APPROXIMATION FOR RATIO ESTIMATORS OF THE MEAN AND VARIANCE

2.1 Introduction

In many studies the first order approximation for the theoretical mean square error has been used for ratio type estimators for the population mean and variance. Bivariate normality is another commonly used assumption. The main focus of this chapter is on examining the adequacy of the first order approximation and also on examining the robustness of ratio estimators against departure from bivariate normality. We have calculated the theoretical mean square errors for many ratio type estimators, based on first order approximation, and the corresponding empirical mean square errors. We observed that the first order approximation for the ratio type mean and variance estimators generally works well as long as the sampling fraction is small. We also observed that departure from the assumption of bivariate normality is not a serious handicap for large samples. We will use the terminology introduced in Chapter I.

2.2 Some Ratio Estimators of the Mean

Often the characteristic Y under study is closely related to an auxiliary variable X, and summary data on X, such as the population mean $\bar{X}$ and the population variance S_x^2, are readily available. In such a situation it is convenient to consider estimators of the population mean $\bar{Y}$ and population variance S_y^2 that use information about X. Those estimators are generally more efficient than those based on a sample of Y alone if the correlation between X and Y is strong. Many modifications of the ratio and product estimators have been done to improve their efficiency by using a number of known parameters of the auxiliary variable such as the coefficient of variation C_x, the coefficient of kurtosis $\beta_2(x)$, standard deviation σ_x, the coefficient of skewness $\beta_1(x)$, the correlation coefficient between the study variable and an auxiliary variable ρ_{yx} and the quartiles Q_i's. Sisodia and Dwivedi (1981)[68] have suggested a modified ratio estimator using the coefficient of variation C_x of an auxiliary variable X for estimating the population mean $\bar{Y}$. Upadhyaya and Singh (1999)[80] suggested another modified ratio estimator using a linear combination of the coefficient of variation C_x and coefficient of the kurtosis $\beta_2(x)$. Singh and Tailor (2003) [62] proposed another estimator using the correlation coefficient ρ_{yx} between X and Y. By using the population variance S_x^2 of an auxiliary variable X, Singh (2003)[56] proposed another modified ratio estimator. Also, Singh used a linear combination of the coefficient of kurtosis $\beta_2(x)$ and standard deviation σ_x, and the coefficient of skewness $\beta_1(x)$ and standard deviation σ_x for estimating the population mean of the study variable $\bar{Y}$. Motivated by Singh (2003)[56], Yan and Tian (2010)[89] used a linear combination of the coefficient of kurtosis $\beta_2(x)$ and the coefficient of skewness $\beta_1(x)$, and the coefficient of variation C_x and the coefficient of skewness $\beta_1(x)$ of the auxiliary variable X. More recently, Subramani and Kumarapandiyan (2013)[76] suggested a new modified ratio estimator using known population median M_d of an auxiliary variable. Subramani and Kumarapandiyan (2012, 2013)[71–73] have also suggested modified ratio estimators using the known median and the coefficient of kurtosis, median and coefficient of skewness, median and the coefficient of variation, and median and the coefficient of correlation.

More detailed discussion about the ratio and product estimators of the population mean and their modification can be also found in Olkin (1958)[37], Pathak (1964)[38], Tin (1965)[79], Murthy (1967)[35], Reddy (1973)[45], David and Sukhatme (1974)[9], Cochran (1977)[7], Ray and Sahai (1980)[43], Naik and Gupta (1991)[36], Kadilar and Cingi (2003)[22], Singh and Espejo (2003)[61], Shabbir and Yaab (2003)[54], Jhajj et al. (2006)[21], Khoshnevisan et al. (2007)[29], Perri (2007)[40], Gupta and Shabbir (2008)[16], Singh and Agnihotri (2008)[60], Koyuncu and Kadilar (2009)[30] and Sharma and Taylor (2010)[55].

We now introduce some of the ratio type estimators and product type estimators with corresponding mean square errors for the purpose of examining the adequacy of the first order of approximation and robustness. These include the classical ratio estimator t_R, the product estimator t_P and the exponential ratio and product type estimator t_{ER}, all introduced in Chapter I. We also include many modified ratio and product estimators with corresponding characterising constants, the bias and the mean square error as given in Table 1. This was also discussed by Zatezalo et al. (2016)[86].

Table 1. Modified Ratio and Product Estimators of Population Mean With the Chracterising Constant, Bias and Mean Square Errors

Estimator	**Constant $\boldsymbol{\theta}_i$**	**Bias**	**Mean squared error**
$t_1 = \bar{y}\left(\frac{\bar{X}+C_x}{\bar{x}+C_x}\right)$	$\theta_1 = \frac{\bar{X}}{\bar{X}+C_x}$	$\lambda\bar{Y}\left[\theta_1 C_x^2\left(\theta_1 - C\right)\right]$	$\lambda\bar{Y}^2\left[C_y^2 + \theta_1 C_x^2\left(\theta_1 - 2C\right)\right]$
$t_2 = \bar{y}\left(\frac{\bar{x}+C_x}{\bar{X}+C_x}\right)$	$\theta_2 = \frac{\bar{X}}{\bar{X}+C_x}$	$\lambda\bar{Y}\left[\theta_2 C_x^2\left(\theta_2 + C\right)\right]$	$\lambda\bar{Y}^2\left[C_y^2 + \theta_2 C_x^2\left(\theta_2 + 2C\right)\right]$
$t_3 = \bar{y}\left(\frac{\beta_2(x)\bar{x}+C_x}{\beta_2(x)\bar{X}+C_x}\right)$	$\theta_3 = \frac{\beta_2(x)\bar{X}}{\beta_2(x)\bar{X}+C_x}$	$\lambda\bar{Y}\left[\theta_3 C_x^2\left(\theta_3 + C\right)\right]$	$\lambda\bar{Y}^2\left[C_y^2 + \theta_3 C_x^2\left(\theta_3 + 2C\right)\right]$

Continued on next page

Estimator	Constant $\boldsymbol{\theta}_i$	Bias	Mean square error
$t_4 = \bar{y}\left(\frac{C_x\bar{x}+\beta_2(x)}{C_x\bar{X}+\beta_2(x)}\right)$	$\theta_4 = \frac{C_x\bar{X}}{C_x\bar{X}+\beta_2(x)}$	$\lambda\bar{Y}\Big[\theta_4 C_x^2\left(\theta_4+C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_4 C_x^2\left(\theta_4+2C\right)\Big]$
$t_5 = \bar{y}\left(\frac{\bar{x}+\sigma_x}{\bar{X}+\sigma_x}\right)$	$\theta_5 = \frac{\bar{X}}{\bar{X}+\sigma_x}$	$\lambda\bar{Y}\Big[\theta_5 C_x^2\left(\theta_5+C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_5 C_x^2\left(\theta_5+2C\right)\Big]$
$t_6 = \bar{y}\left(\frac{\beta_1(x)\bar{x}+\sigma_x}{\beta_1(x)\bar{X}+\sigma_x}\right)$	$\theta_6 = \frac{\beta_1(x)\bar{X}}{\beta_1(x)\bar{X}+\sigma_x}$	$\lambda\bar{Y}\Big[\theta_6 C_x^2\left(\theta_6+C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_6 C_x^2\left(\theta_6+2C\right)\Big]$
$t_7 = \bar{y}\left(\frac{\beta_2(x)\bar{x}+\sigma_x}{\beta_2(x)\bar{X}+\sigma_x}\right)$	$\theta_7 = \frac{\beta_2(x)\bar{X}}{\beta_2(x)\bar{X}+\sigma_x}$	$\lambda\bar{Y}\Big[\theta_7 C_x^2\left(\theta_7+C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_7 C_x^2\left(\theta_7+2C\right)\Big]$
$t_8 = \bar{y}\left(\frac{\bar{X}+\rho_{yx}}{\bar{x}+\rho_{yx}}\right)$	$\theta_8 = \frac{\bar{X}}{\bar{X}+\rho_{yx}}$	$\lambda\bar{Y}\Big[\theta_8 C_x^2\left(\theta_8-C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_8 C_x^2\left(\theta_8-2C\right)\Big]$
$t_9 = \bar{y}\left(\frac{\bar{x}+\rho_{yx}}{\bar{X}+\rho_{yx}}\right)$	$\theta_9 = \frac{\bar{X}}{\bar{X}+\rho_{yx}}$	$\lambda\bar{Y}\Big[\theta_9 C_x^2\left(\theta_9+C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_9 C_x^2\left(\theta_9+2C\right)\Big]$
$t_{10} = \bar{y}\left(\frac{\bar{X}+\beta_2(x)}{\bar{x}+\beta_2(x)}\right)$	$\theta_{10} = \frac{\bar{X}}{\bar{X}+\beta_2(x)}$	$\lambda\bar{Y}\Big[\theta_{10} C_x^2\left(\theta_{10}-C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_{10} C_x\left(\theta_{10}-2C\right)\Big]$
$t_{11} = \bar{y}\left(\frac{\bar{x}+\beta_2(x)}{\bar{X}+\beta_2(x)}\right)$	$\theta_{11} = \frac{\bar{X}}{\bar{X}+\beta_2(x)}$	$\lambda\bar{Y}\Big[\theta_{11} C_x^2\left(\theta_{11}+C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_{11} C_x^2\left(\theta_{11}+2C\right)\Big]$
$t_{12} = \bar{y}\left(\frac{\beta_2\bar{X}+\beta_1(x)}{\beta_2\bar{x}+\beta_1(x)}\right)$	$\theta_{12} = \frac{\beta_2\bar{X}}{\beta_2(x)\bar{X}+\beta_1(x)}$	$\lambda\bar{Y}\Big[\theta_{12} C_x^2\left(\theta_{12}-C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_{12} C_x^2\left(\theta_{12}-2C\right)\Big]$
$t_{13} = \bar{y}\left(\frac{\bar{X}+\sigma_x}{\bar{x}+\sigma_x}\right)$	$\theta_{13} = \frac{\bar{X}}{\bar{X}+\sigma_x}$	$\lambda\bar{Y}\Big[\theta_{13} C_x^2\left(\theta_{13}-C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_{13} C_x^2\left(\theta_{13}-2C\right)\Big]$
$t_{14} = \bar{y}\left(\frac{\bar{X}+M_d}{\bar{x}+M_d}\right)$	$\theta_{14} = \frac{\bar{X}}{\bar{X}+M_d}$	$\lambda\bar{Y}\Big[\theta_{14} C_x^2\left(\theta_{14}-C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_{14} C_x^2\left(\theta_{14}-2C\right)\Big]$
$t_{15} = \bar{y}\left(\frac{\beta_2(x)\bar{X}+C_x}{\beta_2(x)\bar{x}+C_x}\right)$	$\theta_{15} = \frac{\beta_2(x)\bar{X}}{\beta_2(x)\bar{X}+C_x}$	$\lambda\bar{Y}\Big[\theta_{15} C_x^2\left(\theta_{15}-C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_{15} C_x^2\left(\theta_{15}-2C\right)\Big]$
$t_{16} = \bar{y}\left(\frac{C_x\bar{X}+\rho_{xy}}{C_x\bar{x}+\rho_{xy}}\right)$	$\theta_{16} = \frac{C_x\bar{X}}{C_x\bar{X}+\rho_{xy}}$	$\lambda\bar{Y}\Big[\theta_{16} C_x^2\left(\theta_{16}-C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_{16} C_x^2\left(\theta_{16}-2C\right)\Big]$
$t_{17} = \bar{y}\left(\frac{C_x\bar{X}+M_d}{C_x\bar{x}+M_d}\right)$	$\theta_{17} = \frac{C_x\bar{X}}{C_x\bar{X}+M_d}$	$\lambda\bar{Y}\Big[\theta_{17} C_x^2\left(\theta_{17}-C\right)\Big]$	$\lambda\bar{Y}^2\Big[C_y^2+\theta_{17} C_x^2\left(\theta_{17}-2C\right)\Big]$

Continued on next page

Estimator	Constant $\boldsymbol{\theta}_i$	Bias	Mean square error
$t_{18} = \bar{y}\left(\frac{\beta_1(x)\bar{X}+\rho_{yx}}{\beta_1(x)\bar{x}+\rho_{yx}}\right)$	$\theta_{18} = \frac{\beta_1(x)\bar{X}}{\beta_1(x)\bar{X}+\rho_{xy}}$	$\lambda\bar{Y}\left[\theta_{18}C_x^2\left(\theta_{18}-C\right)\right]$	$\lambda\bar{Y}^2\left[C_y^2+\theta_{18}C_x^2\left(\theta_{18}-2C\right)\right]$
$t_{19} = \bar{y}\left(\frac{\rho_{yx}\bar{X}+\sigma_x}{\rho_{yx}\bar{x}+\sigma_x}\right)$	$\theta_{19} = \frac{\rho_{yx}\bar{X}}{\rho_{yx}\bar{X}+\sigma_x}$	$\lambda\bar{Y}\left[\theta_{19}C_x^2\left(\theta_{19}-C\right)\right]$	$\lambda\bar{Y}^2\left[C_y^2+\theta_{19}C_x^2\left(\theta_{19}-2C\right)\right]$
$t_{20} = \bar{y}\left(\frac{\rho_{xy}\bar{X}+M_d}{\rho_{xy}\bar{x}+M_d}\right)$	$\theta_{20} = \frac{\rho_{xy}\bar{X}}{\rho_{xy}\bar{X}+M_d}$	$\lambda\bar{Y}\left[\theta_{20}C_x^2\left(\theta_{20}-C\right)\right]$	$\lambda\bar{Y}^2\left[C_y^2+\theta_{20}C_x^2\left(\theta_{20}-2C\right)\right]$
$t_{21} = \bar{y}\left(\frac{\sigma_x\bar{X}+M_d}{\sigma_x\bar{x}+M_d}\right)$	$\theta_{21} = \frac{\sigma_x\bar{X}}{\sigma_x\bar{X}+M_d}$	$\lambda\bar{Y}\left[\theta_{21}C_x^2\left(\theta_{21}-C\right)\right]$	$\lambda\bar{Y}^2\left[C_y^2+\theta_{21}C_x^2\left(\theta_{21}-2C\right)\right]$

The bias and theoretical mean square errors for the estimators t_1 to t_{21}, up to first order of approximation, can be represented in a single expressions as:

$$Bias(t_i) = \lambda\bar{Y}\left[\theta_i C_x^2\left(\theta_i \pm C\right)\right], \quad i = 1 \dots 21 \tag{2.1}$$

and

$$MSE(t_i) = \lambda\bar{Y}^2\left[C_y^2 + \theta_i C_x^2\left(\theta_i \pm 2C\right)\right], \quad i = 1 \dots 21, \tag{2.2}$$

where (+) sign is used for the product estimators and (−) sign is used for the ratio estimators. Also,

$$C_y = \frac{S_y}{\bar{Y}},\ C_x = \frac{S_x}{\bar{X}},\ S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2,\ S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2, \tag{2.3}$$

$$\lambda = \frac{1-f}{n},\ f = \frac{n}{N}\ \rho_{xy} = \frac{S_{yx}}{S_y S_x},\ S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})(X_i - \bar{X}),\ C = \rho_{xy}\frac{Cy}{Cx}.$$

2.3 Some Ratio Estimators of the Variance

The ratio type variance estimators are used to improve the precision of the sample variance estimator when the study variable Y is positively correlated with an auxiliary variable X. Isaki (1983)[20] proposed a ratio type variance estimator of the population variance S_y^2 when the population variance S_x^2 of an auxiliary variable X is known. Further improvements over the classical ratio estimator are also achieved by introducing a number of modified ratio estimators with the use of known parameters such as the coefficient of variation C_x and coefficient of kurtosis $\beta_2(x)$. The problem of constructing efficient estimators for the population variance has been widely discussed by various authors such as Das and Tripathi (1978)[8], Wolter (1985)[85], Prasad and Singh (1990)[41], Garcia and Cebrain (1997)[10], Upadhyaya and Singh (2006)[81], Gupta and Shabbir (2008)[17], Bhushan (2012)[4], Subramani and Kumarapandiyan (2012b, 2012c)[74, 75], and Singh et al.(1988,2003)[57, 58]. Motivated by Sisoda and Dwivedi (1981)[68], Uphadhyaya and Singh (1999)[80] and Singh et al.(2004)[63], Kadilar and Cingi (2006)[24,25] suggested four types of variance estimators using known values of the coefficient of variation C_x and the coefficient of kurtosis $\beta_2(x)$ of an auxiliary variable X. Singh et al. (2011)[59] proposed the exponential ratio type estimator for the population variance with the aim to improve the efficiency of the existing ratio estimators. Also, Subramani and Kumarpandiyan (2012c)[75] suggested the modified ratio type estimators using the quartiles of the auxiliary variable. The modified ratio type estimators are biased, but have smaller mean squared errors compared to the traditional ratio type variance estimator. Following Kadilar and Cingi (2006)[25], Subramani and Kumarapandiyan (2013)[76] proposed the ratio type estimators of the population variance S_y^2 using known values of the coefficient of variation C_x and the population median Q_2 of an auxiliary variable X. Recently Khan and Shabbir (2013)[28] proposed another ratio type estimator of the population variance using known values of the coefficient of correlation ρ_{yx} and the population upper quartile Q_3 of an auxiliary variable. Following Singh et al. (2011)[59] and motivated by Upadhyaya et al. (2011)[82], Yadav and Kadilar (2013)[88] proposed an improved generalized ratio exponential type estimator of the population variance. We discus some of these estimators below.

The sample variance estimator of the population variance is defined as:

$$\hat{S}_y^2 = s_y^2, \tag{2.4}$$

which is an unbiased estimator. Its variance is given by:

$$V(\hat{S}_y^2) = \gamma S_y^4 \left(\lambda_{40} - 1\right), \tag{2.5}$$

where

$$\lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{\frac{r}{2}}\mu_{20}^{\frac{s}{2}}} \quad \mu_{rs} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^r(X - \bar{X})^s, \quad \text{and} \quad \gamma = \frac{1}{n}. \tag{2.6}$$

The classical ratio type estimator for the population variance S_y^2, when the population variance S_x^2 of an auxiliary variable X is known is proposed by Isaki (1983) [20], and is given by:

$$\hat{S}_R^2 = s_y^2 \frac{S_x^2}{s_x^2}, \tag{2.7}$$

where s_y^2 is the sample variance of the study variable and s_x^2 is the sample mean of the auxiliary variable. The bias and mean square error of this estimator, up to the first order of approximation, are given by:

$$Bias(\hat{S}_R^2) \approx \frac{1}{n} S_y^2 \left[(\beta_2(x) - 1) - (\lambda_{22} - 1) \right], \quad \text{and} \tag{2.8}$$

$$MSE(\hat{S}_R^2) \approx \frac{1}{n} S_y^4 \Big[(\beta_2(y) - 1) + (\beta_2(x) - 1) - 2\left(\lambda_{22} - 1\right) \Big], \tag{2.9}$$

respectively, where

$$\beta_2(y) = \frac{\mu_{04}}{\mu_{02}^2}, \qquad \beta_2(x) = \frac{\mu_{40}}{\mu_{20}^2}, \quad \text{and} \quad \lambda_{22} = \frac{\mu_{22}}{\mu_{02}\mu_{20}}. \tag{2.10}$$

Singh et al.(2011) [59] proposed an exponential ratio type estimator for the population variance which is given by:

$$\hat{S}_{EXP}^2 = s_y^2 \exp\left(\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2} \right). \tag{2.11}$$

Its bias and mean square error respectively, up to first order of approximation, are given by:

$$Bias(\hat{S}^2_{EXP}) \approx \frac{1}{n} S^2_y \left[\frac{3}{8} \left(\beta_2(x) - 1 \right) - \frac{1}{2} \left(\lambda_{22} - 1 \right) \right], \quad \text{and} \tag{2.12}$$

$$MSE(\hat{S}^2_{EXP}) \approx \frac{1}{n} S^4_y \left[(\beta_2(y) - 1) + \left(\frac{\beta_2(x) - 1}{4} \right) - (\lambda_{22} - 1) \right]. \tag{2.13}$$

The following table gives various modified ratio estimators of the population variance using known population parameters of an auxiliary variable. For the ease of presentation, the following notations are used:

$$\boldsymbol{\beta} = \beta_{2(x)} - \mathit{1} \quad \text{and} \quad \boldsymbol{\lambda} = \lambda_{22} - \mathit{1}. \tag{2.14}$$

Table 2. The Modified Ratio Estimators of Population Variance With the Corresponding Characterising Constant, Bias and the Mean Square Error

Estimator	**Constant R_i**	**Bias**	**Mean square error**
$\hat{S}^2_1 = s^2_y \left[\frac{S^2_x + C_x}{s^2_x + C_x} \right]$	$R_1 = \frac{S^2_x}{S^2_x + C_x}$	$\frac{1}{n} S^2_y R_1 \left[R_1 \boldsymbol{\beta} - \boldsymbol{\lambda} \right]$	$\frac{1}{n} S^4_y \left[\boldsymbol{\beta} + R^2_1 \boldsymbol{\beta} - 2R_1 \boldsymbol{\lambda} \right]$
$\hat{S}^2_2 = s^2_y \left[\frac{S^2_x + \beta_{2(x)}}{s^2_x + \beta_{2(x)}} \right]$	$R_2 = \frac{S^2_x}{S^2_x + \beta_{2(x)}}$	$\frac{1}{n} S^2_y R_2 \left[R_2 \boldsymbol{\beta} - \boldsymbol{\lambda} \right]$	$\frac{1}{n} S^4_y \left[\boldsymbol{\beta} + R^2_2 \boldsymbol{\beta} - 2R_2 \boldsymbol{\lambda} \right]$

Continued on next page

Estimator	Constant R_i	Bias	Mean square error
$\hat{S}_3^2 = s_y^2\left[\frac{S_x^2\beta_{2(x)}+C_x}{s_x^2\beta_{2(x)}+C_x}\right]$	$R_3 = \frac{S_x^2\beta_{2(x)}}{S_x^2\beta_{2(x)}+C_x}$	$\frac{1}{n}S_y^2R_3\left[R_3\boldsymbol{\beta} - \boldsymbol{\lambda}\right]$	$\frac{1}{n}S_y^4\left[\boldsymbol{\beta} + R_3^2\boldsymbol{\beta} - 2R_3\boldsymbol{\lambda}\right]$
$\hat{S}_4^2 = s_y^2\left[\frac{S_x^2C_x+\beta_{2(x)}}{s_x^2C_x+\beta_{2(x)}}\right]$	$R_4 = \frac{S_x^2C_x}{S_x^2C_x+\beta_{2(x)}}$	$\frac{1}{n}S_y^2R_4\left[R_4\boldsymbol{\beta} - \boldsymbol{\lambda}\right]$	$\frac{1}{n}S_y^4\left[\boldsymbol{\beta} + R_4^2\boldsymbol{\beta} - 2R_4\boldsymbol{\lambda}\right]$
$\hat{S}_5^2 = s_y^2\left[\frac{S_x^2+Q_a}{s_x^2+Q_a}\right]$	$R_5 = \frac{S_x^2}{S_x^2+Q_a}$	$\frac{1}{n}S_y^2R_5\left[R_5\boldsymbol{\beta} - \boldsymbol{\lambda}\right]$	$\frac{1}{n}S_y^4\left[\boldsymbol{\beta} + R_5^2\boldsymbol{\beta} - 2R_5\boldsymbol{\lambda}\right]$
$\hat{S}_6^2 = s_y^2\left[\frac{S_x^2C_x+M_d}{s_x^2C_x+M_d}\right]$	$R_6 = \frac{S_x^2}{S_x^2+M_d}$	$\frac{1}{n}S_y^2R_6\left[R_6\boldsymbol{\beta} - \boldsymbol{\lambda}\right]$	$\frac{1}{n}S_y^4\left[\boldsymbol{\beta} + R_6^2\boldsymbol{\beta} - 2R_6\boldsymbol{\lambda}\right]$

For convenience, the biases and mean square errors, up to first order of approximation, of the modified ratio type variance estimators $\hat{S}_i^2$ shown in Table 2 are represented in a single expressions as:

$$\text{Bias}(S_i^2) \approx \frac{1}{n}S_y^2R_i\left[R_i\left(\beta_{2(x)} - 1\right) - (\lambda_{22} - 1)\right], \quad i = 1, \dots 11 \tag{2.15}$$

$$\text{MSE}(S_i^2) \approx \frac{1}{n}S_y^4\left[\left(\beta_{2(y)} - 1\right) + R_i^2\left(\beta_{2(x)} - 1\right) - 2R_i\left(\lambda_{22} - 1\right)\right], \quad i = 1, ..11 \tag{2.16}$$

For both the mean and variance estimators, we will now check how the empirical mean square errors and the approximated theoretical mean square errors compare.

2.4 Comparisons of the Theoretical and Empirical Mean Square Errors

In this section we compare the empirical and approximate theoretical mean square errors of various ratio type mean and variance estimators by carrying out a simulation study. We calculate the ratios of two mean square errors using the expressions:

$$R(t_i) = 100 \times \frac{MSEE(t_i)}{MSET(t_i)} \quad \text{and} \quad R(\hat{S}_i^2) = 100 \times \frac{MSEE(\hat{S}_i^2)}{MSET(\hat{S}_i^2)} \tag{2.17}$$

where $MSEE$ is the empirical mean square error and $MSET$ is the corresponding theoretical mean square error, correct to first order of approximation. In order to study the effect of departure from bivariate normal assumption on these comparisons, we consider three distributions - bivariate normal, bivariate Poisson and bivariate gamma for (X, Y) with parameters as given in Tables 3 and 4. We generated bivariate normal distributions with mean as $\mu = [4 \quad 6]$ and the standard deviation as $\sigma = [2 \quad 3]$. We also used three correlation levels between X and Y as $\rho_{yx} = 0.8;\ 0.2;\ 0.5$. For the purpose of simulation, we generated 10,000 values from each distribution and used that as our finite population. In doing so, our means, standard deviations and the coefficient of correlation shift a little bit from the original distribution values. The same approach was used for generating bivariate Poisson and gamma distributions. Tables 5 and 6 give the ratios between the empirical and approximated theoretical mean square errors for the ratio and product type mean estimators, respectively. Table 7 does the same for the variance estimators. The population size used is $N = 5000$ with sample size $n = 100, 200$ and 500. The results are averaged over 10,000 trials. These distributions were generated using the software package R and the code is given in Appendix A. Note that the ratios for all estimators are greater than one hundred, indicating that the first order approximations underestimate the true mean square errors. Also, the first order approximation works better when the sampling fraction is smaller. For the variance estimators, almost all ratios are close to 100, indicating that the first order approximations for the theoretical mean square errors for variance estimators are generally good.

Table 3. Population Statistics for Various Bivariate Distributions with Positive Coefficient of Correlations

	Normal distribution			**Poisson distribution**			**Gamma distribution**		
	Pop.1	**Pop.2**	**Pop.3**	**Pop.1**	**Pop.2**	**Pop.3**	**Pop.1**	**Pop.2**	**Pop.3**
$\boldsymbol{\rho_{yx}}$	0.7929	0.1742	0.5077	0.7779	0.2027	0.4931	0.7931	0.1965	0.4905
$\boldsymbol{\bar{Y}}$	5.9468	6.0264	5.9642	5.9960	5.9538	6.0286	5.9882	5.9554	5.9699
$\boldsymbol{\bar{X}}$	3.9783	4.0038	4.0078	3.9906	4.0156	4.0072	4.0268	4.0403	4.0350
$\boldsymbol{S_y^2}$	8.5642	3.8957	8.9001	5.9960	6.0720	6.2066	8.8218	8.7676	8.7965
$\boldsymbol{S_x^2}$	3.9425	8.7966	3.9877	4.0581	4.0877	4.0827	3.9351	3.9307	3.9339
$\boldsymbol{C_y}$	0.4921	0.4921	0.5002	0.4083	0.4138	0.4132	0.4959	0.4971	0.4968
$\boldsymbol{C_x}$	0.4991	0.4929	0.4982	0.5048	0.5034	0.5042	0.4926	0.4907	0.4915
$\boldsymbol{\beta_{1(x)}}$	0.0417	0.0437	0.0514	0.5110	0.5046	0.5508	0.9461	0.9445	0.4915
$\boldsymbol{\beta_{1(y)}}$	-0.0028	0.0153	-0.0141	0.4334	0.4007	0.4231	0.9998	1.0532	1.0337
$\boldsymbol{\beta_{2(x)}}$	2.9266	2.9253	3.0240	3.2403	3.2322	3.3187	4.1891	4.1458	4.1813
$\boldsymbol{\beta_{2(y)}}$	3.1279	3.0059	2.9742	3.2559	3.1523	3.0777	4.6017	4.9625	4.8376

Table 4. Population Statistics for Various Bivariate Distributions with Negative Coefficients of Correlation

	Normal distribution			**Poisson distribution**			**Gamma distribution**		
	Pop.1	**Pop.2**	**Pop.3**	**Pop.1**	**Pop.2**	**Pop.3**	**Pop.1**	**Pop.2**	**Pop.3**
ρ_{yx}	-0.7960	-0.2168	-0.4854	-0.7654	-0.2045	-0.4714	-0.7245	-0.1823	-0.4572
$\bar{Y}$	6.0463	5.9800	5.9969	5.9998	5.9626	6.0132	5.9121	5.9382	5.9257
$\bar{X}$	3.9785	3.9914	3.9938	4.0258	3.967	3.9696	4.0498	4.0454	4.0481
S_y^2	8.8926	8.9801	8.8556	6.0694	6.0080	5.8517	8.8091	8.7193	8.6725
S_x^2	4.0426	4.1489	4.0230	4.0699	4.0243	4.0234	3.9177	3.9248	3.9202
C_y	0.4932	0.5011	0.4962	0.4106	0.4111	0.4022	0.4962	0.4972	0.4969
C_x	0.5053	0.5103	0.5022	0.5011	0.5056	0.5053	0.4962	0.4897	0.4891
$\beta_{1(x)}$	-0.0150	-0.0221	0.0155	0.4817	0.5429	0.5283	0.9343	0.9385	0.9342
$\beta_{1(y)}$	0.0299	0.0032	-0.0028	0.4791	0.5019	0.3383	1.0226	1.0594	1.0487
$\beta_{2(x)}$	2.9358	3.0516	3.0137	3.2173	3.2514	3.4321	4.0620	4.0732	4.0351
$\beta_{2(y)}$	2.9369	3.0078	2.9492	3.3854	3.3816	3.0074	4.6990	4.9880	4.9007

Table 5. The Ratio of the Empirical Mean Square Errors and the Theoretical Mean Square Errors for Some Ratio Estimators of Population Mean

Estimator	**Distribution**	n	$\rho_{yx} = 0.8$	$\rho_{yx} = 0.2$	$\rho_{yx} = 0.5$
t_R	normal	100	100.2321	103.2854	103.8200
		200	105.0483	104.3566	105.8758
		500	112.0853	113.84	110.0194
	poisson	100	102.4274	103.6298	101.1414
		200	104.5659	105.8989	106.7388
		500	110.9086	111.1239	111.2031

Continued on next page

Estimator	Distribution	n	$\rho_{yx} = 0.8$	$\rho_{yx} = 0.2$	$\rho_{yx} = 0.5$
	gamma	100	103.4866	103.5309	103.8604
		200	105.7128	106.5458	105.6162
		500	111.2953	109.2713	108.9397
t_{ER}	normal	100	99.38368	102.4767	102.8346
		200	104.3322	103.0252	106.6609
		500	109.5924	114.084	112.0102
	poisson	100	103.9735	102.7874	100.6577
		200	104.9174	104.7845	106.7614
		500	110.6656	112.2963	109.7985
	gamma	100	102.4864	102.9836	103.0244
		200	105.5767	105.387	105.7716
		500	110.4721	109.686	109.0082
t_1	normal	100	99.91239	102.9766	103.5206
		200	104.9526	103.9713	106.004
		500	111.6351	113.9252	112.0102
	poisson	100	102.2853	103.2257	100.7495
		200	104.3092	105.2909	106.9095
		500	110.547	111.3049	110.8211
	gamma	100	103.3101	103.3526	103.6305
		200	105.6329	106.3251	105.6657
		500	111.9976	109.2644	108.7708
t_8	normal	100	99.79767	103.1607	103.516

Continued on next page

Estimator	Distribution	n	$\rho_{yx} = 0.8$	$\rho_{yx} = 0.2$	$\rho_{yx} = 0.5$
		200	104.9001	104.2067	106.0066
		500	111.3602	113.8724	110.4547
	poisson	100	102.3332	103.4441	100.7559
		200	104.8723	105.7712	106.9063
		500	110.4137	111.2337	110.8286
	gamma	100	103.2178	103.4494	103.6309
		200	105.6183	106.4524	105.6656
		500	111.0081	109.2629	108.7711
t_{10}	normal	100	99.49451	102.4014	102.8629
		200	104.5105	103.0878	106.524
		500	109.936	114.07	111.7853
	poisson	100	103.6949	102.5935	100.4773
		200	104.3297	104.7673	106.8626
		500	110.4786	112.0259	109.8373
	gamma	100	102.4247	102.8487	102.9197
		200	105.6019	105.3316	105.7317
		500	110.4758	109.6692	109.0421
t_{12}	normal	100	100.2444	103.2739	103.8076
		200	105.0515	104.3431	105.8796
		500	112.0971	113.8429	110.0345
	poisson	100	102.3442	103.4837	100.9874
		200	105.2978	105.2216	106.7997

Continued on next page

Estimator	**Distribution**	n	$\rho_{yx} = 0.8$	$\rho_{yx} = 0.2$	$\rho_{yx} = 0.5$
		500	110.7798	111.2238	111.0676
	gamma	100	103.3994	103.4378	103.7428
		200	105.6642	106.4382	105.6381
		500	111.2953	109.2624	108.8474
t_{13}	normal	100	99.57341	102.5288	103.0421
		200	104.6762	103.3053	106.3713
		500	110.4195	114.0579	111.4054
	poisson	100	103.0156	102.7045	100.4017
		200	104.6184	104.0080	105.0635
		500	110.2619	111.7367	110.1086
	gamma	100	102.8818	103.0759	103.2414
		200	105.6227	105.8209	105.7647
		500	110.7136	109.4011	108.7263
t_{14}	normal	100	99.43659	102.3166	102.7426
		200	104.3603	102.9423	106.6093
		500	109.5952	114.0473	112.008
	poisson	100	103.9902	102.5738	100.5521
		200	105.2389	104.5556	106.7036
		500	110.6473	112.2362	109.765
	gamma	100	102.5017	102.8907	102.9737
		200	105.6104	105.418	105.7475
		500	110.5067	109.6157	108.9717

Continued on next page

Estimator	Distribution	n	$\rho_{yx} = 0.8$	$\rho_{yx} = 0.2$	$\rho_{yx} = 0.5$
t_{15}	normal	100	100.1001	103.1645	103.7071
		200	105.0131	104.2113	105.9148
		500	111.9379	113.8713	110.1646
	poisson	100	102.345	103.4841	100.9994
		200	104.8734	105.0988	106.7947
		500	110.7813	111.2237	111.0787
	gamma	100	103.4393	103.4799	103.7964
		200	105.6844	106.4887	105.6272
		500	111.2492	109.2652	108.8884
t_{16}	normal	100	99.62308	103.0511	103.3062
		200	104.7514	104.0687	106.1474
		500	110.6907	113.9032	110.8416
	poisson	100	102.7141	103.2923	100.5444
		200	104.5655	105.2231	107.0185
		500	110.2492	111.2798	110.5362
	gamma	100	102.9826	103.3795	103.4657
		200	105.6195	106.3627	105.7135
		500	110.7885	109.2624	108.6981
t_{17}	normal	100	99.31717	102.184	102.4745
		200	104.0298	102.763	106.6999
		500	109.0659	113.8599	112.3249
	poisson	100	104.6171	102.6304	100.9547

Continued on next page

Estimator	Distribution	n	$\rho_{yx} = 0.8$	$\rho_{yx} = 0.2$	$\rho_{yx} = 0.5$
		200	105.6679	104.7766	105.7763
		500	111.2693	112.8358	109.7016
	gamma	100	102.0796	102.5983	102.6325
		200	105.5362	104.8809	105.5878
		500	110.372	109.9758	109.4835
t_{18}	normal	100	99.36481	102.3168	102.4113
		200	103.4991	102.9427	106.6942
		500	108.7809	114.0474	112.3512
	poisson	100	102.7024	103.2929	100.5716
		200	104.3997	105.8712	105.0036
		500	110.2501	111.2796	110.5812
	gamma	100	103.2043	103.445	103.6204
		200	105.6174	106.4471	105.6685
		500	110.9938	109.2627	108.7648
t_{19}	normal	100	99.52536	102.1462	102.749
		200	104.5819	102.7529	106.6053
		500	110.1281	113.7218	111.997
	poisson	100	103.3636	102.6733	100.5634
		200	104.2235	104.9812	106.6796
		500	110.3442	113.0128	109.7573
	gamma	100	102.7525	102.4615	102.9356
		200	105.6237	104.6741	105.7368

Continued on next page

Estimator	Distribution	n	$\rho_{yx} = 0.8$	$\rho_{yx} = 0.2$	$\rho_{yx} = 0.5$
		500	110.6315	110.1277	109.0208
t_{20}	normal	100	99.39377	102.1017	102.4807
		200	104.2428	102.817	106.6998
		500	109.3767	113.4286	112.3211
	poisson	100	100.287	102.8040	100.9687
		200	105.3788	104.8766	105.7411
		500	110.8782	113.4029	109.7043
	gamma	100	102.3522	102.1922	102.6315
		200	105.5919	104.3254	105.5872
		500	110.4494	110.396	109.4851
t_{21}	normal	100	99.57239	102.5207	103.0416
		200	104.6744	103.2917	106.3718
		500	110.4135	110.0596	111.4067
	poisson	100	102.9972	102.7115	100.4024
		200	104.8978	104.9876	107.0661
		500	110.2593	111.7243	110.1212
	gamma	100	102.9103	103.092	103.2612
		200	105.6219	105.8558	105.762
		500	110.7338	109.386	108.7168

The Table 6 is given below:

Table 6. The Ratio of the Empirical Mean Square Errors and the Theoretical Mean Square Errors for Some Product Estimators of Population Mean

Estimator	Distribution	n	$\rho_{yx} = -0.8$	$\rho_{yx} = -0.2$	$\rho_{yx} = -0.5$
t_P	normal	100	100.7915	102.4633	102.924
		200	104.3677	103.1064	105.2511
		500	109.6164	110.7636	111.7535
	poisson	500	102.5464	100.7025	100.4814
		200	104.6183	105.0594	103.6489
		500	111.0469	108.6186	112.588
	gamma	100	100.4532	100.4111	102.8959
		200	104.186	103.6635	104.9024
		500	110.5243	109.9268	108.8254
t_{EP}	normal	100	102.6857	102.6083	102.1444
		200	102.8142	102.7849	103.6092
		500	109.9355	111.3013	112.0234
	poisson	100	102.239	99.46373	101.3195
		200	103.5144	104.2884	102.1749
		500	108.0874	109.676	110.333
	gamma	100	99.84798	99.98673	102.9191
		200	103.1241	104.4126	104.2602
		500	112.3393	110.814	109.1679
t_2	normal	100	102.2507	102.4889	102.7886

Continued on next page

Estimator	Distribution	n	$\rho_{yx}=-0.8$	$\rho_{yx}=-0.2$	$\rho_{yx}=-0.5$
		200	103.6927	102.9077	104.9994
		500	109.5666	110.8968	111.824
	poisson	100	102.6845	100.4653	100.5029
		200	104.3465	104.9586	103.3465
		500	110.3771	108.7617	112.1137
	gamma	100	100.1934	100.2999	102.8969
		200	103.6414	103.7711	104.7753
		500	110.9058	110.0245	108.6992
t_3	normal	100	102.6067	102.4167	102.8792
		200	104.1152	103.0319	105.1696
		500	109.5909	110.8102	111.7785
	poisson	100	102.5997	100.6269	100.4818
		200	104.5343	105.0295	103.5587
		500	110.8402	108.6615	112.4456
	gamma	100	100.3809	100.3809	102.8958
		200	104.0563	103.6894	104.8699
		500	110.6187	109.9488	108.7841
t_4	normal	100	102.4197	102.6283	101.959
		200	103.0729	102.9747	104.2773
		500	110.0457	111.3666	112.0318
	poisson	100	101.7583	99.24205	101.8122
		200	103.4367	103.9872	104.89

Continued on next page

Estimator	**Distribution**	n	$\rho_{yx} = -0.8$	$\rho_{yx} = -0.2$	$\rho_{yx} = -0.5$
		500	108.0828	110.0665	110.0189
	gamma	100	100.0999	100.0955	102.9719
		200	104.4244	104.8327	104.0842
		500	112.6744	111.3918	109.9204
t_5	normal	100	102.9923	102.6381	102.4300
		200	105.8345	104.6867	104.2223
		500	109.6884	111.159	111.956
	poisson	100	102.606	99.91742	100.7893
		200	103.7097	104.6424	102.634
		500	108.7927	109.1969	111.0331
	gamma	100	99.86107	100.1111	102.9124
		200	102.179	104.0898	104.4759
		500	111.8073	110.3774	108.7577
t_6	normal	100	102.5952	102.1691	101.6073
		200	104.3076	105.0189	103.2927
		500	110.4568	111.1449	111.9344
	poisson	100	102.1137	99.54014	101.2400
		200	103.434	104.3345	102.1657
		500	108.0895	109.6098	110.3643
	gamma	100	99.85309	100.1022	102.9142
		200	102.0757	104.1151	104.4547
		500	111.8658	110.4081	108.7816

Continued on next page

Estimator	Distribution	n	$\rho_{yx} = -0.8$	$\rho_{yx} = -0.2$	$\rho_{yx} = -0.5$
t_7	normal	100	100.0679	102.5509	102.745
		200	103.5048	102.8624	104.9155
		500	109.5654	110.9326	111.8439
	poisson	100	102.7039	100.4166	100.511
		200	104.2795	104.9354	103.3014
		500	110.2112	108.7943	112.0436
	gamma	100	100.194	100.3005	102.8969
		200	103.6431	103.7703	104.7749
		500	110.9047	110.0238	108.699
t_9	normal	100	100.479	100.4261	103.0476
		200	105.6751	103.2183	105.4638
		500	100.2139	110.6987	111.6723
	poisson	100	102.1747	100.8053	100.5172
		200	104.978	105.0968	103.9425
		500	111.9261	108.5645	113.0575
	gamma	100	100.9835	100.4608	102.8984
		200	104.7845	103.625	105.0288
		500	110.0227	109.8964	109.0511
t_{11}	normal	100	102.7542	102.6559	102.2548
		200	104.7896	104.7075	103.9014
		500	109.8039	111.2546	111.9976
	poisson	100	102.3278	99.61978	101.1517

Continued on next page

Estimator	Distribution	n	$\rho_{yx} = -0.8$	$\rho_{yx} = -0.2$	$\rho_{yx} = -0.5$
		200	103.4949	104.4081	102.2342
		500	108.2487	109.5119	110.4586
	gamma	100	99.88689	100.0447	102.9379
		200	101.1153	104.4351	104.251
		500	112.373	110.8222	109.186

Table 7. The Ratio of the Empirical Mean Square Errors and the Theoretical Mean Square Errors for Some Ratio Estimators of Population Variance

Estimator	Distribution	n	$\rho_{yx} = 0.8$	$\rho_{yx} = 0.2$	$\rho_{yx} = 0.5$
$\hat{S}^2_R$	normal	100	102.6034	100.1753	100.2699
		200	100.0565	101.8364	103.8158
		500	101.3344	102.3967	101.6947
	poisson	100	100.8989	100.32592	100.3245
		200	102.8938	100.6621	99.9786
		500	101.7800	101.2609	100.3021
	gamma	100	100.6676	101.1580	100.5213
		200	100.7265	102.1337	102.2242
		500	101.9045	100.9870	100.6419
$\hat{S}^2_{EXP}$	normal	100	99.14575	99.1708	101.5117
		200	99.86147	99.66083	99.85453
		500	102.5893	102.1652	100.7433
	poisson	100	101.7435	99.7546	99.4442
		200	100.6136	100.7263	99.9876

Continued on next page

Estimator	Distribution	n	$\rho_{yx} = 0.8$	$\rho_{yx} = 0.2$	$\rho_{yx} = 0.5$
		500	100.2186	100.3465	100.1730
	gamma	100	101.8146	100.5106	99.8945
		200	102.2507	100.9564	101.0135
		500	100.5558	101.9936	100.8295
$\hat{S}_1^2$	normal	100	102.2496	100.9403	103.8919
		200	99.45776	101.5443	102.4047
		500	102.3915	102.1696	101.2589
	poisson	100	100.93423	101.3237	99.89328
		200	102.0616	100.1365	99.3841
		500	101.3153	99.8032	100.0044
	gamma	100	100.6576	100.6597	102.1390
		200	100.9911	101.7138	101.8544
		500	102.4915	101.9453	102.3454
$\hat{S}_2^2$	normal	100	98.75806	102.4983	100.7022
		200	99.46771	102.3254	99.73651
		500	102.5981	101.9765	100.5811
	poisson	100	100.5542	99.6985	99.2566
		200	100.5009	99.5748	101.5866
		500	100.2759	100.0909	100.3321
	gamma	100	101.6496	100.2611	99.7620
		200	101.1869	100.7350	100.8760
		500	100.3941	101.9003	100.1419

Continued on next page

Estimator	Distribution	n	$\rho_{yx} = 0.8$	$\rho_{yx} = 0.2$	$\rho_{yx} = 0.5$
$\hat{S}_3^2$	normal	100	103.6515	101.2764	105.2261
		200	99.7874	102.3254	103.2793
		500	102.3452	102.3012	101.5259
	poisson	100	100.9856	100.5016	100.1651
		200	102.5866	100.2570	99.3830
		500	101.6146	99.8975	100.1987
	gamma	100	100.6440	101.0134	100.4092
		200	100.7819	102.0199	102.1244
		500	102.8929	100.9120	101.4923
$\hat{S}_4^2$	normal	100	98.15762	101.8818	100.2661
		200	99.92439	98.8897	98.9731
		500	102.4372	102.0219	100.4837
	poisson	100	100.5581	100.8977	99.8810
		200	99.88426	99.9987	99.6844
		500	99.86211	100.2740	100.3467
	gamma	100	100.02922	100.5105	99.8400
		200	102.3324	100.5051	100.5786
		500	100.1122	102.4766	100.4888
$\hat{S}_5^2$	normal	100	98.326	101.9862	100.5991
		200	99.2388	98.93484	99.526
		500	102.4762	101.8357	100.6721
	poisson	100	100.9719	99.6923	99.3741

Continued on next page

Estimator	**Distribution**	n	$\rho_{yx} = 0.8$	$\rho_{yx} = 0.2$	$\rho_{yx} = 0.5$
		200	100.3009	100.8454	100.5001
		500	99.86105	101.9002	99.5100
	gamma	100	101.3022	100.1843	99.8754
		200	101.9050	100.9605	101.8705
		500	100.9568	101.5943	100.9341
$\hat{S}_6^2$	normal	100	109.2415	91.42963	93.9322
		200	111.478	88.76051	92.56915
		500	113.919	91.64913	94.16197
	poisson	100	111.0124	89.1428	90.0183
		200	110.2576	88.9592	93.4996
		500	112.6517	89.8431	92.9386
	gamma	100	112.0999	90.1924	93.8665
		200	111.4981	90.2828	94.5559
		500	100.8711	91.9146	92.5532

2.5 Robustness Against Departure From the Bivariate Normal Assumption

We have considered bivariate normal distributions with the three positive and three negative coefficients of correlation ρ_{yx} whose statistics are given in Tables 3 and 4. Also, we have considered the bivariate Poisson and bivariate Gamma distributions for similar coefficients of correlation ρ_{yx}. It was shown that the departure from the bivariate normal assumption does not produce any serious issue if the sample size is large. Tables 5 and 6 are for the ratio and product type mean estimators respectively, and Table 7 is for the ratio type variance estimators. We can see that corresponding ratios for the six bivariate distributions are very similar. Thus we can say that these estimators are robust with respect to the assumption of bivariate normality.

CHAPTER III

THE NEW GENERALIZED MIXTURE ESTIMATORS OF THE MEAN

3.1 Introduction

In this chapter we propose new generalized mixture estimators of the population mean of the study variable by utilizing an auxiliary variable. The aim is to get a more efficient estimator than the existing mixture estimators. As was mentioned before, many modification have been done on the ratio, product, and regression estimators of the population mean of the study variable Y using an auxiliary variable X to improve efficiency of these estimators. Bahl & Tuteja [2] introduced the exponential ratio and product type estimators which we will also use in the proposed estimators. Combining modified ratio type estimators and the exponential ratio type estimator, Singh et al. (2009)[65] suggested a generalized ratio type estimator. The special cases of this estimator are exponential ratio type, exponential product type and also Bedi (1996)[3] transformed estimators. Also, many authors have suggested several transformed ratio-type estimators for estimating the finite population mean by utilizing auxiliary information. Khoshnevisan et al. (2007)[29] proposed a general class of estimators that includes several modified ratio type estimators. Shabbir and Gupta (2010)[52] proposed a regression ratio type exponential estimator by combining Rao's (1991)[42] and Bedi's (1996)[3] estimators. Following these works, Grover & Kaur (2011)[12] introduced a regression exponential type estimator. Subramani (2013)[77] proposed a generalized modified ratio estimator for estimation of finite population mean. The ordinary ratio estimator, the linear regression estimator and the existing modified ratio estimators are special cases of that estimator. Also, more recently Grover & Kaur (2014)[13] proposed a generalized class of ratio type exponential estimators by combining Rao's (1991)[42] and Singh's et al. (2009)[65] generalized ratio type exponential estimator. Our approach is as follows.

Let $U = \{U_1, \ldots\ldots U_N\}$ be a finite population of size N and let (y_i, x_i) be the value of the study variable Y and the auxiliary variable X on ith unit U_i, $i = 1, \ldots N$. Let $\bar{Y}$ and $\bar{X}$ be population means of the study variable Y and the auxiliary variable X respectively. We assume that the population mean $\bar{X}$ and the population variance S_x^2 of the auxiliary variable are known. Let S_y^2 be the population variance of the study variable Y. Let the correlation coefficient between the study variable and the auxiliary variable be ρ_{yx}. Also, let $C_y = \frac{S_y}{\bar{Y}}$ and $C_x = \frac{S_x}{\bar{X}}$ be the coefficients of variation of the study variable Y and the auxiliary variable X, and $C_{yx} = \frac{S_{yx}}{\bar{Y}\bar{X}}$ be the coefficient of covariance between Y and X with $S_{yx} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})(x_i - \bar{X})$. To obtain the asymptotic properties of the estimators, we define the following error terms, as in Sukathme and Sukathme (1970)[78]:

$$e_y = \frac{\bar{y} - \bar{Y}}{\bar{Y}} \qquad \text{and} \qquad e_x = \frac{\bar{x} - \bar{X}}{\bar{X}}, \tag{3.1}$$

for which the following holds true:

$$\begin{gathered} E(e_y) = E(e_x) = 0, \qquad E(e_y^2) = \lambda C_y^2, \qquad E(e_x^2) = \lambda C_x^2, \\ E(e_y e_x) = \lambda C_{yx} = \lambda \rho_{yx} C_y C_x, \quad \text{where} \quad \lambda = \frac{1-f}{n} \quad \text{and} \quad f = \frac{n}{N}. \end{gathered} \tag{3.2}$$

We will use these expressions to derive the bias and the mean square error of the proposed estimators, up to the first order of approximation.

3.2 Some Existing Generalized Estimators

There are many generalized estimators of the population mean of the study variable utilizing an auxiliary variable. We discuss a few of these generalized estimators first.

3.2.1 Rao (1991) Regression Estimator

Rao (1991)[42] introduced the generalized regression type estimator to improve efficiency of the ordinary regression estimator. The estimator is given by:

$$\hat{\mu}_{R,Reg} = k_1\bar{y} + k_2\left(\bar{X} - \bar{x}\right), \tag{3.3}$$

where k_1 and k_2 are suitably chosen constants. The minimum mean square error of this estimator, up to the first order of approximation, with optimum values of k_1 and k_2 i.e.,

$$k_{1(opt)} = \frac{1}{1+\lambda\left(1-\rho_{yx}^2\right)C_y^2}, \tag{3.4}$$

$$k_{2(opt)} = k_{1(opt)}\frac{\bar{Y}}{\bar{X}}\frac{\rho_{yx}C_y}{C_x}, \tag{3.5}$$

is given by:

$$MSE_{min}(\hat{\mu}_{R,Reg}) \approx \bar{Y}^2\left[1 - \frac{1}{1+\lambda\left(1-\rho_{yx}^2\right)C_y^2}\right]. \tag{3.6}$$

3.2.2 Singh et al. (2008) Estimator

Following Bahl & Tuteja (1991)[2], Singh et al. (2008)[64] proposed a ratio product type exponential estimator given by:

$$\hat{\mu}_S = \bar{y}\left[\alpha \exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right) + (1-\alpha)\exp\left(\frac{\bar{x}-\bar{X}}{\bar{X}+\bar{x}}\right)\right], \tag{3.7}$$

where α is suitably chosen constant. The minimum mean square error, up to the first order of approximation, at optimum value of α, i.e.,

$$\alpha_{(opt)} = \frac{1}{2} + \frac{\rho_{yx}C_y}{C_x}, \tag{3.8}$$

is given by:

$$MSE_{min}(\hat{\mu}_S) \approx \lambda \bar{Y}^2\left(1-\rho_{yx}^2\right)C_y^2 = MSE(\hat{\mu}_{Reg}). \tag{3.9}$$

3.2.3 Grover & Kaur (2011) Estimator

Following Rao (1991) [42] and Bahl & Tuteja [2], Grover & Kaur (2011) [12] suggested a regression exponential type estimator given by:

$$\hat{\mu}_{GK} = \left[l_1\bar{y} + l_2\left(\bar{X}-\bar{x}\right)\right]\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right), \tag{3.10}$$

where l_1 and l_2 are suitably chosen constants.

The minimum mean square error of this estimator, at the optimum values of l_1 and l_2, i.e.,

$$l_{1(opt)} = \frac{-1 + \frac{1}{8}\lambda C_x^2}{\left[-1 + \lambda(1 - \rho_{yx}^2)C_y^2\right]}, \tag{3.11}$$

$$l_{2(opt)} = \frac{\bar{Y}}{\bar{X}}\left[\frac{1}{2} - l_{1(opt)}\left(1 - \rho_{yx}\frac{C_y}{C_x}\right)\right], \tag{3.12}$$

is given by:

$$MSE_{min}(\hat{\mu}_{GK}) \approx \frac{\lambda \bar{Y}^2\left[\lambda C_x^4 - 16\left(\rho_{yx}^2 - 1\right)\left(-4 + \lambda C_x^2\right)C_y^2\right]}{64\left[-1 + \lambda\left(\rho_{yx}^2 - 1\right)C_y^2\right]}. \tag{3.13}$$

It turns out that

$$MSE_{min}(\hat{\mu}_{GK}) \approx MSE(\hat{\mu}_{Reg}) - \frac{\lambda^2 \bar{Y}^2\left[C_x^2 + 8\left(1 - \rho_{yx}^2\right)C_y^2\right]^2}{64\left[1 + \lambda\left(\rho_{yx}^2 - 1\right)C_y^2\right]}. \tag{3.14}$$

We can see that Grover & Kaur (2011)[12] estimator is more efficient than the linear regression estimator $\hat{\mu}_{Reg}$ if

$$1 + \lambda\left(\rho_{yx}^2 - 1\right)C_y^2 > 0, \text{ or if } \quad \rho_{yx}^2 > 1 - \frac{1}{\lambda C_y^2}. \tag{3.15}$$

This condition is very likely to hold true since $\left(1 - \frac{1}{\lambda C y^2}\right)$ is typically small. For example, if $N = 5000, n = 200$ and $C_y = 1.5$, this expression equals -150. Hence (3.15) will hold true for all correlation values. Since the linear regression estimator is always better than the sample mean, ratio, product, and exponential estimators, we can say that Grover & Kaur (2011)[12] estimator is also always better than these estimators.

3.3 The Proposed Generalized Mixture Estimator I

In this section we propose a new generalized mixture estimator by combining the ratio, product, regression, and exponential ratio type estimators. The estimator is given as:

$$\hat{\mu}_{GM} = \left\{ d_1 \bar{y} \left[\frac{1}{2} \left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}} \right) \right]^{\alpha} + d_2 \left(\bar{X} - \bar{x} \right) \right\} \exp \left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}} \right) \tag{3.16}$$

where d_i $(i = 1, 2)$ and α are suitably chosen constants. We will consider two values of α ($\alpha = 1$ and $\alpha = 2$). Using error terms (3.1), this generalized mixture estimator can be written as:

$$\hat{\mu}_{GM} = \left[d_1 \bar{Y} \left(1 + e_y\right) \frac{1}{2^{\alpha}} \left[\left(1 + e_x\right)^{-1} + 1 + e_x \right]^{\alpha} - \bar{X} d_2 e_x \right] \exp \left[\left(-\frac{e_x}{2} \right) \left(1 + \frac{e_x}{2} \right)^{-1} \right]. \tag{3.17}$$

Using first order approximation, this can be written as:

$$\begin{aligned} \hat{\mu}_{GM} &\approx \left[d_1 \bar{Y} \left(1 + e_y\right) \frac{1}{2^{\alpha}} \left(2 + e_x^2\right)^{\alpha} - d_2 \bar{X} e_x \right] \left(1 - \frac{1}{2} e_x + \frac{3}{8} e_x^2 \right) \\ &\approx \left[d_1 \bar{Y} \left(1 + e_y\right) \left(1 + \alpha \frac{e_x^2}{2} \right) - d_2 \bar{X} e_x \right] \left(1 - \frac{1}{2} e_x + \frac{3}{8} e_x^2 \right) \\ &\approx d_1 \bar{Y} + \frac{\alpha}{2} d_1 \bar{Y} e_x^2 + d_1 \bar{Y} e_y - d_2 \bar{X} e_x - \frac{1}{2} d_1 \bar{Y} e_x - \frac{1}{2} d_1 \bar{Y} e_y e_x + \frac{1}{2} d_2 \bar{X} e_x^2 + \frac{3}{8} d_1 \bar{Y} e_x^2. \end{aligned}$$

Thus it follows

$$\hat{\mu}_{GM} - \bar{Y} \approx (d_1 - 1)\,\bar{Y} + d_1\bar{Y}\left(e_y - \frac{1}{2}e_x - \frac{1}{2}e_y e_x + Ae_x^2\right) - d_2\bar{X}\left(e_x - \frac{1}{2}e_x^2\right), \quad (3.18)$$

where

$$A = \frac{\alpha}{2} + \frac{3}{8}. \quad (3.19)$$

By taking expectation of (3.18), the bias of the proposed generalized mixture estimator, up to the first order of approximation, is given by:

$$Bias(\hat{\mu}_{GM}) \approx (d_1 - 1)\,\bar{Y} + \lambda d_1\bar{Y}\left(AC_x^2 - \frac{1}{2}\rho_{yx}C_yC_x\right) + \frac{d_2}{2}\lambda\bar{X}C_x^2. \quad (3.20)$$

By squaring equation (3.18) and keeping terms only up to the first order of approximation, we have:

$$\begin{aligned}(\hat{\mu}_{GM} - \bar{Y})^2 \approx\ & (d_1 - 1)^2 + d_1^2\bar{Y}^2\left(e_y - \frac{1}{2}e_x - \frac{1}{2}e_y e_x + Ae_x^2\right) + d_2^2\bar{X}^2\left(e_x - \frac{1}{2}e_x^2\right)^2 \\ & + 2d_1\,(d_1 - 1)\,\bar{Y}^2\left(e_y - \frac{1}{2}e_x - \frac{1}{2}e_y e_x + Ae_x^2\right) - 2d_2\,(d_1 - 1)\,\bar{X}\bar{Y}\left(e_x - \frac{1}{2}e_x^2\right) \\ & - 2d_1d_2\bar{X}\bar{Y}\left(e_y - \frac{1}{2}e_x - \frac{1}{2}e_y e_x + Ae_x^2\right)\left(e_x - \frac{1}{2}e_x^2\right) \\ & = (d_1 - 1)^2\,\bar{Y}^2 + d_1^2\bar{Y}^2\left(e_y^2 - e_y e_x + \frac{1}{4}e_x^2\right) + d_2^2\bar{X}^2e_x^2 \\ & + 2d_1^2\bar{Y}^2\left(e_y - \frac{1}{2}e_x - \frac{1}{2}e_y e_x + Ae_x^2\right) - 2d_1\bar{Y}^2\left(e_y - \frac{1}{2}e_y - \frac{1}{2}e_y e_x + Ae_x^2\right) \\ & - 2d_1d_2\bar{X}\bar{Y}\left(e_y e_x - e_x^2 + e_x\right) + d_2\bar{X}\bar{Y}\left(2e_x - e_x^2\right). \end{aligned} \quad (3.21)$$

This can be further simplified to

$$
\begin{aligned}
(\hat{\mu}_{GM}-\bar{Y})^2 \approx (d_1-1)^2 + d_1^2\bar{Y}^2\left[2e_y - e_x - 2e_ye_x + e_y^2 + \left(2A+\frac{1}{4}\right)e_x^2\right] \\
+ d_2^2\bar{X}^2e_x^2 - d_1\bar{Y}^2\left(2e_y - e_x - e_ye_x + 2Ae_x^2\right) \\
- 2d_1d_2\bar{X}\bar{Y}\left(e_x + e_ye_x - e_x^2\right) + d_2\bar{X}\bar{Y}\left(2e_x - e_x^2\right).
\end{aligned} \tag{3.22}
$$

Taking expectation of (3.22), the mean square error of the proposed estimator, up to the first order of approximation, is given as:

$$
\begin{aligned}
MSE(\hat{\mu}_{GM}) \approx (d_1-1)^2\bar{Y}^2 + \lambda d_1^2\bar{Y}^2\left[C_y^2 - 2\rho_{yx}C_yC_x + \left(2A+\frac{1}{4}\right)C_x^2\right] \\
+ \lambda d_2^2\bar{X}^2C_x^2 - \lambda d_1\bar{Y}^2\left(2AC_x^2 - \rho_{yx}C_yC_x\right) \\
- 2\lambda d_1d_2\bar{X}\bar{Y}\left(\rho_{yx}C_yC_x - C_x^2\right) - \lambda d_2\bar{X}\bar{Y}C_x^2
\end{aligned} \tag{3.23}
$$

Taking partial derivatives with respect to d_1 and d_2, we have:

$$
\begin{aligned}
\frac{\partial MSE(\hat{\mu}_{GM})}{\partial d_1} = 2\left(d_1-1\right)\bar{Y}^2 + 2\lambda d_1\bar{Y}^2\left[C_y^2 - 2\rho_{yx}C_yC_x + \left(2A+\frac{1}{4}\right)C_x^2\right] \\
- \lambda\bar{Y}^2\left(2AC_x^2 - \rho_{yx}C_yC_x\right) - 2\lambda d_2\bar{X}\bar{Y}\left(\rho_{yx}C_yC_x - C_x^2\right),
\end{aligned}
$$

$$
\frac{\partial MSE(\hat{\mu}_{GM})}{\partial d_2} = 2\lambda d_2\bar{X}^2C_x^2 - 2\lambda d_1\bar{X}\bar{Y}\left(\rho_{yx}C_yC_x - C_x^2\right) - \lambda\bar{X}\bar{Y}C_x^2. \tag{3.24}
$$

Setting these derivatives equal to zero, the optimum values of d_1 and d_2 are given as:

$$d_{1(opt)} = \frac{1+\lambda\left(A-\frac{1}{2}\right)C_x^2}{1+\lambda\left[\left(2A-\frac{3}{4}\right)C_x^2+\left(1-\rho_{yx}^2\right)C_y^2\right]}, \quad \text{and} \tag{3.25}$$

$$d_{2(opt)} = \frac{\bar{Y}}{\bar{X}}\left[\frac{1}{2} - d_{1(opt)}\left(1-\rho_{yx}\frac{C_y}{C_x}\right)\right]. \tag{3.26}$$

Substituting the optimum value of d_2 in (3.23) we get:

$$\begin{aligned}
MSE_{min}(\hat{\mu}_{GM}) \approx{}& (d_1-1)^2\bar{Y}^2 + \lambda d_1^2\bar{Y}^2\left[\left(2A+\frac{1}{4}\right)C_x^2 - 2\rho_{yx}C_xC_y + C_y^2\right] \\
&+ \lambda\bar{X}^2C_x^2\frac{\bar{Y}^2}{\bar{X}^2}\left[\frac{1}{2}-d_{1(opt)}\left(1-\rho_{yx}\frac{C_y}{C_x}\right)\right]^2 - \lambda d_1\bar{Y}^2\left[2AC_x^2 - \rho_{yx}C_xC_y\right] \\
&- 2\lambda\bar{X}\bar{Y}\frac{\bar{Y}}{\bar{X}}d_1\left[\frac{1}{2}-d_1\left(1-\rho_{yx}\frac{C_y}{C_x}\right)\right]\left(\rho_{yx}C_xC_y - C_x^2\right) \\
&- \lambda\bar{X}\bar{Y}C_x^2\frac{\bar{Y}}{\bar{X}}\left[\frac{1}{2}-d_1\left(1-\rho_{yx}\frac{C_y}{C_x}\right)\right] \\
={}& d_1^2\bar{Y}^2 - 2d_1\bar{Y}^2 + \bar{Y}^2 + \lambda d_1^2\left[\left(2A+\frac{1}{4}\right)C_x^2 - 2\rho_{yx}C_yC_x + C_y^2\right] \\
&+ \lambda\bar{Y}^2C_x^2\left[\frac{1}{4}-d_1\left(1-\rho_{yx}\frac{C_y}{C_x}\right)+d_1^2\left(1-\rho_{yx}\frac{C_y}{C_x}\right)^2\right] \\
={}& d_1^2\bar{Y}^2\left\{1+\lambda\left[\left(2A-\frac{3}{4}\right)Cx^2+\left(1-\rho_{yx}^2\right)C_y^2\right]\right\} \\
&- 2d_1\bar{Y}^2\left[1+\lambda\left(A-\frac{1}{2}\right)C_x^2\right] + \bar{Y}^2 - \frac{1}{4}\lambda\bar{Y}^2C_x^2.
\end{aligned} \tag{3.27}$$

Substituting the optimum value of d_1 in (3.27), the minimum meann square error is given by:

$$MSE_{min}(\hat{\mu}_{GM}) \approx \bar{Y}^2 \left\{ \left(1 - \frac{1}{4}\lambda C_x^2\right) - \frac{\left[1 + \lambda\left(A - \frac{1}{2}\right) C_x^2\right]^2}{1 + \lambda\left[\left(2A - \frac{3}{4}\right) C_x^2 + \left(1 - \rho_{yx}^2\right) C_y^2\right]} \right\}.$$

For $\alpha = 1$ this generalized mixture estimator becomes

$$\hat{\mu}_{GM1} = \left\{ d_1 \bar{y} \left[\frac{1}{2}\left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}}\right)\right] + d_2\left(\bar{X} - \bar{x}\right) \right\} \exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right) \tag{3.28}$$

The optimum values of d_1 and d_2, for this estimator, are given by:

$$d_{1(opt)} = \frac{1 + \frac{3}{8}\lambda C_x^2}{1 + \lambda\left[C_x^2 + (1 - \rho_{yx}^2) C_y^2\right]}, \quad \text{and} \tag{3.29}$$

$$d_{2(opt)} = \frac{\bar{Y}}{\bar{X}}\left[\frac{1}{2} - d_{1(opt)}\left(1 - \rho_{yx}\frac{C_y}{C_x}\right)\right]. \tag{3.30}$$

The minimum mean squared error, up to the first order of approximation, is given by:

$$MSE_{min}(\hat{\mu}_{GM1}) \approx \bar{Y}^2 \left\{ \left(1 - \frac{1}{4}\lambda C_x^2\right) - \frac{\left[1 + \lambda \frac{3}{8} C_x^2\right]^2}{1 + \lambda\left[C_x^2 + \left(1 - \rho_{xy}^2\right) C_y^2\right]} \right\}. \tag{3.31}$$

When $\alpha = 2$, the generalized mixture estimator is given by:

$$\hat{\mu}_{GM2} = \left\{ d_1 \bar{y} \left[\frac{1}{2}\left(\frac{\bar{X}}{\bar{x}} + \frac{\bar{x}}{\bar{X}}\right)\right]^2 + d_2\left(\bar{X} - \bar{x}\right) \right\} \exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right). \tag{3.32}$$

The optimum values of d_1 and d_2 are given by:

$$d_{1(opt)} = \frac{1 + \frac{7}{8}\lambda C_x^2}{1 + \lambda\Big[2C_x^2 + (1 - \rho_{yx}^2)C_y^2\Big]}, \quad \text{and} \tag{3.33}$$

$$d_{2(opt)} = \frac{\bar{Y}}{\bar{X}}\left[\frac{1}{2} - d_{1(opt)}\left(1 - \rho_{yx}\frac{C_y}{C_x}\right)\right]. \tag{3.34}$$

The minimum mean square of this estimator, up to the first order of approximation, is given by:

$$MSE_{min}(\hat{\mu}_{GM2}) \approx \bar{Y}^2\left\{\left(1 - \frac{1}{4}\lambda C_x^2\right) - \frac{\left[1 + \frac{7}{8}\lambda C_x^2\right]^2}{1 + \lambda\left[2C_x^2 + \left(1 - \rho_{yx}^2\right)C_y^2\right]}\right\}. \tag{3.35}$$

Since in most cases, the correlation coefficient between the study variable Y and the auxiliary variable X is positive, we also propose a generalized mixture estimator without the product term in the next section.

3.4 The Proposed Generalized Mixture Estimator II

By combining the ratio, regression, and the exponential estimators, we propose a second generalized mixture estimator without the product term:

$$\hat{\mu}_{GMR} = \left[k_1\bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)^{\alpha} + k_2\left(\bar{X} - \bar{x}\right)\right]\exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right) \tag{3.36}$$

where k_i $(i = 1, 2)$ and α are suitably chosen constants. We will consider two values of α ($\alpha = 1$ and $\alpha = 2$).

Rewriting $\hat{\mu}_{GMR}$ in terms e_y and e_x, the estimator (3.36) can be written as:

$$\hat{\mu}_{GMR} = k_1\bar{Y}\left[(1+e_y)(1+e_x)^{-\alpha} - k_2\bar{X}e_x\right]\exp\left[\left(-\frac{e_x}{2}\right)\left(1+\frac{e_x}{2}\right)^{-1}\right] \tag{3.37}$$

Up to the first order of approximation, this estimator can be expressed as:

$$\begin{aligned}
\hat{\mu}_{GMR} &\approx \left[k_1\bar{Y}(e_y+1)\left(1-\alpha e_x+\frac{1}{2}\alpha(\alpha+1)e_x^2\right) - k_2\bar{X}e_x\right]\left(1-\frac{1}{2}e_x+\frac{3}{8}e_x^2\right)\\
&= k_1\bar{Y} - k_1\alpha\bar{Y}e_x + \frac{1}{2}k_1\bar{Y}\alpha(\alpha+1)e_x^2 + k_1\bar{Y}e_y - k_1\alpha e_y e_x - k_2\bar{X}e_x\\
&- \frac{1}{2}k_1\bar{Y}e_x + \frac{1}{2}k_1\alpha\bar{Y}e_x^2 - \frac{1}{2}k_1\bar{Y}e_y e_x + \frac{1}{2}k_2\bar{X}e_x^2 + \frac{3}{8}k_1\bar{Y}e_x^2\\
&= k_1\bar{Y} - k_1\bar{Y}\left(\alpha+\frac{1}{2}\right)e_x + k_1\bar{Y}\left[\frac{1}{2}\alpha(\alpha+1)+\frac{1}{2}\alpha+\frac{3}{8}\right]e_x^2\\
&+ k_1\bar{Y}e_y - k_1\bar{Y}\left(\alpha+\frac{1}{2}\right)e_x e_y - k_2\bar{X}\left(e_x-\frac{1}{2}e_x^2\right)\\
&= \bar{Y} + (k_1-1)\bar{Y} + k_1\bar{Y}\left\{e_y - \left(\alpha+\frac{1}{2}\right)e_x - \left(\alpha+\frac{1}{2}\right)e_x e_y\right.\\
&\left.+ \left[\frac{1}{2}\alpha(\alpha+2)+\frac{3}{8}\right]e_x^2\right\} - k_2\bar{X}\left(e_x-\frac{1}{2}e_x^2\right)
\end{aligned} \tag{3.38}$$

Using the substitutions

$$A = \alpha + \frac{1}{2} \quad \text{and} \quad B = \frac{1}{2}\alpha(\alpha+2)+\frac{3}{8}, \tag{3.39}$$

we get

$$\hat{\mu}_{GMR} - \bar{Y} = (k_1-1)\bar{Y} + k_1\bar{Y}\left(e_y - Ae_x - Ae_x e_y + Be_x^2\right) - k_2\bar{X}\left(e_x - \frac{1}{2}e_x^2\right) \tag{3.40}$$

By taking expectation on both sides of (3.40), the bias of the proposed estimator, up to the first order of approximation, is given by:

$$Bias(\hat{\mu}_{GMR}) = (k_1 - 1)\,\bar{Y} + \lambda k_1 \bar{Y}\left(BC_x^2 - A\rho_{yx}C_xC_y\right) + \frac{1}{2}k_2\lambda\bar{X}C_x^2. \tag{3.41}$$

By squaring Equation (3.40), and keeping terms only up to the the first order of approximation, we get:

$$\begin{aligned}
(\hat{\mu}_{GMR} - \bar{Y})^2 &\approx (k_1 - 1)^2\,\bar{Y}^2 + k_1^2\bar{Y}^2\left(e_y - Ae_x - Ae_xe_y + Be_x^2\right)^2 \\
&+ k_2^2\bar{X}^2\left(e_x - \frac{1}{2}e_x^2\right)^2 + 2k_1\,(k_1 - 1)\,\bar{Y}^2\left(e_y - Ae_x - Ae_xe_y + Be_x^2\right) \\
&- 2k_1k_2\bar{X}\bar{Y}\left(e_x - \frac{1}{2}e_x^2\right)\left(e_y - Ae_x - Ae_xe_y + Be_x^2\right) - 2k_2\,(k_1 - 1)\,\bar{X}\bar{Y}\left(e_x - \frac{1}{2}e_x^2\right) \\
&= (k_1 - 1)^2\,\bar{Y}^2 + k_1^2\bar{Y}^2\left(e_y^2 - 2Ae_ye_x + A^2e_x^2\right) + k_2^2\bar{X}^2e_x^2 \\
&+ 2k_1^2\bar{Y}^2\left(e_y - Ae_x - Ae_xe_y + Be_x^2\right) - 2k_1\bar{Y}^2\left(e_y - Ae_x - Ae_xe_y + Be_x^2\right) \\
&- 2k_1k_2\bar{X}\bar{Y}\left(e_xe_y - Ae_x^2\right) - 2k_1k_2\bar{X}\bar{Y}\left(e_x - \frac{1}{2}e_x^2\right) + 2k_2\bar{X}\bar{Y}\left(e_x - \frac{1}{2}e_x^2\right) \\
&= (k_1 - 1)^2\,\bar{Y}^2 + k_1^2\bar{Y}^2\left[2e_y - 2Ae_x - 4Ae_xe_y + e_y^2 + \left(A^2 + 2B\right)e_x^2\right] \\
&+ k_2^2\bar{X}^2e_x^2 - 2k_1\bar{Y}^2\left(e_y - Ae_x - Ae_ye_x + Be_x^2\right) \\
&- 2k_1k_2\bar{X}\bar{Y}\left[e_x + e_xe_y - \left(A + \frac{1}{2}\right)e_x^2\right] + k_2\bar{X}\bar{Y}\left(2e_x - e_x^2\right).
\end{aligned} \tag{3.42}$$

Using the substitution:

$$C = A^2 + 2B, \tag{3.43}$$

we get

$$
\begin{aligned}
(\hat{\mu}_{GMR} - \bar{Y})^2 \approx & (k_1 - 1)^2 \bar{Y}^2 + k_1^2 \bar{Y}^2 \left(2e_y - 2Ae_x - 4Ae_x e_y + e_y^2 + Ce_x^2\right) \\
& + k_2^2 \bar{X}^2 e_x^2 - 2k_1 \bar{Y}^2 \left(e_y - Ae_x - Ae_y e_x + Be_x^2\right) \\
& - 2k_1 k_2 \bar{X}\bar{Y} \left[e_x + e_x e_y - \left(A + \frac{1}{2}\right) e_x^2\right] + k_2 \bar{X}\bar{Y} \left(2e_x - e_x^2\right)
\end{aligned} \tag{3.44}
$$

By taking expectation on both sides, the mean square error of the proposed estimator, up to the first order of approximation, is given by:

$$
\begin{aligned}
MSE(\hat{\mu}_{GMR}) \approx & (k_1 - 1)^2 \bar{Y}^2 + \lambda k_1^2 \bar{Y}^2 \left(CC_x^2 - 4A\rho_{yx} C_x C_y + C_y^2\right) \\
& + \lambda k_2^2 \bar{X}^2 C_x^2 - \lambda k_1 \bar{Y}^2 \left(2BC_x^2 - 2A\rho_{yx} C_x C_y\right) \\
& - 2\lambda k_1 k_2 \; \bar{X}\bar{Y} \left[\rho_{xy} C_x C_y - \left(A + \frac{1}{2}\right) C_x^2\right] - \lambda k_2 \bar{X}\bar{Y} C_x^2.
\end{aligned} \tag{3.45}
$$

Partially differentiating (3.45) with respect to k_1 and k_2, we get:

$$
\begin{aligned}
\frac{\partial MSE(\hat{\mu}_{GMR})}{\partial k_1} = & 2(k_1 - 1)\bar{Y}^2 + 2\lambda k_1 \bar{Y}^2 (CC_x^2 - 4A\rho_{xy} C_x C_y + C_y^2) \\
& - \lambda \bar{Y}^2 \left(2BC_x^2 - 2A\rho_{xy} C_x C_y\right) - 2\lambda k_2 \bar{X}\bar{Y} \left[\rho_{xy} C_x C_y - \left(A + \frac{1}{2}\right) C_x^2\right]
\end{aligned} \tag{3.46}
$$

$$
\frac{\partial MSE(\hat{\mu}_{GMR})}{\partial k_2} = 2\lambda k_2 \bar{X}^2 C_x^2 - 2\lambda k_1 \bar{X}\bar{Y} \left[\rho_{xy} C_x C_y - \left(A + \frac{1}{2}\right) C_x^2\right] - \lambda \bar{X}\bar{Y} C_x^2.
$$

Setting the first derivatives equal to zero, the optimum values of k_1 and k_2 are given by:

$$k_{1(opt)} = \frac{1+\lambda\left[\left(B-\frac{1}{2}A-\frac{1}{4}\right)C_x^2+\left(\frac{1}{2}-A\right)\rho_{yx}C_xC_y\right]}{1+\lambda\left\{\left[C-\left(A+\frac{1}{2}\right)^2\right]C_x^2+(1-2A)\rho_{yx}C_xC_y+(1-\rho_{yx}^2)C_y^2\right\}}, \tag{3.47}$$

$$\text{and} \qquad k_{2(opt)} = \frac{\bar{Y}}{\bar{X}}\left\{\frac{1}{2}-k_{1(opt)}\left[\left(A+\frac{1}{2}\right)-\rho_{yx}\frac{C_y}{C_x}\right]\right\}. \tag{3.48}$$

By substituting the optimum value of k_2 in the expression (3.45), we get:

$$\begin{aligned}
MSE_{min}(\hat{\mu}_{GMR}) &\approx k_1^2\bar{Y}^2-2k_1\bar{Y}^2+\bar{Y}^2+\lambda k_1^2\bar{Y}^2\left(CC_x^2-4A\rho_{yx}C_yC_x+C_y^2\right)\\
&+\lambda\bar{X}^2C_x^2\frac{\bar{Y}^2}{\bar{X}^2}\left\{\frac{1}{2}-k_1\left[\left(A+\frac{1}{2}\right)-\rho_{yx}\frac{C_y}{C_x}\right]\right\}^2-\lambda k_1\bar{Y}^2\left(2BC_x^2-2A\rho_{yx}C_yC_x\right)\\
&-2\lambda\bar{X}\bar{Y}k_1\frac{\bar{Y}}{\bar{X}}\left\{\frac{1}{2}-k_1\left[\left(A+\frac{1}{2}\right)-\rho_{yx}\frac{C_y}{C_x}\right]\left[\rho_{yx}C_yC_x-\left(A+\frac{1}{2}\right)C_x^2\right]\right\}\\
&-\lambda\bar{X}\bar{Y}C_x^2\frac{\bar{Y}}{\bar{X}}\left\{\frac{1}{2}-k_1\left[\left(A+\frac{1}{2}\right)-\rho_{yx}\frac{C_y}{C_x}\right]\right\} \qquad (3.49)\\
&=k_1^2\bar{Y}^2\left\{1+\lambda\left[\left\{C-\left(A+\frac{1}{2}\right)^2\right\}C_x^2+(1-2A)\rho_{yx}C_yC_x+\left(1-\rho_{yx}^2\right)C_x^2\right]\right\}\\
&-2k_1\bar{Y}^2\left\{1+\lambda\left[\left(B-\frac{1}{2}A-\frac{1}{4}C_x^2\right)+\left(\frac{1}{2}-A\right)\rho_{yx}C_yC_x\right]\right\}+\bar{Y}^2-\frac{1}{4}\lambda\bar{Y}^2C_x^2
\end{aligned}$$

Then substituting the optimum value of k_1 in (3.49), the minimum mean square error of the proposed estimator, up to the first order of approximation , is given by:

$$MSE_{min}(\hat{\mu}_{GMR}) \approx \bar{Y}^2 \Bigg\{ \left(1 - \frac{1}{4}\lambda C_x^2\right) - \frac{\left[1+\lambda\left[\left(B-\frac{1}{2}A-\frac{1}{4}\right)C_x^2+\left(\frac{1}{2}-A\right)\rho_{xy}C_xC_y\right]\right]^2}{1+\lambda\left[\left(2B-A-\frac{1}{4}\right)C_x^2+(1-2A)\rho_{xy}C_xC_y+\left(1-\rho_{xy}^2\right)C_y^2\right]} \Bigg\}. \tag{3.50}$$

For $\alpha = 1$ this generalized mixture estimator is given by:

$$\hat{\mu}_{GMR1} = \left[k_1\bar{y}\left(\frac{\bar{X}}{\bar{x}}\right) + k_2\left(\bar{X}-\bar{x}\right)\right]\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right). \tag{3.51}$$

The optimum values of k_1 and k_2 are given by:

$$k_{1(opt)} = \frac{1+\lambda\left[\frac{7}{8}C_x^2-\rho_{xy}C_xC_y\right]}{1+\lambda\left[2C_x^2-2\rho_{xy}C_xC_y+(1-\rho_{xy}^2)C_y^2\right]}, \quad \text{and} \tag{3.52}$$

$$k_{2(opt)} = \frac{\bar{Y}}{\bar{X}}\left[\frac{1}{2} - k_{1(opt)}\left(2-\rho_{xy}\frac{C_y}{C_x}\right)\right], \tag{3.53}$$

and the minimum mean square error, up to the first order of approximation, is given by:

$$MSE_{min}(\hat{\mu}_{GMR1}) = \bar{Y}^2\left\{\left(1-\frac{1}{4}\lambda C_x^2\right) - \frac{\left[1+\lambda\left(\frac{7}{8}C_x^2-\rho_{yx}C_xC_y\right)\right]^2}{1+\lambda\left[2C_x^2-2\rho_{yx}C_xC_y+\left(1-\rho_{yx}^2\right)C_y^2\right]}\right\}.$$

When $\alpha = 2$, this generalized mixture estimator is given by:

$$\hat{\mu}_{GMR2} = \left[k_1\bar{y}\left(\frac{\bar{X}}{\bar{x}}\right)^2 + k_2\left(\bar{X}-\bar{x}\right)\right]\exp\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right) \tag{3.54}$$

The optimum values for k_1 and k_2 are given as:

$$k_{1(opt)} = \frac{1+\lambda\left[\frac{23}{8}C_x^2 - 2\rho_{yx}C_xC_y\right]}{1+\lambda\left[6C_x^2 - 4\rho_{yx}C_xC_y + (1-\rho_{yx}^2)C_y^2\right]}, \quad \text{and} \tag{3.55}$$

$$k_{2(opt)} = \frac{\bar{Y}}{\bar{X}}\left[\frac{1}{2} - k_{1(opt)}\left(3 - \rho_{yx}, \frac{C_y}{C_x}\right)\right], \tag{3.56}$$

and the minimum mean square error is given by:

$$MSE_{min}(\hat{\mu}_{GMR2}) = \bar{Y}^2\left\{\left(1 - \frac{1}{4}\lambda C_x^2\right) - \frac{\left[1+\lambda\left(\frac{23}{8}C_x^2 - 2\rho_{yx}C_xC_y\right)\right]^2}{1+\lambda\left[6C_x^2 - 4\rho_{yx}C_xC_y + \left(1-\rho_{yx}^2\right)C_y^2\right]}\right\}.$$

In the next section we will derive the conditions under which our estimators perform better than ordinary sample mean, ratio, exponential and regression type estimators.

3.5 Efficiency Comparisons

In this section efficiency of the second proposed estimator is compared with some of the commonly used estimators. We did not use the first proposed estimator because it was constantly less efficient than the second proposed estimator in our simulations and numerical examples presented in next sections. Conditions under which the proposed estimator is more efficient are given below:

$MSE(\hat{\mu}_{GMR}) < MSE(\bar{y})$ if

$$\lambda C_y^2 - \left\{ \left(1 - \frac{1}{4}\lambda C_x^2\right) - \frac{\left\{1+\lambda\left[\left(B-\frac{1}{2}A-\frac{1}{4}\right)C_x^2+\left(-A+\frac{1}{2}\right)C_{xy}\right]\right\}^2}{1+\lambda\left[\left(2B-A-\frac{1}{4}\right)C_x^2+(-2A+1)\rho_{yx}C_xC_y+\left(1-\rho_{xy}^2\right)C_y^2\right]} \right\} > 0 \tag{3.57}$$

$MSE(\hat{\mu}_{GMR}) < MSE(\hat{\mu}_R)$ if

$$\lambda\left(C_x-\rho_{xy}C_y\right)^2+\lambda\left(1-\rho_{xy}^2\right)C_y^2 - \left\{ \left(1 - \frac{1}{4}\lambda C_x^2\right) - \frac{\left\{1+\lambda\left[\left(B-\frac{1}{2}A-\frac{1}{4}\right)C_x^2+\left(-A+\frac{1}{2}\right)C_{xy}\right]\right\}^2}{1+\lambda\left[\left(2B-A-\frac{1}{4}\right)C_x^2+(-2A+1)\rho_{yx}C_xC_y+\left(1-\rho_{xy}^2\right)C_y^2\right]} \right\} > 0 \tag{3.58}$$

$MSE(\hat{\mu}_{GMR}) < MSE(\hat{\mu}_{Reg})$ if the following inequality holds:

$$\lambda\bar{Y}^2C_y^2\left(1-\rho_{yx}^2\right)-\left\{\left(1-\frac{1}{4}\lambda C_x^2\right)\right. \tag{3.59}$$
$$\left.-\frac{\left\{1+\lambda\left[\left(B-\frac{1}{2}A-\frac{1}{4}\right)C_x^2+\left(-A+\frac{1}{2}\right)C_{xy}\right]\right\}^2}{1+\lambda\left\{\left(2B-A-\frac{1}{4}\right)C_x^2+(-2A+1)\rho_{yx}C_xC_y+\left(1-\rho_{xy}^2\right)C_y^2\right\}}\right\}>0$$

$MSE(\hat{\mu}_{GMR})<MSE(\hat{\mu}_{ER})$ if

$$\lambda\left(\frac{1}{2}C_x-\rho_{xy}C_y\right)^2+\lambda\left(1-\rho_{xy}^2\right)C_y^2-\left\{\left(1-\frac{1}{4}\lambda C_x^2\right)\right. \tag{3.60}$$
$$\left.-\frac{\left\{1+\lambda\left[\left(B-\frac{1}{2}A-\frac{1}{4}\right)C_x^2+\left(-A+\frac{1}{2}\right)C_{xy}\right]\right\}^2}{1+\lambda\left\{\left(2B-A-\frac{1}{4}\right)C_x^2+(-2A+1)\rho_{yx}C_xC_y+\left(1-\rho_{xy}^2\right)C_y^2\right\}}\right\}>0$$

Numerical examples and simulation results show that these conditions are generally true, and hence the proposed estimator may be preferred over the existing estimators when these conditions hold true.

3.6 Numerical Examples

In this section we compare the performances of different estimators with the proposed generalized mixture estimators using some real data sets whose summary statistics are in Table 8. Table 9 shows the Theoretical Percent Relative Efficiencies of the estimators as compared to the ordinary sample mean which is calculated from the following expression:

$$PRET(\hat{\mu}_i)=\frac{MSET(\bar{y})}{MSET(\hat{\mu}_i)}\times 100 \tag{3.61}$$

where $i = R; Reg; ER; S, RP; R, Reg; GK$, and $MSET$ is the theoretical mean square error. From Table 9 we can confirm that all the percent relative efficiencies are greater than 100 indicating that all estimators are better than the sample mean estimator. The proposed generalized estimators are more efficient than other estimators given in Table 9.

Table 8. Summary Statistics for the Real Populations Used in Comparing $\hat{\mu}_{GM}$ and $\hat{\mu}_{GMR}$ with other Mean Estimators

Parameters	***Population 1***	***Population 2***	***Population 3***	***Population 4***
$\boldsymbol{N}$	70	34	256	204
$\boldsymbol{n}$	25	20	100	50
$\boldsymbol{\rho_{yx}}$	0.7293	0.4491	0.887	0.71
$\overline{\boldsymbol{Y}}$	96.7	856.4118	56.47	966
$\overline{\boldsymbol{X}}$	175.2671	208.8824	44.45	26441
$\boldsymbol{C_y}$	0.6254	0.8561	1.42	2.4739
$\boldsymbol{C_x}$	0.8037	0.7205	1.40	1.7171
$\boldsymbol{f}$	0.3571	0.5882	0.3906	0.2451

(1) **Population 1 [Source: Singh and Chaudhary (1986), pp.108]**

(2) **Population 2 [Source: Singh and Chaudhary(1986), pp. 177]**

(3) **Population 3 [Source: Cochran (1977), pp. 196]**

(4) **Population 4 [Source: Kadilar & Cingi (2005)]**

Table 9. The Theoretical Percent Relative Efficiency for Various Mean Estimators

Estimators	***Population 1***	***Population 2***	***Population 3***	***Population 4***
$\hat{\mu}_Y$	**100**	**100**	**100**	**100**
$\hat{\mu}_R$	**128.6948**	**105.0011**	**448.3998**	**201.5302**
$\hat{\mu}_{Reg}$	**213.6198**	**125.2647**	**468.975**	**201.6536**
$\hat{\mu}_{ER}$	**210.2398**	**125.1392**	**271.3702**	**159.1883**
$\hat{\mu}_{S,RP}$	**213.6198**	**125.2647**	**468.975**	**201.6536**
$\hat{\mu}_{R,Reg}$	**214.6255**	**126.7737**	**470.2037**	**210.9342**
$\hat{\mu}_{GK}$	**215.7188**	**127.1322**	**472.0147**	**213.4533**
$\hat{\mu}_{GM1}$	**224.2375**	**128.9346**	**487.6321**	**226.2338**
$\hat{\mu}_{GM2}$	**240.5612**	**131.3791**	**518.698**	**226.2338**
$\hat{\mu}_{GMR1}$	**222.6746**	**128.7937**	**474.0536**	**215.9437**
$\hat{\mu}_{GMR2}$	**298.3243**	**137.8524**	**536.5996**	**242.7579**

3.7 Conclusion

In this chapter, we have proposed two generalized mixture estimators in simple random sampling without replacement by using information on an auxiliary variable. The proposed estimators are a mixture of some of the commonly known estimators. We have derived the minimum mean square errors up to the first order of approximation. Also we discussed two special cases $\alpha = 1$ and $\alpha = 2$. Numerical examples with real data show that both of the proposed estimators are more efficient than other estimators considered here. Also the estimators for $\alpha = 2$ perform better than the estimators with $\alpha = 1$.

We may note that at a theoretical level, one may be tempted to optimize α. Our goal though was to have a general family of estimators where many of the existing estimators become special cases of the proposed estimator with specific choice of α. For example, with $\alpha = 0$ our generalized mixture estimator II becomes combination of the regression and exponential ratio type estimators. For $\alpha = 1$, it involves the ratio term also. For $\alpha = -1$, it involves the product term.

In the next chapter we will use the proposed estimator $\hat{\mu}_{GMR}$ in the situation when the study variable is sensitive in nature and can not be observed directly, and a non- sensitive auxiliary variable is available.

CHAPTER IV

IMPROVED GENERALIZED MIXTURE ESTIMATORS WITH RRT MODELS

4.1 Introduction

Randomized response technique (RRT) is used to estimate the mean of a sensitive variable Y when direct observation is not possible. In this chapter, our focus is on estimation of the mean of a sensitive variable Y which cannot be observed directly using a non-sensitive auxiliary variable X. Sousa et al.(2010)[69] introduced the ratio type estimator and Gupta et al.(2012)[19] proposed the ordinary regression and a generalized regression-cum-ratio estimators based on RRT models. Following Bahl & Tuteja (1991) [2], Koyucu et al. (2014)[32] proposed the generalized exponential ratio type estimators to improve the efficiency of the mean estimator based on RRT models. In this chapter we propose an ordinary exponential ratio type estimator and a generalized mixture estimators where RRT estimators of the population mean of the study variable is further improved by using information about an auxiliary variable. We will use the following notations.

Let Y be the study sensitive variable which cannot be observed directly. Let X be a non sensitive auxiliary variable which has a strong positively correlation with Y, and let S be a scrambling variable. Assume that S is independent of Y and X. Also, assume that the population mean and the population variance of the scrambling variable are known and given as $\mu_s = 0$ and σ_S^2. The population mean and the population variance of the non-sensitive auxiliary variable are known and given as $\bar{X}$ and S_x^2. The population mean and the population variance of the study variable are unknown and given as $\bar{Y}$ and S_Y^2. Let a random sample of size n be drawn without replacement from a finite population $U = (U_1, U_2, \ldots, U_N)$. For ith unit, let y_i and x_i, respectively be the values of the study variable Y and the auxiliary variable X. The respondent is asked to report a scrambled response for Y given by:

$$Z = Y + S, \tag{4.1}$$

but is asked to provide the true response for X. Note that from (4.1) $\bar{Z} = \bar{Y}$ and,

$$C_z^2 = C_y^2 + \frac{\sigma_s^2}{\bar{Y}^2} \tag{4.2}$$

where C_z and C_y are the coefficients of the variation of the reported variable Z and the study variable Y, respectively. We will use the same error terms as in Sukhatme and Sukhatme (1970), given as:

$$e_z = \frac{\bar{z} - \bar{Z}}{\bar{Z}} \quad \text{and} \quad e_x = \frac{\bar{x} - \bar{X}}{\bar{X}}, \tag{4.3}$$

for which the following holds true:

$$\begin{gathered} E(e_z) = E(e_x) = 0, \quad E(e_z^2) = \lambda C_z^2, \quad E(e_x^2) = \lambda C_x^2 \\ E(e_z e_x) = \lambda C_{zx} = \lambda \rho_{zx} C_z C_x, \quad \text{where} \quad \lambda = \left(\frac{1}{n} - \frac{1}{N}\right). \end{gathered} \tag{4.4}$$

4.2 The Proposed Generalized Mixture Estimator in RRT

Following Bahl & Tuteja (1991) we propose the ordinary exponential ratio type estimator for estimating the population mean of the sensitive characteristic Y when non sensitive auxiliary variable X is used. This estimator is given by:

$$\hat{\mu}_{ER} = \bar{z} \exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right), \tag{4.5}$$

where $\bar{z}$ and $\bar{x}$ are the sample means of the reported responses and an auxiliary variable, respectively. Using error terms (4.3), this estimator can be written as:

$$\hat{\mu}_{ER} = \bar{Z}\left(1 + e_z\right) \exp\left[-e_x\left(e_x + 2\right)^{-1}\right]. \tag{4.6}$$

Up to the first order of approximation, this estimator can be written as:

$$\hat{\mu}_{ER} - \bar{Z} \approx \bar{Z}\left(e_z - \frac{1}{2}e_x - \frac{1}{2}e_z e_x + \frac{3}{8}e_x^2\right). \tag{4.7}$$

Recognizing that $\bar{Z} = \bar{Y}$ in (4.7), the bias of this estimator up to the first order of approximation, is given as:

$$Bias(\hat{\mu}_{ER}) \approx \lambda\bar{Y}\left(\frac{3}{8}C_x^2 - \frac{1}{2}\rho_{zx}C_z C_x\right) \tag{4.8}$$

By squaring equation (4.7) and using first order of approximation, we get:

$$\left(\hat{\mu}_{ER} - \bar{Z}\right)^2 \approx \bar{Z}^2\left(e_z^2 - e_z e_x + \frac{1}{4}e_x^2\right). \tag{4.9}$$

Taking expectation, the mean square error of the proposed estimator, up to the first order of approximation, is given by:

$$MSE(\hat{\mu}_{ER}) \approx \lambda \bar{Y}^2 \frac{1}{4}\left(4C_z^2 - 4\rho_{zx}C_zC_x + C_x^2\right). \tag{4.10}$$

The conditions under which the proposed estimator is more efficient than the ordinary sample mean and RRT ratio estimators are given below:

$$a) \quad MSE(\hat{\mu}_{ER}) < MSE(\hat{\mu}_Y) \qquad \text{if} \qquad \rho_{zx} > \frac{1}{4}\frac{C_x}{C_z}, \tag{4.11}$$

$$b) \quad MSE(\hat{\mu}_{ER}) < MSE(\hat{\mu}_R) \qquad \text{if} \qquad \rho_{zx} < \frac{3}{4}\frac{C_x}{C_z}. \tag{4.12}$$

By combining the regression, ratio and exponential estimators we further generalized the estimator in (4.5) and propose a generalized mixture estimator given by:

$$\hat{\mu}_{GRR} = \left[d_1\bar{z}\left(\frac{\bar{X}}{\bar{x}}\right)^{\alpha} + d_2\left(\bar{X} - \bar{x}\right)\right]\exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right), \tag{4.13}$$

where d_i $(i = 1, 2)$ and α are suitably chosen constants. We will consider two values for α ($\alpha = 1$ and $\alpha = 2$). Using error terms (4.3), this generalized mixture estimator $\hat{\mu}_{GRR}$ can be written as:

$$\hat{\mu}_{GRR} = \left[d_1\bar{Z}\left(1 + e_z\right)\left(1 + e_x\right)^{-\alpha} - d_2\bar{X}e_x\right]\exp\left[-\frac{e_x}{2}\left(1 + \frac{e_x}{2}\right)^{-1}\right] \tag{4.14}$$

Up to the first order of approximation, this estimator can be written as:

$$\begin{aligned}
\hat{\mu}_{GRR} &\approx \left[d_1\bar{Z}\left(e_z+1\right)\left(1-\alpha e_x+\frac{1}{2}\alpha\left(\alpha+1\right)e_x^2\right)-d_2\bar{X}e_x\right]\left(1-\frac{1}{2}e_x+\frac{3}{8}e_x^2\right)\\
&= d_1\bar{Z}-d_1\alpha\bar{Z}e_x+\frac{1}{2}d_1\bar{Z}\alpha\left(\alpha+1\right)e_x^2+d_1\bar{Z}e_y-d_1\alpha e_z e_x-d_2\bar{X}e_x\\
&-\frac{1}{2}d_1\bar{Z}e_x+\frac{1}{2}d_1\alpha\bar{Z}e_x^2-\frac{1}{2}d_1\bar{Z}e_z e_x+\frac{1}{2}d_2\bar{X}e_x^2+\frac{3}{8}d_1\bar{Z}e_x^2\\
&= d_1\bar{Z}-d_1\bar{Z}\left(\alpha+\frac{1}{2}\right)e_x+d_1\bar{Z}\left[\frac{1}{2}\alpha\left(\alpha+1\right)+\frac{1}{2}\alpha+\frac{3}{8}\right]e_x^2\\
&+d_1\bar{Y}e_z-d_1\bar{Y}\left(\alpha+\frac{1}{2}\right)e_x e_z-d_2\bar{X}\left(e_x-\frac{1}{2}e_x^2\right)\\
&= \bar{Z}+\left(d_1-1\right)\bar{Z}+d_1\bar{Z}\Bigg\{e_z-\left(\alpha+\frac{1}{2}\right)e_x-\left(\alpha+\frac{1}{2}\right)e_z e_x\\
&+\left[\frac{1}{2}\alpha\left(\alpha+2\right)+\frac{3}{8}\right]e_x^2\Bigg\}-d_2\bar{X}\left(e_x-\frac{1}{2}e_x^2\right). \qquad (4.15)
\end{aligned}$$

Using the substitutions

$$A=\alpha+\frac{1}{2} \quad \text{and} \quad B=\frac{1}{2}\alpha(\alpha+2)+\frac{3}{8}, \qquad (4.16)$$

we get

$$\hat{\mu}_{GRR}-\bar{Z}\approx(d_1-1)\bar{Z}+d_1\bar{Z}\left(e_z-Ae_x-Ae_z e_x+Be_x^2\right)-d_2\bar{X}\left(e_x-\frac{1}{2}e_x^2\right). \qquad (4.17)$$

By taking expectation on both sides of (4.17) and recognizing that $\bar{Z}=\bar{Y}$, the bias of the proposed estimator, up to the first order of approximation, is given by:

$$Bias(\hat{\mu}_{GRR})\approx(d_1-1)\bar{Y}+\lambda d_1\bar{Y}\left(BC_x^2-A\rho_{zx}C_zC_x\right)+\lambda d_2\bar{X}\ \frac{1}{2}C_x^2 \qquad (4.18)$$

By squaring Equation (4.17), and keeping terms only up to the the first order of approximation, we get:

$$
\begin{aligned}
(\hat{\mu}_{GRR}-\bar{Z})^2 \approx\ & (d_1-1)^2\bar{Z}^2 + d_1^2\bar{Z}^2\left(e_z - Ae_x - Ae_ze_x + Be_x^2\right)^2 + d_2^2\bar{X}^2\left(e_x-\frac{1}{2}e_x^2\right)^2 \\
& + 2d_1(d_1-1)\bar{Z}^2\left(e_z - Ae_x - Ae_ze_x + Be_x^2\right) \\
& - 2d_1d_2\bar{X}\bar{Z}\left(e_x-\frac{1}{2}e_x^2\right)\left(e_z - Ae_x - Ae_ze_x + Be_x^2\right) - 2d_2(d_1-1)\bar{X}\bar{Z}\left(e_x-\frac{1}{2}e_x^2\right) \\
=\ & (d_1-1)^2\bar{Z}^2 + d_1^2\bar{Z}^2\left(e_z^2 - 2Ae_ze_x + A^2e_x^2\right) + d_2^2\bar{X}^2e_x^2 \qquad (4.19) \\
& + 2d_1^2\bar{Y}^2\left(e_z - Ae_x - Ae_ze_x + Be_x^2\right) - 2d_1\bar{Z}^2\left(e_z - Ae_x - Ae_ze_x + Be_x^2\right) \\
& - 2d_1d_2\bar{X}\bar{Z}\left(e_ze_x - Ae_x^2\right) - 2d_1d_2\bar{X}\bar{Z}\left(e_x-\frac{1}{2}e_x^2\right) + 2d_2\bar{X}\bar{Z}\left(e_x-\frac{1}{2}e_x^2\right).
\end{aligned}
$$

This can be further simplified to:

$$
\begin{aligned}
(\hat{\mu}_{GRR}-\bar{Z})^2 \approx\ & (d_1-1)^2\bar{Z}^2 + d_1^2\bar{Z}^2\left[2e_z - 2Ae_x - 4Ae_ze_x + e_z^2 + \left(A^2+2B\right)e_x^2\right] \\
& + d_2^2\bar{X}^2e_x^2 - 2d_1\bar{Z}^2\left[e_z - Ae_x - Ae_ze_x + Be_x^2\right] \\
& - 2d_1d_2\bar{X}\bar{Z}\left[e_x + e_ze_x - \left(A+\frac{1}{2}\right)e_x^2\right] + 2d_2\bar{X}\bar{Z}\left(e_x-\frac{1}{2}e_x^2\right). \qquad (4.20)
\end{aligned}
$$

By taking expectation of (4.20) and recognizing that $\bar{Z}=\bar{Y}$, the mean square error of the proposed estimator, up to the first order of approximation, is given by:

$$
\begin{aligned}
MSE(\hat{\mu}_{GRR}) \approx\ & (d_1-1)^2\bar{Y}^2 + \lambda d_1^2\bar{Y}^2\left[\left(A^2+2B\right)C_x^2 - 4A\rho_{zx}C_zC_x + C_z^2\right] \\
& + \lambda d_2^2\bar{X}^2C_x^2 - 2\lambda d_1\bar{Y}^2\left(BC_x^2 - A\rho_{zx}C_zC_x\right) \\
& - 2\lambda d_1d_2\bar{X}\bar{Y}\left[\rho_{zx}C_zC_x - \left(A+\frac{1}{2}\right)C_x^2\right] - \lambda d_2\bar{X}\bar{Y}C_x^2. \qquad (4.21)
\end{aligned}
$$

Partially differentiating (4.21) with respect to d_2 and d_2, we get:

$$\frac{\partial MSE(\hat{\mu}_{GRR})}{\partial d_1} = 2\left(d_1 - 1\right)\bar{Y}^2 + 2\lambda d_1 \bar{Y}^2\Big[(A^2+2B)C_x^2 - 4A\rho_{zx}C_zC_x + C_z^2\Big] \quad (4.22)$$
$$- 2\lambda\bar{Y}^2\Big[BC_x^2 - A\rho_{zx}C_zC_x\Big] - 2\lambda d_2\bar{X}\bar{Y}\Big[\rho_{zx}C_zC_x - \left(A+\frac{1}{2}\right)C_x^2\Big],$$

$$\frac{\partial MSE(\hat{\mu}_{GRR})}{\partial d_2} = 2\lambda d_2\bar{X}^2C_x^2 - 2\lambda d_1\bar{X}\bar{Y}\left[\rho_{zx}C_zC_x - \left(A+\frac{1}{2}\right)\right] - \lambda\bar{X}\bar{Y}C_x^2. \quad (4.23)$$

Setting the first derivatives equal to zero, the optimum value of d_1 and d_2 are given by:

$$d_{1(opt)} = \frac{1+\lambda\left[\left(B-\frac{1}{2}A-\frac{1}{4}\right)C_x^2 + \left(\frac{1}{2}-A\right)\rho_{zx}C_zC_x\right]}{1+\lambda\left[\left(2B-A-\frac{1}{4}\right)C_x^2 + \left(1-2A\right)\rho_{zx}C_zC_x + (1-\rho_{zx}^2)C_z^2\right]}, \quad (4.24)$$

$$\text{and} \quad d_{2(opt)} = \frac{\bar{Y}}{\bar{X}}\left\{\frac{1}{2} - d_{1(opt)}\left[\left(A+\frac{1}{2}\right) - \rho_{zx}\frac{C_z}{C_x}\right]\right\}. \quad (4.25)$$

By substituting the optimum value of d_2 in the expression (4.21), we get:

$$
\begin{aligned}
MSE_{min}(\hat{\mu}_{GRR}) &\approx d_1^2\bar{Y}^2 - 2d_1\bar{Y}^2 + \bar{Y}^2 + \lambda d_1^2\bar{Y}^2\Big[\left(A^2+2B\right)C_x^2 - 4A\rho_{zx}C_zC_x + C_z^2\Big] \\
&+ \lambda\bar{X}^2C_x^2\frac{\bar{Y}^2}{\bar{X}^2}\left\{\frac{1}{2} - d_1\left[\left(A+\frac{1}{2}\right) - \rho_{zx}\frac{C_z}{C_x}\right]\right\}^2 \\
&- 2\lambda d_1\bar{Y}^2\Big(BC_x^2 - A\rho_{zx}C_zC_x\Big) \\
&- 2\lambda\bar{X}\bar{Y}d_1\frac{\bar{Y}}{\bar{X}}\left\{\frac{1}{2} - d_1\left[\left(A+\frac{1}{2}\right) - \rho_{zx}\frac{C_z}{C_x}\right]\left[\rho_{zx}C_zC_x - \left(A+\frac{1}{2}\right)C_x^2\right]\right\} \\
&- \lambda\bar{X}\bar{Y}C_x^2\frac{\bar{Y}}{\bar{X}}\left\{\frac{1}{2} - d_1\left[\left(A+\frac{1}{2}\right) - \rho_{yx}\frac{C_y}{C_x}\right]\right\} \\
&= d_1^2\bar{Y}^2\Bigg\{1 + \left[\left(2B - A - \frac{1}{4}\right)C_x^2 + \left(1-2A\right)\rho_{zx}C_zC_x + \left(1-\rho_{zx}^2\right)C_z^2\right]\Bigg\} \\
&+ \bar{Y}^2 - \frac{1}{4}\lambda\bar{Y}^2C_x^2 - 2d_1\bar{Y}^2\Bigg\{1 + \lambda\left[\left(B - \frac{1}{2}A - \frac{1}{4}\right)C_x^2 + \left(\frac{1}{2} - A\right)\rho_{zx}C_zC_x\right]\Bigg\}.
\end{aligned}
$$

Then substituting the optimum value of d_2 in (4.26), the minimum mean square error of the proposed estimator, up to the first order of approximation , is given by:

$$
\begin{aligned}
MSE_{min}(\hat{\mu}_{GRR}) \approx \bar{Y}^2\Bigg\{&\left(1 - \frac{1}{4}\lambda C_x^2\right) \\
&- \frac{\left\{1 + \lambda\left[\left(B - \frac{1}{2}A - \frac{1}{4}\right)C_x^2 + \left(\frac{1}{2} - A\right)\rho_{zx}C_zC_x\right]\right\}^2}{1 + \lambda\left[\left(2B - A - \frac{1}{4}\right)C_x^2 + \left(1-2A\right)\rho_{zx}C_zC_x + \left(1-\rho_{zx}^2\right)C_z^2\right]}\Bigg\}.
\end{aligned} \tag{4.26}
$$

For $\alpha = 1$ the generalized mixture estimator is given by:

$$
\hat{\mu}_{GRR1} = \left[d_1\bar{z}\left(\frac{\bar{X}}{\bar{x}}\right) + d_2\left(\bar{X} - \bar{x}\right)\right]\exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right). \tag{4.27}
$$

The optimum values of d_1 and d_2 are given by:

$$d_{1GRR1(opt)} = \frac{1 + \left[\frac{7}{8}C_x^2 - \rho_{zx}C_zC_x\right]}{1 + \lambda\left[2C_x^2 - 2\rho_{zx}C_zC_x + (1-\rho_{zx}^2)C_z^2\right]}, \quad \text{and} \tag{4.28}$$

$$d_{2GRR1(opt)} = \frac{\bar{Y}}{\bar{X}}\left[\frac{1}{2} - d_{1(opt)}\left(2 - \rho_{zx}\frac{C_z}{C_x}\right)\right], \tag{4.29}$$

and the minimum mean square error, up to the first order of approximation, is given by:

$$MSE_{min}(\hat{\mu}_{GRR1}) \approx \bar{Y}^2\left\{\left(1 - \frac{1}{4}\lambda C_x^2\right) - \frac{\left[1 + \lambda\left(\frac{7}{8}C_x^2 - \rho_{zx}C_zC_x\right)\right]^2}{1 + \lambda\left[2C_x^2 - 2\rho_{zx}C_zC_x + \left(1-\rho_{zx}^2\right)C_z^2\right]}\right\}.$$

For $\alpha = 2$ the generalized mixture estimator is given by:

$$\hat{\mu}_{GRR2} = \left[d_1\bar{z}\left(\frac{\bar{X}}{\bar{x}}\right)^2 + d_2\left(\bar{X} - \bar{x}\right)\right]\exp\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right). \tag{4.30}$$

The optimum values of d_1 and d_2 are given by:

$$d_{1GRR2(opt)} = \frac{1 + \lambda\left[\frac{23}{8}C_x^2 - 2\rho_{zx}C_zC_x\right]}{1 + \lambda\left[6C_x^2 - 4\rho_{zx}C_zC_x + (1-\rho_{zx}^2)C_z^2\right]}, \quad \text{and} \tag{4.31}$$

$$d_{2GRR2(opt)} = \frac{\bar{Y}}{\bar{X}}\left[\frac{1}{2} - d_{1(opt)}\left(3 - \rho_{zx}\frac{C_z}{C_x}\right)\right], \tag{4.32}$$

and the minimum mean square error up to the first order of approximation, is given as:

$$MSE_{min}(\hat{\mu}_{GRR2}) \approx \bar{Y}^2 \left\{ \left(1 - \frac{1}{4}\lambda C_x^2\right) - \frac{\left[1 + \lambda\left(\frac{23}{8}C_x^2 - 2\rho_{zx}C_zC_x\right)\right]^2}{1 + \lambda\left[6C_x^2 - 4\rho_{zx}C_zC_x + \left(1 - \rho_{zx}^2\right)C_z^2\right]} \right\}.$$

4.3 Efficiency Comparisons

In this section efficiency of the proposed estimator is compared with some of the commonly used RRT estimators. Conditions under which the generalized mixture RRT estimator is more efficient are given below:

$$MSE(\hat{\mu}_{GRR}) < MSE(\mu_Y) \quad \text{if}$$

$$\lambda C_z^2 - \left\{ \left(1 - \frac{1}{4}\lambda C_x^2\right) - \frac{\left\{1 + \lambda\left[\left(B - \frac{1}{2}A - \frac{1}{4}\right)C_x^2 + \left(\frac{1}{2} - A\right)\rho_{zx}C_zC_x\right]\right\}^2}{1 + \lambda\left[\left(2B - A - \frac{1}{4}\right)C_x^2 + (1 - 2A)\rho_{zx}C_zC_x + \left(1 - \rho_{zx}^2\right)C_y^2\right]} \right\} > 0 \tag{4.33}$$

$$MSE(\hat{\mu}_{GRR}) < MSE(\hat{\mu}_R) \quad \text{if}$$

$$\lambda\left(C_x - \rho_{zx}C_z\right)^2 + \lambda(1 - \rho_{zx}^2)C_z^2 - \left\{ \left(1 - \frac{1}{4}\lambda C_x^2\right) - \frac{\left\{1 + \lambda\left[\left(B - \frac{1}{2}A - \frac{1}{4}\right)C_x^2 + \left(\frac{1}{2} - A\right)\rho_{zx}C_zC_x\right]\right\}^2}{1 + \lambda\left[\left(2B - A + \frac{1}{4}\right)C_x^2 + (1 - 2A)\rho_{zx}C_zC_x + \left(1 - \rho_{zx}^2\right)C_z^2\right]} \right\} > 0 \tag{4.34}$$

It also holds:

$$MSE(\hat{\mu}_{GRR}) < MSE(\hat{\mu}_{Reg}) \quad \text{if}$$

$$\lambda\bar{Y}^2C_z^2\left(1-\rho_{zx}^2\right) - \left\{\left(1-\frac{1}{4}\lambda C_x^2\right) - \frac{\left\{1+\lambda\left[\left(B-\frac{1}{2}A-\frac{1}{4}\right)C_x^2+\left(\frac{1}{2}-A\right)\rho_{zx}C_zC_x\right]\right\}^2}{1+\lambda\left[\left(2B-A-\frac{1}{4}\right)C_x^2+(1-2A)\rho_{zx}C_zC_x+\left(1-\rho_{zx}^2\right)C_z^2\right]}\right\} > 0 \tag{4.35}$$

$$MSE(\hat{\mu}_{GRR}) < MSE(\hat{\mu}_{ER}) \quad \text{if}$$

$$\lambda\left(\frac{1}{2}C_x-\rho_{zx}C_z\right)^2 + \lambda\left(1-\rho_{zx}^2\right)C_z^2 - \left\{\left(1-\frac{1}{4}\lambda C_x^2\right) - \frac{\left\{1+\lambda\left[\left(B-\frac{1}{2}A-\frac{1}{4}\right)C_x^2+\left(\frac{1}{2}-A\right)\rho_{zx}C_zC_x\right]\right\}^2}{1+\lambda\left[\left(2B-A+\frac{1}{4}\right)C_x^2+(-2A+1)\rho_{zx}C_zC_x+\left(1-\rho_{zx}^2\right)C_z^2\right]}\right\} > 0 \tag{4.36}$$

We will use the real data and simulatied data to show that these conditions are generaly true, and hence the proposed estimator may be preferred over the existing estimators.

4.4 Numerical Examples

In this section,we compare the efficiency of proposed estimators with other existing RRT mean estimators using real data. The Population Statistics for the real data are given in Table 10. The scrambling variable S is assume to be a normal distribution with mean zero and standard deviation equal to 2. The reported response is given by $Z = Y + S$. Table 11 gives Theoretical Percent Relative Efficiency for various RRT estimators based on the first order of approximation. The Theoretical Percent

Relative Efficiency of the estimators as compared to the ordinary RRT sample mean are calculated from the following equation:

$$PRET(\hat{\mu}_i) = 100 \times \frac{MSET(\hat{\mu}_y)}{MSET(\hat{\mu}_i)} \tag{4.37}$$

where $i = R, Reg, ER, GRR, GER, GRR1$, and $GRR2$.

Table 10. Summary Statistics for the Real Populations Used in Comparing $\hat{\mu}_{GRR}$ with other RRT Mean Estimators

Parameters	***Population 1***	***Population 2***	***Population 3***	***Population 4***
$\boldsymbol{N}$	70	34	256	204
$\boldsymbol{n}$	25	20	100	50
$\boldsymbol{\rho_{yx}}$	0.7293	0.4491	0.887	0.71
$\boldsymbol{\rho_{zx}}$	0.81079	0.44909	0.8867	0.7099
$\boldsymbol{\bar{Y}}$	96.7	856.4118	56.47	966
$\boldsymbol{\bar{X}}$	175.2671	208.8824	44.45	26441
$\boldsymbol{S_x^2}$	19842.15	22650.18	3872.573	2061327175
$\boldsymbol{S_y^2}$	3657.368	537544.3	6430.019	5711084
$\boldsymbol{\sigma_s^2}$	3.67395	3.67395	3.67395	3.67395
$\boldsymbol{C_y}$	0.6254	0.8561	1.42	2.4739
$\boldsymbol{C_x}$	0.8037	0.7205	1.40	1.7171
$\boldsymbol{C_z}$	0.6257	0.8561	1.4204	2.4739
$\boldsymbol{f}$	0.3571	0.5882	0.3906	0.2451

Population 1 [Source: Singh and Chaudhary (1986), pp.108]

Population 2 [Source: Singh and Chaudhary(1986), pp. 177]

Population 3 [Source: Cochran (1977), pp. 196]

Population 4 [Source: Kadilar & Cingi (2005)]

Table 11. The Theoretical Percent Relative Efficiency for RRT Mean Estimators

Estimators	***Population 1***	***Population 2***	***Population 3***	***Population 4***
$\hat{\mu}_Y$	100	100	100	100
$\hat{\mu}_R$	1 76.3753	105.001	447.5094	201.5505
$\hat{\mu}_{Reg}$	291.875	125.2645	467.9889	201.6534
$\hat{\mu}_{ER}$	269.5187	125.1390	449.1049	159.3275
$\hat{\mu}_{GRR}$	292.8943	126.7898	472.3173	211.3242
$\hat{\mu}_{GER}$	294.468	127.1320	478.3395	213.413
$\hat{\mu}_{GRR1}$	303.6344	128.7935	485.3493	212.9479
$\hat{\mu}_{GRR2}$	431.1358	137.8521	775.2617	242.964

4.5 Conclusion

In this chapter, we proposed the generalized mixture estimator and the ordinary exponential ratio type estimator for the mean of a sensitive variable in simple random sampling without replacement by using information about a non-sensitive auxiliary variable. The proposed generalized mixture estimator is a mixture of some of the commonly known RRT estimators. Numerical examples show that for the proposed estimators all the percent relative efficiencies are greater 100 indicating that all these estimators are better than the RRT ordinary mean estimator and RRT ratio type estimator. We also note that the proposed generalized mixture estimator is more efficient than the other estimators considered here. The estimator for $\alpha = 2$ performs better than the one with $\alpha = 1$ for the numerical examples.

CHAPTER V

SIMULATION RESULTS

5.1 Introduction

In this chapter we compare the performance of different estimators with the proposed generalized mixture estimators in the situations when the study variable is non-sensitive, and when the study variable is sensitive and can not be observed directly. In both cases a non-sensitive auxiliary variable is available. The simulated populations we use have the same characteristics as the real data sets considered in the previous chapters. We will consider three finite populations of size $N = 5000$ each with summary statistics as presented in Table 12. The scrambling variable S is taken to be a normal variable with mean zero and standard deviation equal to 2. The reported response is given by $Z = Y + S$. For each population, we consider the sample sizes $n = 100, 200$ and 500. The empirical mean square error is estimated based on 10,000 samples selected from each populations. The R-code for the simulation study is given at the end of this dissertation.

Included in our comparisons will be the Sousa et al. (2010) transformed ratio type estimator given by:

$$\hat{\mu}_{TR} = \bar{z}\left(\frac{c\bar{X} + d}{c\bar{x} + d}\right), \tag{5.1}$$

where c and d are the unit-free parameters. The mean square error of this estimator, up to the first order of approximation, is given by:

$$MSE(\hat{\mu}_{TR}) \approx \lambda \bar{Y}^2 \left(\eta^2 C_x^2 - 2\eta \rho_{zx} C_z C_x + C_z^2 \right), \tag{5.2}$$

where $\eta = \frac{c\bar{X}}{c\bar{X}+d}$.

We will consider the four special cases given by:

$$\begin{aligned}
\hat{\mu}_{TR1} &= \bar{z}\frac{\bar{X} + \beta_1(x)}{\bar{x} + \beta_(x)}, \\
\hat{\mu}_{TR2} &= \bar{z}\frac{\bar{X} + \beta_2(x)}{\bar{x} + \beta_2(x)}, \\
\hat{\mu}_{TR3} &= \bar{z}\frac{\beta_1(x)\bar{X} + \beta_2(x)}{\beta_1(x)\bar{x} + \beta_2(x)}, \\
\hat{\mu}_{TR4} &= \bar{z}\frac{\beta_2(x)\bar{X} + \beta_1(x)}{\beta_2(x)\bar{x} + \beta_1(x)}.
\end{aligned}$$

where $\beta_1(x)$ is the coefficient of skewness and $\beta_2(x)$ is the coefficient of kurtosis.

5.2 Simulation Results

First we will show the Empirical Percent Relative Efficiencies and the Theoretical Percent Relative Efficiencies (in bold) for the non-RRT mean estimators based on the first order of approximation. The results are given in the Table 13. For the RRT mean estimators, the results are given in Table 14.

We can confirm that all the percent relative efficiencies (except for the ordinary ratio estimator $\hat{\mu}_R$) are greater than 100 indicating that all these estimators are better than the ordinary sample mean estimator. The Theoretical Percent Relative Efficiencies suggest that estimators with $\alpha = 2$ perform better than the proposed estimators with $\alpha = 1$.

For the RRT mean estimators the proposed generalized mixture estimators perform just like the non-RRT mean estimators.

Table 12. Summary Statistics Used in Comparing $\hat{\mu}_{GM}$, $\hat{\mu}_{GMR}$ and $\hat{\mu}_{GRR}$ with other Mean Estimators

Parameters	***Population 1***	***Population 2***	***Population 3***
N	5000	5000	5000
$\boldsymbol{\rho_{yx}}$	0.306173	0.6043378	0.8518795
$\boldsymbol{\rho_{zx}}$	0.3063796	0.6043924	0.8517596
$\boldsymbol{\bar{Y}}$	94.13349	94.54217	95.00613
$\boldsymbol{\bar{X}}$	61.48666	61.27170	60.99362
$\boldsymbol{S_y^2}$	6295.845	6301.445	6299.421
$\boldsymbol{S_x^2}$	3491.867	3503.012	3516.064
$\boldsymbol{\sigma_s^2}$	4.040761	4.040761	4.040761
$\boldsymbol{C_y}$	0.84291	0.83964	0.83541
$\boldsymbol{C_x}$	0.96105	0.96596	0.97217
$\boldsymbol{C_z}$	0.84294	0.8397219	0.8355349
$\boldsymbol{\beta_{1(x)}}$	0.00977	0.00896	0.00582
$\boldsymbol{\beta_{1(y)}}$	-0.01039	-0.01044	-0.00816482
$\boldsymbol{\beta_{2(x)}}$	2.92869	2.92334	2.925049
$\boldsymbol{\beta_{2(y)}}$	3.073645	3.049695	3.010963

The Theoretical and Empirical Percent Relative Efficiency for the non-RRT Mean Estimators:

Table 13. The Theoretical Percent Relative Efficiency (PRET) and the Empirical Percent Relative Efficiency (PREE) for the non-RRT Mean Estimators

Estimators	n	***Population 1***		***Population 2***		***Population 3***	
		PREE	**PRET**	PREE	**PRET**	PREE	**PRET**
$\hat{\mu}_Y$	100	100	**100**	100	**100**	100	**100**
$\hat{\mu}_R$		58.03119	**62.43019**	103.8177	**107.18**	258.5247	**269.1498**
$\hat{\mu}_{Reg}$		108.9269	**110.3438**	156.777	**157.5359**	361.5658	**364.5627**
$\hat{\mu}_{ER}$		100.4883	**102.469**	156.7632	**157.3259**	284.8589	**288.0067**
$\hat{\mu}_{S,RP}$		108.2584	**110.3438**	154.1957	**157.5359**	342.6528	**364.5627**
$\hat{\mu}_{R,Reg}$		107.1271	**111.0401**	152.4903	**158.2268**	329.7561	**365.2466**
$\hat{\mu}_{GK}$		107.0532	**111.3146**	153.5348	**158.6368**	345.8642	**366.3567**
$\hat{\mu}_{GM1}$		106.7733	**112.8668**	149.5855	**161.2368**	302.5733	**376.2145**
$\hat{\mu}_{GM2}$		103.7155	**115.1912**	142.0656	**165.4793**	266.2331	**395.8243**
$\hat{\mu}_{GMR1}$		107.6861	**113.8431**	155.9745	**161.0671**	339.9154	**370.5307**
$\hat{\mu}_{GMR2}$		100.60872	**127.3194**	149.3095	**178.9323**	271.671	**430.9847**
$\hat{\mu}_Y$	200	100	**100**	100	**100**	100	**100**
$\hat{\mu}_R$		60.08251	**62.43019**	105.7923	**107.18**	258.7645	**269.1498**
$\hat{\mu}_{Reg}$		109.8632	**110.3438**	157.5753	**157.5359**	357.2337	**364.5627**
$\hat{\mu}_{ER}$		101.4283	**102.469**	157.2676	**157.3259**	284.2063	**288.0067**
$\hat{\mu}_{S,RP}$		109.5522	**110.3438**	156.3128	**157.5359**	349.7985	**364.5627**
$\hat{\mu}_{R,Reg}$		108.8401	**110.6849**	155.5144	**157.8743**	343.7958	**364.8977**
$\hat{\mu}_{GK}$		108.7607	**110.8187**	155.9329	**158.0744**	351.5238	**365.4400**
$\hat{\mu}_{GM1}$		108.6089	**111.5749**	154.1839	**159.3396**	329.7909	**370.2103**
$\hat{\mu}_{GM2}$		107.2027	**112.6952**	150.2984	**161.3735**	307.7655	**379.39**

Estimators	***n***	***Population 1***		***Population 2***		***Population 3***	
		PREE	**PRET**	PREE	**PRET**	PREE	**PRET**
$\hat{\mu}_{GMR1}$		108.9538	**112.0474**	157.6905	**159.2574**	350.1518	**367.4717**
$\hat{\mu}_{GMR2}$		108.7825	**118.2626**	143.0767	**167.5455**	274.3496	**394.8204**
$\hat{\mu}_{Y}$	500	100	**100**	100	**100**	100	**100**
$\hat{\mu}_{R}$		61.0523	**62.4302**	106.2602	**107.180**	269.9781	**269.1498**
$\hat{\mu}_{Reg}$		110.1817	**110.3438**	156.1066	**157.5359**	365.7146	**364.5627**
$\hat{\mu}_{ER}$		101.9275	**102.469**	155.4934	**157.3259**	288.2839	**288.0067**
$\hat{\mu}_{S,RP}$		109.8801	**110.3438**	155.3295	**157.5359**	361.9954	**364.5627**
$\hat{\mu}_{R,Reg}$		109.6644	**110.4717**	155.1292	**157.6628**	359.6135	**364.6883**
$\hat{\mu}_{GK}$		109.6368	**110.5218**	155.4442	**157.7377**	362.7923	**364.8914**
$\hat{\mu}_{GM1}$		109.6038	**110.8045**	154.5626	**158.2103**	353.8349	**366.6674**
$\hat{\mu}_{GM2}$		109.2878	**111.2206**	152.8622	**158.9633**	343.9008	**370.0193**
$\hat{\mu}_{GMR1}$		109.8011	**110.9804**	156.2002	**158.1797**	362.6607	**365.6504**
$\hat{\mu}_{GMR2}$		109.0028	**113.2309**	150.7239	**161.1911**	331.9076	**375.4523**

Table 14. The Theoretical Percent Relative Efficiency (PRET) and the Empirical Percent Relative Efficiency (PREE) for the RRT Mean Estimators

Estimators	n	*Population 1*		*Population 2*		*Population 3*	
		PREE	**PRET**	PREE	**PRET**	PREE	**PRET**
$\hat{\mu}_Y$	100	100	**100**	100	**100**	100	**100**
$\hat{\mu}_R$		58.71567	**62.48659**	98.51078	**107.2565**	255.0575	**269.1183**
$\hat{\mu}_{TR1}$		58.72834	**62.49837**	98.53273	**107.2776**	255.1092	**269.1685**
$\hat{\mu}_{TR2}$		62.42388	**65.93512**	105.443	**113.9063**	279.7681	**293.0594**
$\hat{\mu}_{TR3}$		108.7746	**108.9033**	122.5912	**123.0173**	124.6077	**124.8461**
$\hat{\mu}_{TR4}$		58.7200	**62.49061**	98.51829	**107.2637**	255.0752	**269.1355**
$\hat{\mu}_{Reg}$		109.1013	**110.4212**	153.724	**157.6236**	357.1331	**364.4144**
$\hat{\mu}_{ER}$		100.9350	**102.5526**	153.4705	**157.4118**	283.6987	**287.9487**
$\hat{\mu}_{GRR}$		106.6441	**111.1243**	154.1921	**158.3213**	342.5125	**365.1052**
$\hat{\mu}_{GER}$		107.852	**111.3926**	151.8842	**158.7251**	329.1788	**366.2081**
$\hat{\mu}_{GRR1}$		107.8473	**113.9221**	153.2379	**161.156**	336.7513	**370.3777**
$\hat{\mu}_{GRR2}$		101.6116	**127.4049**	136.203	**179.0270**	273.0817	**430.7435**
$\hat{\mu}_Y$	200	100	**100**	100	**100**	100	**100**
$\hat{\mu}_R$		62.44458	**62.48659**	104.9488	**107.2565**	258.9979	**269.1183**
$\hat{\mu}_{TR1}$		62.45694	**62.49837**	104.9705	**107.2776**	259.0479	**269.1685**
$\hat{\mu}_{TR2}$		66.05773	**65.93512**	111.7769	**113.9063**	282.9415	**293.0594**
$\hat{\mu}_{TR3}$		109.1664	**108.9033**	122.9115	**123.0173**	124.6378	**124.8461**
$\hat{\mu}_{TR4}$		62.4488	**62.49061**	104.9563	**107.2637**	259.015	**269.1355**
$\hat{\mu}_{Reg}$		110.5068	**110.4212**	156.5281	**157.6236**	355.9265	**364.4144**
$\hat{\mu}_{ER}$		103.2897	**102.5526**	156.6336	**157.4118**	284.161	**287.9487**
$\hat{\mu}_{GRR}$		109.0753	**110.764**	156.7457	**157.9637**	350.5665	**364.7511**
$\hat{\mu}_{GER}$		109.6907	**110.8964**	155.9092	**158.1624**	343.4442	**365.2916**

Estimators	***n***	***Population 1***		***Population 2***		***Population 3***	
		PREE	**PRET**	PREE	**PRET**	PREE	**PRET**
$\hat{\mu}_{GRR1}$		109.4142	**112.1255**	156.6565	**159.3457**	348.1898	**367.3212**
$\hat{\mu}_{GRR2}$		109.9778	**118.3439**	150.6148	**167.6366**	310.2307	**394.6324**
$\hat{\mu}_{Y}$	500	100	**100**	100	**100**	100	**100**
$\hat{\mu}_{R}$		62.28899	**62.48659**	103.5708	**107.2565**	267.0283	**269.1183**
$\hat{\mu}_{TR1}$		62.3010	**62.49837**	103.5917	**107.2776**	267.0775	**269.1685**
$\hat{\mu}_{TR2}$		65.80321	**65.93512**	110.155	**113.9063**	290.4467	**293.0594**
$\hat{\mu}_{TR3}$		108.9663	**108.9033**	122.5472	**123.0173**	124.4845	**124.8461**
$\hat{\mu}_{TR4}$		62.29309	**62.49061**	103.5779	**107.2637**	267.0451	**269.1355**
$\hat{\mu}_{Reg}$		110.5143	**110.4212**	154.6682	**157.6236**	359.0418	**364.4144**
$\hat{\mu}_{ER}$		102.6987	**102.5526**	154.7359	**157.4118**	283.2339	**287.9487**
$\hat{\mu}_{GRR}$		110.3119	**110.5494**	154.9472	**157.7508**	356.7733	**364.5403**
$\hat{\mu}_{GER}$		110.4029	**110.5993**	154.5534	**157.8255**	353.5611	**364.7431**
$\hat{\mu}_{GRR1}$		110.4402	**111.058**	155.0007	**158.2676**	355.6599	**365.5013**
$\hat{\mu}_{GRR2}$		109.636	**113.3097**	153.0042	**161.2800**	341.3199	**375.2902**

CHAPTER VI

OVERALL CONCLUSIONS AND FUTURE WORK

In this dissertation we proposed some generalized mixture estimators which are mixtures of the commonly used mean estimators. We also extended the proposed estimators to the situation when the study variable is sensitive and a non-sensitive auxiliary variable is available.

One may extend the proposed estimators to the more complex optional RRT situations. The parameters one need to estimate for optimal models are the population mean of the sensitive study variable and the proportion of respondents who consider the question sensitive (the sensitivity level W of the study variable). Since in the optional RRT models two parameters need to be estimated, a larger sample size is needed.

We have examined adequacy of the first order approximation for the theoretical mean square errors in the ratio type mean estimators and the ratio type variance estimators. One can examine the adequacy of second order of approximation and robustness in the ratio and regression type estimators as well.

The proposed estimators may also be extended to more complex sampling designs as well such stratified sampling and multi-stage sampling.

One could also consider the multivariate case where more than one non-sensitive auxiliary variables are available.

REFERENCES

[1] Asghar A.,Aullah, A. and Hanif, M. (2014). Generalized exponential type estimator for population variance in survey sampling *Revista Colombiana de Estadistica,Vol* 37(1), 211-222

[2] Bahl, S., Tuteja, R.K.(1991). Ratio and product type exponential estimators *Information and Optimization Sciences, Vol* 12(1):159-163

[3] Bedi P.K.(1996). Efficient utilization of auxiliary information at estimation stage *Biometrical Journal* 38(8), 973-976

[4] Bhushan, S.(2012). Some efficient sampling strategies based on ratio type estimators *Electronic Journal of Applied Statistical Analysis,Vol* 5(1), 74-78

[5] Chaudhuri, A. (2012). Unbiased estimation of a sensitive proportion in general sampling by three non-randomized response schemes. *Journal of Statistical Theory and Practice, Vol* 6(2), 376-381

[6] Cochran W.G. (1940). The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce. *Journal of Agricultural Science, Vol* 59, 1225-1226

[7] Cochran W.G. (1977).Sampling Techniques. *Third Edition, Wiley Eastern Limited*

[8] Das, A.K. and Tripathi, T.P.(1978). Use of auxiliary information in estimating the finite population variance *Sankhya, Vol* 40, 139-148

[9] David, I. P. and Sukhatme, B. V.(1974). On the bias and mean square error of the ratio estimator *Journal of the American Statistical Association,Theory and Methods Section, Vol* 69, 464-466

[10] Garcia, M. K. and Cebrain, A.A. (1999). Variance estimation using auxiliary information: An almost unbiased multivariate ratio estimation. *Metrika, Vol.* 45, 171-178

[11] Grover, L.K. (2010). A correction note on improvement in variance estimation using auxiliary information . *Communications in Statistics - Theory and Methods, Vol.* 39, 753-764

[12] Grover, L.K. and Kaur,P. (2011).An improved exponential estimator of finite population mean in simple random sampling using an auxiliary information. *Applied mathematics and computation, Vol.* 218, 3093-3099

[13] Grover, L.K. and Kaur,P. (2014). A generalized class of ratio type exponential estimators of population mean under linear transformation of auxiliary variable *Communications in Statistics- Simulation and Computation, Vol.* 43(7), 1552-1574

[14] Gupta S., Gupta B. and Singh. S. (2002). Estimation of sensitivity of personal interview survey questions. *Journal of Statistical Planning and Inference, Vol.* 100(2), 239-247

[15] Gupta S.,Thornton B., Shabbir J. and Singhal S. (2006). A comparison of multiplicative and additive optional RRT models *Journal of Statistical Theory and Applications, Vol.* 5(3), 226-239

[16] Gupta S. and Shabbir J.(2008). On improvement in estimating the population mean in simple random sampling *Journal of Applied Statistics, Vol.* 35(5), 556-5569

[17] Gupta S. and Shabbir J.(2008). Variance estimation in simple random sampling using auxiliary information *Hacettepe Journal of Mathematics and Statistics, Vol.* 37, 57-67

[18] Gupta S., Shabbir J. and Sehra S. (2010). Mean and sensitivity estimation in optional randomized response model *Journal of Statistical Planning and Inference, Vol.* 140 (10), 2870-2874

[19] Gupta S., Shabbir J., Sousa R. and Corte-Real P. (2012). Estimation of the mean of a sensitive variable in the presence of auxiliary information. *Communications in Statistics-Theory and Methods, Vol.* 41 (13), 2394-2404

[20] Isaki C.T. (1983). Variance estimation using auxiliary information. *Journal of American Statistical Association, Vol* 78, 117-123

[21] Jhajj, H.S., Sharma, M.K. and Grover, L.K. (2006). Dual of ratio estimators of finite population mean obtained on using linear transformation to auxiliary variable *Journal of Japan Statistical Society, Vol* 36, 107-119

[22] Kadilar, C. and Cingi,H. (2003). A study on the chain ratio type estimator *Hacettepe Journal of Mathematical and Statistics, Vol.* 32, 105-108

[23] Kadilar, C. and Cingi,H. (2004). Ratio estimators in simple random sampling. *Applied mathematics and computation, Vol.* 151, pp.893-902

[24] Kadilar, C. and Cingi, H.(2006). Improvement in variance estimation using auxiliary information. *Hacettepe Journal of Mathematicas and Statistics, Vol.* 35(1), 111-115

[25] Kadilar, C. and Cingi, H.(2006). Ratio estimators for the population variance in sample and stratified random sampling. *Applied Mathematics and Computation, Vol.* 173, 1047-1059

[26] Kadilar C., Candan M. and Cingi H.(2007). Ratio estimators using robust regression *Hacettepe Journal of Mathematics and Statistics, Vol.* 36(2), 81-188

[27] Kalucha G., Gupta S. and Dass, B.K.(2015). Ratio estimation of finite population mean using optional randomized response models *Journal of Statistical Theory and Practice, Vol.* 9(3), 633-645

[28] Khan, M. and Shabbir, J.(2013). A ratio-type estimator for the estimation of population variance using quartiles of an auxiliary variable *Journal of Statistics Applications and Probability, Vol.* 2(3), 319-325

[29] Khoshnevisan M., Singh R., Chauhan P., Sawn N. and Smarandache F. (2007). A general family of estimators for estimating population mean using known value of some population parameter(s) *Far East Journal of Theoretical Statistics, Vol.* 22, 181-191

[30] Koyuncu, N. and Kadilar, C. (2009). Efficient Estimators for the population mean *Hacettepe Journal of Mathematics and Statistics, Vol.* 38(2), 217-225

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

[32] Koyuncu, N., Gupta S. and 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, Vol.* 43, 1583-1594

[33] Mohanty, S. (1967). Combination of Regression and Ratio Estimate. *Journal of Indian Statistical Association, Vol.* 5, 16-19

[34] Murthy, M.N. (1964). Product method of estimation. *The Indian Journal of Statistic A, Vol.* 26, 69-74

[35] Murthy, M.N. (1967). Sampling Theory and Methods. *Calcutta, India: Statistical Publishing Society*

[36] Naik, V.D. and Gupta, P.C.(1991). A general class of estimators for estimating population mean using auxiliary information *Metrica, Vol.* 38, 11-17

[37] Olkin, I.(1958). Multivariate ratio estimation for finite populations *Biometrica, Vol* 45, 154-165

[38] Pathak, P.K.(1964). On sampling schemes providing unbiased ratio estimators *The Annals of Mathematical Statistics, Vol* 35(1), 222-231

[39] Perri P.F (2004). On the efficient use regression-in-ratio estimator in simple random sampling *Atti della XLII Riunione Scientifica della Societa Italiana di Statistica Bari, Vol* 9(11), 537-540.

[40] Perri P.F (2007). Improved ratio-cum-product type estimators *Statistics in Transition, Vol* 8(1), 51-69.

[41] Prasad,B. and Singh, H.P.(1990). Some improved ratio type estimators of finite population variance in sample surveys *Communication in statistics - Theory and Methods, Vol* 19, 1127-1139

[42] Rao, T.J.(1991). On certain methods of improving ratio and regression estimator. *Communication in statistics - Theory and Methods, Vol* 20(10), 3325-3340

[43] Ray, S.K. and Sahai, A.(1980). Efficient families of ratio and product type estimators. *Biometrica, Vol* 67, 211-215

[44] Ray, S.K. and Singh R.K. (1981). Difference-cum-ratio type estimators *Journal of Indian Statistical Association, Vol* 19, 147-151

[45] Reddy, V. N. (1973). On ratio and product methods of estimation. *Journal of Indian Statistical Association, Vol* 35 (3), 211-215

[46] Robinson, J. (1987). Conditioning ratio estimates under simple random sampling. *Journal of American Statistical Association, Vol* 82(399), 826-831.

[47] Robson, D.S.(1957). Applications of multivariate polykays to the theory of unbiased ratio-type estimation. *Journal of American Statistical Association, Vol* 52 (282), 491-508.

[48] Saha, A. (2008) A randomized response technique for quantitative data under unequal probability sampling. *Journal of Statistical Theory and Practice,Vol* 2(4), 589-596

[49] Samiuddin, M. and Hanif, M. (2006). Estimation of population means in a single and two phase sampling with or additional information. *Pakistan Journal of Statistics,Vol* 23, 99-118

[50] Sanaullah, A., Ali H.A., Noor ul Amin M. and Hanif M. (2006). Generalized exponential chain ratio estimators under stratified two-phase random sampling. *Applied Mathematics and Computation,Vol* 226, 541-547

[51] Shabbir J. and Gupta S.(2007). On improvement in variance estimation using auxiliary information . *Communication in statistic- Theory and Methods,Vol* 36(12), 2177-2185

[52] Shabbir J. and Gupta S.(2010). A new difference-cum-exponential type estimator of finite population mean in simple random sampling. *Colombian Journal of Statistics,Vol* 37(1), 199-211

[53] Shabbir J. and Gupta S.(2015). A note on generalized exponential type estimator for population variance in survey sampling *Revista Colombiana de Estadistia,Vol* 38(2)

[54] Shabbir J. and Yaab, M.Z.(2003). Improvement over transformed auxiliary variable in estimating the finite populations mean. *Biometrical Journal,Vol* 45(6), 723-729

[55] Sharma,B. and Tailor,R (2010). A new ratio-cum-dual to ratio estimator of finite population mean in simple random sampling *Global Journal of Science Frontier Research, Vol* 10(1), 27-31

[56] Singh, G.N.(2003). On the improvement of product method of estimation in sample surveys *Journal of the Indian Society of Agricultural Statistic, Vol* 56(3), 267-275

[57] Singh H.P., Upadhyaya U.D. and Namjoshi U.D.(1988). Estimation of finite population variance *Current Science, Vol* 57, 1331-1334

[58] Singh, H.P., Chandra,P. and Singh, S. (2003). Variance estimation using multi-auxiliary information for random non-response in survey sampling *STATISTICA, anno LXIII* (1), 23-40

[59] Singh R., Chauhan P., Sawan N. and Smarandache, F. (2011). Improved exponential estimator for population variance using two auxiliary variables *Italian Journal of Pure and Applied Mathematics, Vol* 28, 101-108

[60] Singh,H.P. and Agnihotri, N.(2008). A general procedure of estimating population mean using auxiliary information in sample surveys *Statistics in Transition, Vol* 9(1), 71-78

[61] Singh,H.P. and Espejo,M.R.(2003). On linear regression and ratio product estimation of a finite population mean *The Statistician, Vol* 52, 59-67

[62] Singh, H.P. and Tailor R. (2003). Use of known correlation coefficient in estimating the finite population mean . *Statistics in Transition, Vol* 6(4), 555-560

[63] Singh, H.P., Tailor R.,Tailor R. and Kakran, M.S.(2004). An improved estimator of population mean using power transformation. *Journal of the Indian Society of Agricultural Statistics, Vol* 58(2), 223-230

[64] Singh, R., Chauhan, P. and Sawan, N.(2008). On linear combination of ratio-product type exponential estimator for estimating finite population mean. *Statistic in transition, Vol* 9 (1), 105-115.

[65] Singh R., Chauhan, Sawan N. and Smarandache F. (2009). Improvement in estimating the population mean using exponential estimator in simple random sampling *Journal of Statistics and Economics, Vol* 3(A09), 13-18

[66] Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2011). Improved exponential estimator for population variance using two auxiliary variables. *Italian Journal of Pure and Applied Mathematics, Vol* 28, 101-108

[67] Singh, P. and K. Vishwakarman, K. (2007). Modified exponential ratio and product estimators for finite population mean in double sampling. *Australian Journal of Statistics, Vol* 52, 59-67

[68] Sisoda, B.V., Dwivedi V.K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable . *Journal Indian society of agricultural statistics, Vol* 33(1), 13-18

[69] Sousa R., Shabbir J., Real P.C. and Gupta S. (2010). Ratio estimation of the mean of a sensitive variable in the presence of auxiliary information. *Journal of Statistical Theory and Practice, Vol* 4 (3), 495-507.

[70] Surendra K. Srivastava (1971). A Generalized Estimator for the Mean of a Finite Population Using Multi-Auxiliary Information . *Journal of the American Statistical Association , Vol* 66 (334), 404-407.

[71] Subramani, J. and Kumarapandiyan G. (2012b). Modified ratio estimators using known median and coefficient of kurtosis *American Journal of Mathematics and Statistics, Vol* 2(4), 95-100

[72] Subramani, J. and Kumarapandiyan G. (2012b). Estimation of population mean using known median and coefficient of skewness. *American Journal of Mathematics and Statistics, Vol* 2 (5), 101-107

[73] Subramani, J. and Kumarapandiyan G. (2012c). Estimation of population mean using coefficient of variation and median of an auxiliary variable . *International Journal of Probability and Statistic, Vol* 1(4), 111-118

[74] Subramani, J. and Kumarapandiyan G. (2012b). Variance estimation using median of the auxiliary variable. *International Journal of Probability and Statistic, Vol* 1(3), 36-40

[75] Subramani, J. and Kumarapandiyan G. (2012c). Variance estimation using quartiles and their functions of the auxiliary variable. *International Journal of Statistics and Applications, Vol* 2 (5), 67-72

[76] Subramani, J. and Kumarapandiyan G. (2013). A new modified ratio estimator for estimation of population mean when median of the auxiliary variable is known *Pakistan Journal of Statistics and Operation Research, Vol.* 9 (2), 137-145

[77] Subramani, J. (2013). Generalized modified ratio estimator for estimation of finite population mean. *Journal of modern applied statistical method, Vol.* 12(2), 121-155

[78] Sukhatme, P., Sukhatme, B. and Sukhatme, S.(1970). Sampling theory of surveys with applications. Third edition, *Iowa state university press,*

[79] Tin, M.(1965) .Comparison of some ratio estimators *ournal of the American Statistical Association, Vol* 60, 294-307

[80] Upadhyaya, L.N. and Singh H.P. (1999). Use of transformed auxiliary variable in estimating the finite population mean . *Biometrical Journal, Vol* 41(5), 627-636

[81] Upadhyaya, L.N. and Singh H.P. (2006). Almost unbiased ratio and product type estimators of finite population variance in sample surveys *Statistics in Transition, Vol* 7(5), 1087-1096

[82] Upadhyaya L., Singh H., Chatterjee S. and Yadav, R.(2011). Improved ratio and product exponential type estimators *Journal of Statistical Theory and Practice, Vol* 5(2), 285-302

[83] Warner S.L.(1965). Randomized response: a survey technique for eliminating evasive answer bias *Journal of the American Statistical Association, Vol* 60 (309), 63-69

[84] Warner S.L.(1971). The linear randomized resonse model *Journal of the American Statistical Association, Vol* 66 (336), 884-888

[85] Wolter, K.M.(1985). Introduction to Variance Estimation *New York, NY: Springer-Verlag*

[86] Zatezalo, T., Gupta S., Yadav S.K. and Shabbir J. (2016). Assessing the Adequacy of First Order Approximations in Ratio Type Estimators *Journal of Interdisciplinary Mathematics,* accepted

[87] Zatezalo, T., Gupta S. and Shabbir J. (2016). A Generalized Mixture Estimator of Finite Population Mean in Simple Random Sampling *Communication in Statistics - Theory and Methods,* submitted

[88] Yadav, S.K. and Kadilar, C. (2013). Improved exponential type ratio estimator of population variance . *Revista Colombiana de Estadistica, Vol* 36(1), 145-152

[89] Yan, Z. and Tian B. (2010). Ratio method to the mean estimation using co-efficient of skewness of auxiliary variable. *ICICA, Part II, CCIS Vol* 106, 103-110

APPENDIX A

R CODE I

```
library("copula")

# Generate N=5,000 observation of bivariate normal, bivariate Poisson and

# bivariate  gamma distributions with the population mean (X,Y)=(4,6) and

# the population variance (4,9)

#distributions with mean(x)=4 and mean(y)=6.

# mv.BN<- mvdc(normalCopula(0.5), c("norm", "norm"),list(list(mean=4,sd=2),

#  list(mean=6,sd=3)))

# lambda1<-4; lambda2<-6

# mv.PP<- mvdc(normalCopula(-0.5), c("pois", "pois"),

# list(list(lambda<-lambda1), list(lambda<-lambda2)))

# alpha1<-4; beta1<-1; alpha2<-4; beta2<-1.5

# mv.GG <- mvdc(normalCopula(0.2), c("gamma", "gamma"),list(list(shape=

#alpha1, scale=beta1), list(shape=alpha2,scale=beta2)))

# population.data<-rMvdc(5,000,mv.BN);  population.data<-rMvdc(5,000,mv.PP)

# and  population.data<-rMvdc(5,000,mv.GG);

population.data<-read.table("population.data.N.0.8",header=TRUE)

X<-population.data$X ; Y<-population.data$Y


# # # # #         THE SUMMARY STATISTICS         # # # # #


N<-nrow(population.data); rhoyx=cor(X,Y);  mux<-mean(X); muy<-mean(Y);

 varx<-var(X); vary<-var(Y); stdevx<-sd(X); stdevy<-sd(Y);

library("moments") #install packages for skewness and kurtosis
```

```
beta1x<-skewness(X); beta2x<-kurtosis(X)# skewness and kurtosis for X
beta1y<-skewness(Y); beta2y<-kurtosis(Y)# skewness and kurtosis for Y
cx<-stdevx/mux; cy<-stdevy/muy; md<-quantile(X,0.5) # coeff. of var.

# # # #   T H E   S I M U L A T I O N   P A R T     # # # #

xbar<-numeric(b); ybar<-numeric(b); tR<-numeric(b); tP<-numeric(b);
tER<-numeric(b); tEP<-numeric(b); t1<-numeric(b); t2<-numeric(b);
t3<-numeric(b); t4<-numeric(b); t5<-numeric(b); t6<-numeric(b);
t7<-numeric(b); t8<-numeric(b); t9<-numeric(b); t10<-numeric(b);
t11<-numeric(b); t12<-numeric(b); t13<-numeric(b); t14<-numeric(b);
t15<-numeric(b); t16<-numeric(b); t17<-numeric(b); t18<-numeric(b);
t19<-numeric(b); t20<-numeric(b); t21<-numeric(b)
mse.tR<-numeric(b); mse.tP<-numeric(b); mse.tEP<-numeric(b);
mse.tER<-numeric(b); mse.t1<-numeric(b); mse.t2<-numeric(b);
mse.t3<-numeric(b); mse.t4<-numeric(b); mse.t5<-numeric(b);
mse.t6<-numeric(b); mse.t7<-numeric(b); mse.t8<-numeric(b);
mse.t9<-numeric(b); mse.t10<-numeric(b); mse.t11<-numeric(b);
mse.t12<-numeric(b); mse.t13<-numeric(b); mse.t14<-numeric(b);
mse.t15<-numeric(b); mse.t16<-numeric(b); mse.t17<-numeric(b);
mse.t18<-numeric(b); mse.t19<-numeric(b); mse.t20<-numeric(b)
mse.t21<-numeric(b); n<-100; b<-10000; f<-n/N # n=100, 200 and 500
for (k in 1:b){s<-sample(N,n,replace=TRUE);
 xbar[k]<-mean(X[s])
 ybar[k]<-mean(Y[s])
 tR[k]<-ybar[k]*(mux/xbar[k])
```

```
tP[k]<-ybar[k]*(xbar[k]/mux)
tER[k]<-ybar[k]*exp((mux-xbar[k])/(mux+xbar[k]))
tEP[k]<-ybar[k]*exp((xbar[k]-mux)/(xbar[k]+mux))
t1[k]<-ybar[k]*((mux+cx)/(xbar[k]+cx))
t2[k]<-ybar[k]*((xbar[k]+cx)/(mux+cx))
t3[k]<-ybar[k]*((beta2x*xbar[k]+cx)/(beta2x*mux+cx))
t4[k]<-ybar[k]*((cx*xbar[k]+beta2x)/(cx*mux+beta2x))
t5[k]<-ybar[k]*((xbar[k]+stdevx)/(mux+stdevx))
t6[k]<-ybar[k]*((beta1x*xbar[k]+stdevx)/(beta1x*mux+stdevx))
t7[k]<-ybar[k]*((beta2x*xbar[k]+stdevx)/(beta2x*mux+stdevx))
t8[k]<-ybar[k]*(mux+rhoyx)/(xbar[k]+rhoyx)
t9[k]<-ybar[k]*(xbar[k]+rhoyx)/(mux+rhoyx)
t10[k]<-ybar[k]*((mux+beta2x)/(xbar[k]+beta2x))
t11[k]<-ybar[k]*((xbar[k]+beta2x)/(mux+beta2x))
t12[k]<-ybar[k]*((beta2x*mux+beta1x)/(beta2x*xbar[k]+beta1x))
t13[k]<-ybar[k]*((mux+stdevx)/(xbar[k]+stdevx))
t14[k]<-ybar[k]*((mux+md)/(xbar[k]+md))
t15[k]<-ybar[k]*((beta2x*mux+cx)/(beta2x*xbar[k]+cx))
t16[k]<-ybar[k]*((cx*mux+rhoyx)/(cx*xbar[k]+rhoyx))
t17[k]<-ybar[k]*((cx*mux+md)/(cx*xbar[k]+md))
t18[k]<-ybar[k]*((beta1x*mux+rhoyx)/(beta1x*xbar[k]+rhoyx))
t19[k]<-ybar[k]*((rhoyx*mux+stdevx)/(rhoyx*xbar[k]+stdevx))
t20[k]<-ybar[k]*((rhoyx*mux+md)/(rhoyx*xbar[k]+md))
t21[k]<-ybar[k]*((stdevx*mux+md)/(stdevx*xbar[k]+md))
mse.tR[k]<-(tR[k]-muy)^2;mse.tR.E<-mean(mse.tR)
mse.tP[k]<-(tP[k]-muy)^2;mse.tP.E<-mean(mse.tP)
```

```
 mse.tER[k]<-(tER[k]-muy)^2;mse.tER.E<-mean(mse.tER)
 mse.tEP[k]<-(tEP[k]-muy)^2;mse.tEP.E<-mean(mse.tEP)
 mse.t1[k]<-(t1[k]-muy)^2;mse.t1.E<-mean(mse.t1)
 mse.t2[k]<-(t2[k]-muy)^2;mse.t2.E<-mean(mse.t2)
 mse.t3[k]<-(t3[k]-muy)^2;mse.t3.E<-mean(mse.t3)
 mse.t4[k]<-(t4[k]-muy)^2;mse.t4.E<-mean(mse.t4)
 mse.t5[k]<-(t5[k]-muy)^2;mse.t5.E<-mean(mse.t5)
 mse.t6[k]<-(t6[k]-muy)^2;mse.t6.E<-mean(mse.t6)
 mse.t7[k]<-(t7[k]-muy)^2;mse.t7.E<-mean(mse.t7)
 mse.t8[k]<-(t8[k]-muy)^2;mse.t8.E<-mean(mse.t8)
 mse.t9[k]<-(t9[k]-muy)^2;mse.t9.E<-mean(mse.t9)
 mse.t10[k]<-(t10[k]-muy)^2;mse.t10.E<-mean(mse.t10)
 mse.t11[k]<-(t11[k]-muy)^2;mse.t11.E<-mean(mse.t11)
 mse.t12[k]<-(t12[k]-muy)^2;mse.t12.E<-mean(mse.t12)
 mse.t13[k]<-(t13[k]-muy)^2;mse.t13.E<-mean(mse.t13)
 mse.t14[k]<-(t14[k]-muy)^2;mse.t14.E<-mean(mse.t14)
 mse.t15[k]<-(t15[k]-muy)^2;mse.t15.E<-mean(mse.t15)
 mse.t16[k]<-(t16[k]-muy)^2;mse.t16.E<-mean(mse.t16)
 mse.t17[k]<-(t17[k]-muy)^2;mse.t17.E<-mean(mse.t17)
 mse.t18[k]<-(t18[k]-muy)^2;mse.t18.E<-mean(mse.t18)
 mse.t19[k]<-(t19[k]-muy)^2;mse.t19.E<-mean(mse.t19)
 mse.t20[k]<-(t20[k]-muy)^2;mse.t20.E<-mean(mse.t20)
 mse.t21[k]<-(t21[k]-muy)^2;mse.t21.E<-mean(mse.t21)
 }
#THE MEANS OVER 10,000 TRIALS FOR THE RATIO MEAN ESTIMATORS
tR.bar<-mean(tR); tP.bar<-mean(tP); tER.bar<-mean(tER);
```

```
tEP.bar<-mean(tEP); t1.bar<-mean(t1); t2.bar<-mean(t2);
t3.bar<-mean(t3); t4.bar<-mean(t4); t5.bar<-mean(t5);
t6.bar<-mean(t6); t7.bar<-mean(t7); t8.bar<-mean(t8);
t9.bar<-mean(t9); t10.bar<-mean(t10); t11.bar<-mean(t11);
t12.bar<-mean(t12); t13.bar<-mean(t13); t14.bar<-mean(t14);
t15.bar<-mean(t15); t16.bar<-mean(t16); t17.bar<-mean(t17);
t18.bar<-mean(t18); t19.bar<-mean(t19); t20.bar<-mean(t20);
t21.bar<-mean(t21)
mean.E<-cbind(tR.bar,tP.bar,tER.bar,tEP.bar,t1.bar,t2.bar,t3.bar,
t4.bar,t5.bar,t6.bar,t7.bar,t8.bar,t9.bar,t10.bar,t11.bar,t12.bar,
t13.bar, t14.bar,t15.bar,t16.bar,t17.bar,t18.bar,t19.bar,t20.bar,
t21.bar)
#vector with the empirical mean square errors for the mean estimators
mse.emp.table<-c(mse.tR.E, mse.tP.E, mse.tER.E, mse.tEP.E, mse.t1.E,
mse.t2.E, mse.t3.E, mse.t4.E, mse.t5.E, mse.t6.E, mse.t7.E, mse.t8.E,
mse.t9.E, mse.t10.E, mse.t11.E, mse.t12.E, mse.t13.E, mse.t14.E,
mse.t15.E, mse.t16.E, mse.t17.E, mse.t18.E, mse.t19.E, mse.t20.E,
mse.t21.E)

# # # # T H E   T H E O R E T I C A L   P A R T   # # # # #

theta1<-theta2<-mux/(mux+cx)
theta3<-theta15<-(beta2x*mux)/(beta2x*mux+cx)
theta4<-(cx*mux)/(cx*mux+beta2x)
theta5<-theta13<-mux/(mux+stdevx)
theta6<-(beta1x*mux)/(beta1x*mux+stdevx)
```

```
theta7<-(beta2x*mux)/(beta2x*mux+stdevx)
theta8<-theta9<-mux/(mux+rhoyx)
theta10<-theta11<-mux/(mux+beta2x)
theta12<-(beta2x*mux)/(beta2x*mux+beta1x)
theta14<-mux/(mux+md)
theta16<-(cx*mux)/(cx*mux+rhoyx)
theta17<-(cx*mux)/(cx*mux+md)
theta18<-(beta1x*mux)/(beta1x*mux+rhoyx)
theta19<-(rhoyx*mux)/(rhoyx*mux+stdevx)
theta20<-(rhoyx*mux)/(rhoyx*mux+md)
theta21<-(stdevx*mux)/(stdevx*mux+md)
mse.ratio<-function(theta){
  ((1-f)/n)*muy^2*(cy^2+theta*cx^2*(theta-2*co))
}
mse.product<-function(theta){
  ((1-f)/n)*muy^2*(cy^2+theta*cx^2*(theta+2*co))
}
co<-rhoyx*(cy/cx); f<-n/N

# THE THEORETICAL MEAN SQUARE ERRORS FOR  THE MEAN ESTIMATORS #
mse.tr.th<-((1-f)/n)*muy^2*(cx^2-2*rhoyx*cx*cy+cy^2)
mse.tp.th<-((1-f)/n)*muy^2*(cx^2+2*rhoyx*cx*cy+cy^2)
mse.tER.th<-((1-f)/n)*(1/4)*muy^2*(cx^2-4*rhoyx*cx*cy+4*cy^2)
mse.tEP.th<-((1-f)/n)*(1/4)*muy^2*(cx^2+4*rhoyx*cx*cy+4*cy^2)
mse.t1.th<-mse.ratio(theta1); mse.t2.th<-mse.product(theta2)
mse.t3.th<-mse.product(theta3); mse.t4.th<-mse.product(theta4)
```

```
mse.t5.th<-mse.product(theta5); mse.t6.th<-mse.product(theta6)
mse.t7.th<-mse.product(theta7); mse.t8.th<-mse.ratio(theta8)
mse.t9.th<-mse.product(theta9); mse.t10.th<-mse.ratio(theta10)
mse.t11.th<-mse.product(theta11); mse.t12.th<-mse.ratio(theta12)
mse.t13.th<-mse.ratio(theta13); mse.t14.th<-mse.ratio(theta14)
mse.t15.th<-mse.ratio(theta15); mse.t16.th<-mse.ratio(theta16)
mse.t17.th<-mse.ratio(theta17); mse.t18.th<-mse.ratio(theta18)
mse.t19.th<-mse.ratio(theta19); mse.t20.th<-mse.ratio(theta20)
mse.t21.th<-mse.ratio(theta21)
# vector with the theoretical mean square errors for the mean estimators
mse.th.table<-c(mse.tr.th,mse.tp.th,mse.tER.th,mse.tEP.th,mse.t1.th,
mse.t2.th,mse.t3.th,mse.t4.th,mse.t5.th,mse.t6.th,mse.t7.th,mse.t8.th,
mse.t9.th, mse.t10.th, mse.t11.th, mse.t12.th, mse.t13.th, mse.t14.th,
mse.t15.th, mse.t16.th, mse.t17.th, mse.t18.th,mse.t19.th,mse.t20.th,
mse.t21.th)
# the ratio of the mean estimators
mse.ratio<-100*(mse.emp.table/mse.th.table)
# the ratio of the empirical and corresponding theoretical mean
# square errors for the ratio type mean estimators
ratio.mean<-cbind(r.R=mse.ratio[1],r.tER=mse.ratio[3],
r.t1.R=mse.ratio[5],
r.t8.R=mse.ratio[12],r.t10.R=mse.ratio[14],r.t12.R=mse.ratio[16],
r.t13.R=mse.ratio[17],r.t14.R=mse.ratio[18],r.t15.R=mse.ratio[19],
r.t16.R=mse.ratio[20],r.t17.R=mse.ratio[21],r.t18.R=mse.ratio[22],
r.t19.R=mse.ratio[23],r.t20.R=mse.ratio[24],r.t21.R=mse.ratio[25])
# the ratio of the empirical and corresponding theoretical mean
```

```
# square errors for the product type mean estimators
product.mean<-cbind(r.P=mse.ratio[2],r.tEP=mse.ratio[4],
r.t2.P=mse.ratio[6],
r.t3.P=mse.ratio[7],r.t4.P=mse.ratio[8],r.t5.P=mse.ratio[9],
r.t6.P=mse.ratio[10],r.t7.P=mse.ratio[11],r.t9.P=mse.ratio[13],
r.t11.P=mse.ratio[15])
# the summary statistics for the variance part
mu40<-(1/N)*sum((Y-muy)^4); mu20<-(1/N)*sum((Y-muy)^2);
Beta2.y<-mu40/(mu20^2);
mu04<-(1/N)*sum((X-mux)^4); mu02<-(1/N)*sum((X-mux)^2);
Beta2.x<-mu04/(mu02^2)
mu22<-(1/N)*sum((Y-muy)^2*(X-mux)^2); lambda22<-mu22/(mu20*mu02)
q2<-quantile(X,0.5); md<-q2; q1<-quantile(X,0.25,names=FALSE);
q3<-quantile(X,0.75,names=FALSE); qa<-(q3+q1)/2; Lambda<-1/n

 # # # #    T H E    S I M U L A T I O N    P A R T     # # # #

s.x<-numeric(b); s.y<-numeric(b); Sr<-numeric(b); Sexp<-numeric(b);
S1<-numeric(b); S2<-numeric(b); S3<-numeric(b); S4<-numeric(b);
S5<-numeric(b); S6<-numeric(b); mse.Sr<-numeric(b);
mse.Sexp<-numeric(b);mse.S1<-numeric(b); mse.S2<-numeric(b);
mse.S3<-numeric(b);mse.S4<-numeric(b);
mse.S5<-numeric(b); mse.S6<-numeric(b)
for (k in 1:b){s<-sample(N,n,replace=TRUE);
  s.x[k]<-var(X[s])
  s.y[k]<-var(Y[s])
```

```
  Sr[k]<-s.y[k]*(varx/s.x[k]) # Isaki (1983)
  Sexp[k]<-s.y[k]*exp((varx-s.x[k])/(varx+s.x[k]))
  S1[k]<-s.y[k]*((varx+cx)/(s.x[k]+cx))
  S2[k]<-s.y[k]*((varx+beta2x)/(s.x[k]+beta2x))
  S3[k]<-s.y[k]*((varx*beta2x+cx)/(s.x[k]*beta2x+cx))
  S4[k]<-s.y[k]*((varx*cx+beta2x)/(s.x[k]*cx+beta2x))
  S5[k]<-s.y[k]*((varx+qa)/(s.x[k]+qa))
  S6[k]<-s.y[k]*((varx*cx+md)/(s.x[k]*cx+md))#
  mse.Sr[k]<-(Sr[k]-vary)^2;mse.Sr.E<-mean(mse.Sr)
  mse.Sexp[k]<-(Sexp[k]-vary)^2; mse.Sexp.E<-mean(mse.Sexp)
  mse.S1[k]<-(S1[k]-vary)^2; mse.S1.E<-mean(mse.S1)
  mse.S2[k]<-(S2[k]-vary)^2; mse.S2.E<-mean(mse.S2)
  mse.S3[k]<-(S3[k]-vary)^2; mse.S3.E<-mean(mse.S3)
  mse.S4[k]<-(S4[k]-vary)^2; mse.S4.E<-mean(mse.S4)
  mse.S5[k]<-(S5[k]-vary)^2; mse.S5.E<-mean(mse.S5)
  mse.S6[k]<-(S6[k]-vary)^2; mse.S6.E<-mean(mse.S6)
}
# vector contains the empirical mean square errors for the var. est.
mse.var.emp<-c(mse.Sr.E, mse.Sexp.E, mse.S1.E, mse.S2.E, mse.S3.E,
 mse.S4.E,mse.S5.E,mse.S6.E)

# # # #   T H E   T H E O R E T I C A L   P A R T   # # # #
R0<-1
R1<-varx/(varx+cx)
R2<-varx/(varx+beta2x)
R3<-varx*beta2x/(varx*beta2x+cx)
```

```
R4<-varx*cx/(varx*cx+beta2x)
R5<-varx/(varx+qa)
R6<-varx/(varx+md)
mse.var<-function(c){
  Lambda*(vary)^2*((beta2y-1)+c^2*(beta2x-1)-2*c*(lambda22-1))
}
#THE THEORETICAL MEAN SQUARE ERRORS  FOR  FOR THE VARIANCE ESTIMATORS
mse.Sexp.th<-Lambda*(vary)^2*((beta2y-1)+(1/4)*(beta2x-1)
-(lambda22-1))
mse.Sr.th<-mse.var(R0)
mse.S1.th<-mse.var(R1)
mse.S2.th<-mse.var(R2)
mse.S3.th<-mse.var(R3)
mse.S4.th<-mse.var(R4)
mse.S5.th<-mse.var(R5)
mse.S6.th<-mse.var(R6)
# vector with the theoretical mean square errors for the ratio
#type variance estimators
mse.var.th<-c(mse.Sr.th,mse.Sexp.th,mse.S1.th,mse.S2.th,mse.S3.th,
mse.S4.th,mse.S5.th,mse.S6.th)
# the ratio of the empirical and corresponding theoretical mean square
#  errors for the ratio type variance estimators
mse.var.ratio<-100*(mse.var.emp/mse.var.th)
ratio.var<-cbind(r.Sr=mse.var.ratio[1], r.Sexp=mse.var.ratio[2],
r.S1=mse.var.ratio[3], r.S2=mse.var.ratio[4],r.S3=mse.var.ratio[5],
r.S4=mse.var.ratio[6], r.S5=mse.var.ratio[7],r.S6=mse.var.ratio[8])
```

APPENDIX B

R CODE II

```
population.data<-read.table("population.data",header=TRUE)
X<-population.data$X; Y<-population.data$Y

# # #     T H E   S U M M A R Y   S T A T I S T I C S     # # #

N<-nrow(population.data); rhoyx=cor(X,Y); mux<-mean(X); muy<-mean(Y);
varx<-var(X); vary<-var(Y); stdevx<-sd(X); stdevy<-sd(Y) ;
beta1x<-skewness(X); beta1y<-skewness(Y); beta2x<-kurtosis(X);
beta2y<-kurtosis(Y); cx<-stdevx/mux; cy<-stdevy/muy

# # #     T H E   S I M U L A T I O N   P A R T     # # #

xbar<-numeric(b);ybar<-numeric(b);var.samplex<-numeric(b);
muER<-numeric(b);
var.sampley<-numeric(b); rhoyx.sample<-numeric(b); muR<-numeric(b);
muReg<-numeric(b); muSRP<-numeric(b); muRReg<-numeric(b);
muGK<-numeric(b);
muGM1<-numeric(b); muGM2<-numeric(b); muGMR1<-numeric(b);
muGMR2<-numeric(b);
mse.ybar<-numeric(b); mse.muER<-numeric(b); mse.muR<-numeric(b);
mse.muReg<-numeric(b); mse.muSRP<-numeric(b);mse.muRReg<-numeric(b);
mse.muGK<-numeric(b); mse.muGM1<-numeric(b);mse.muGM2<-numeric(b)
mse.muGMR1<-numeric(b);mse.muGMR2<-numeric(b); sd.samplex<-numeric(b);
```

```
sd.sampley<-numeric(b);sd.samplexy<-numeric(b);beta.sample<-numeric(b);
alpha.opt<-numeric(b); cy.sample<-numeric(b); k1opt<-numeric(b);
k2opt<-numeric(b); l1opt<-numeric(b); l2optn<-numeric(b);
 l2optd<-numeric(b);
l2opt<-numeric(b); d1optGM1<-numeric(b); d2optGM1<-numeric(b);
d1optGM2<-numeric(b); d2optGM2<-numeric(b); k1optGMR1<-numeric(b);
k2optGMR1<-numeric(b); k1optGMR2<-numeric(b); k2optGMR2<-numeric(b);
n<-100;b<-10000; f<-n/N ; lambda<-(1-f)/n # n=100, 200 and 500
for (k in 1:b){s<-sample(N,n,replace=TRUE); # print(s)
xbar[k]<-mean(X[s]); ybar[k]<-mean(Y[s]); rhoyx.sample[k]<-cor(X[s],Y[s]);
var.samplex[k]<-var(X[s]);var.sampley[k]<-var(Y[s])
sd.sampley[k]<-sd(Y[s]); sd.samplexy[k]<-cov(X[s],Y[s])
cy.sample[k]<-sd.sampley[k]/ybar[k]
beta.sample[k]<-sd.samplexy[k]/var.samplex[k]
alpha.opt[k]<-(0.5)+(rhoyx.sample[k]*(cy.sample[k]/cx))
k1opt[k]<-1/(1+lambda*(1-rhoyx.sample[k]^2)*cy.sample[k]^2)
k2opt[k]<-(ybar[k]/xbar[k])*((rhoyx.sample[k]*cy.sample[k])/cx)
l1opt[k]<-((-8)+lambda*cx^2)(8*(-1+lambda*(rhoyx.sample[k]^2-1)
*cy.sample[k]^2))
l2optn[k]<-ybar[k]*((-8)*rhoyx.sample[k]*cy.sample[k]+cx*(4-lambda*cx^2
-lambda*rhoyx.sample[k]*cx*cy.sample[k]+4*lambda*(rhoyx.sample[k]^2-1)
*cy.sample[k]^2))
l2optd[k]<-8*mux*cx*(-1+lambda*(rhoyx.sample[k]^2-1)*cy.sample[k]^2)
l2opt[k]<-l2optn[k]/l2optd[k]
d1optGM1[k]<-(1+(3/8)*lambda*cx^2)/(1+lambda*cy.sample[k]^2*
(1-rhoyx.sample[k]^2)+lambda*cx^2)
```

```
d2optGM1[k]<-(ybar[k]/xbar[k])*(0.5-d1optGM1[k]*(1-rhoyx.sample[k]
*(cy.sample[k]/cx)))
d1optGM2[k]<-(1+(7/8)*lambda*cx^2)/(1+lambda*(1-rhoyx.sample[k]^2)
*cy.sample[k]^2+2*lambda*cx^2)
d2optGM2[k]<-(ybar[k]/xbar[k])*(0.5-d1optGM2[k]*(1-rhoyx.sample[k]
*(cy.sample[k]/cx)))
k1optGMR1[k]<-(1+(7/8)*lambda*cx^2-lambda*rhoyx.sample[k]*cx*
cy.sample[k])/(1+lambda*cy.sample[k]^2*(1-rhoyx.sample[k]^2)
+2*lambda*cx^2-2*lambda*rhoyx.sample[k]*cx*cy.sample[k])
k2optGMR1[k]<-(ybar[k]/xbar[k])*(0.5-k1optGMR1[k]*(2-rhoyx.sample[k]
*(cy.sample[k]/cx)))
k1optGMR2[k]<-(1+(23/8)*lambda*(cx^2-2*rhoyx.sample[k]*cx*cy.sample[k]))/
(1+lambda*(cy.sample[k]^2*(1-rhoyx.sample[k]^2)+6*cx^2-4*rhoyx.sample[k]
*cx*cy.sample[k]))
k2optGMR2[k]<-(ybar[k]/xbar[k])*(0.5-k1optGMR2[k]*(3-rhoyx.sample[k]
*(cy.sample[k]/cx)))
# The non-RRT Mean Estimators
muR[k]<-ybar[k]*(mux/xbar[k])
muReg[k]<-ybar[k]+beta.sample[k]*(mux-xbar[k])
muER[k]<-ybar[k]*exp((mux-xbar[k])/(mux+xbar[k]))
muSRP[k]<-ybar[k]*(alpha.opt[k]*exp((mux-xbar[k])/(mux+xbar[k]))
+(1-alpha.opt[k])*exp((xbar[k]-mux)/(xbar[k]+mux)))
muRReg[k]<-k1opt[k]*ybar[k]+k2opt[k]*(mux-xbar[k])
muGK[k]<-(l1opt[k]*ybar[k]+l2opt[k]*(mux-xbar[k]))*exp((mux-xbar[k])/
(mux+xbar[k])) # Grover & Kaur (2011)
muGM1[k]<-(d1optGM1[k]*ybar[k]*(0.5*(mux/xbar[k]+xbar[k]/mux))+
```

```
d2optGM1[k] *(mux-xbar[k]))*exp((mux-xbar[k])/(mux+xbar[k]))
muGM2[k]<-(d1optGM2[k]*ybar[k]*(0.5*(mux/xbar[k]+xbar[k]/mux))^2
+d2optGM2[k]*(mux-xbar[k]))*exp((mux-xbar[k])/(mux+xbar[k]))
muGMR1[k]<-(k1optGMR1[k]*ybar[k]*(mux/xbar[k])+k2optGMR1[k]
*(mux-xbar[k]))*exp((mux-xbar[k])/(mux+xbar[k]))
muGMR2[k]<-(k1optGMR2[k]*ybar[k]*(mux/xbar[k])^2+k2optGMR2[k]
*(mux-xbar[k]))*exp((mux-xbar[k])/(mux+xbar[k]))
# the empirical mean square errors  over 10,000 trials
mse.ybar[k]<-(ybar[k]-muy)^2; mse.ybar.E<-mean(mse.ybar)
mse.muR[k]<-(muR[k]-muy)^2; mse.muR.E<-mean(mse.muR)
mse.muER[k]<-(muER[k]-muy)^2; mse.muER.E<-mean(mse.muER)
mse.muReg[k]<-(muReg[k]-muy)^2; mse.muReg.E<-mean(mse.muReg)
mse.muSRP[k]<-(muSRP[k]-muy)^2; mse.muSRP.E<-mean(mse.muSRP)
mse.muRReg[k]<-(muRReg[k]-muy)^2; mse.muRReg.E<-mean(mse.muRReg)
mse.muGK[k]<-(muGK[k]-muy)^2; mse.muGK.E<-mean(mse.muGK)
mse.muGM1[k]<-(muGM1[k]-muy)^2; mse.muGM1.E<-mean(mse.muGM1)
mse.muGM2[k]<-(muGM2[k]-muy)^2; mse.muGM2.E<-mean(mse.muGM2)
mse.muGMR1[k]<-(muGMR1[k]-muy)^2; mse.muGMR1.E<-mean(mse.muGMR1)
mse.muGMR2[k]<-(muGMR2[k]-muy)^2; mse.muGMR2.E<-mean(mse.muGMR2)
}
# The Empirical Mean Square Error for the non-RRT Mean Estimators
 mse.E<-cbind(ybar=mse.ybar.E,muR=mse.muR.E,muReg=mse.muReg.E,
muER=mse.muER.E,muSRP=mse.muSRP.E,muRReg=mse.muRReg.E,muGK=mse.muGK.E,
muGM11=mse.muGM1.E,muGM2=mse.muGM2.E,muGMR1=mse.muGMR1.E,
muGMR2=mse.muGMR2.E)
# The Empirical Percent Relative Efficiency for the Mean estimators
```

```
pre.E<-100*(mse.ybar.E/mse.E)

# # # #      T H E   T H E O R E T I C A L   P A R T      # # # # #

mse.ybar<-lambda*muy^2*cy^2
mse.muR.th<-lambda*muy^2*(cy^2+cx^2-2*rhoyx*cy*cx)
mse.muReg.th<-lambda*muy^2*cy^2*(1-rhoyx^2)
mse.muER.th<-lambda*muy^2*(cy^2+(1/4)*cx^2-rhoyx*cx*cy)
mse.muSRP.th<-lambda*muy^2*(1-rhoyx^2)*cy^2
mse.muRReg.th<-muy^2*(1+(1/(-1+lambda*(rhoyx^2-1)*cy^2)))
tgkn<-lambda*muy^2*(lambda*(cx^4)-16*(rhoyx^2-1)*(lambda*(cx^2)-4)*cy^2)
tgkd<-64*(-1+lambda*(rhoyx^2-1)*cy^2)
mse.muGK.th<-tgkn/tgkd
mse.muGM1.th<-muy^2*((1-0.25*lambda*cx^2)-((1+(3/8)*lambda*cx^2)^2/
(1+lambda*cy^2*(1-rhoyx^2)+lambda*cx^2)))
mse.muGM2.th<-muy^2*((1-0.25*lambda*cx^2)-((1+(7/8)*lambda*cx^2)^2/
(1+lambda*cy^2*(1-rhoyx^2)+2*lambda*cx^2)))
mse.muGMR1.th<-muy^2*((1-0.25*lambda*cx^2)-(1+lambda*((7/8)*cx^2
-rhoyx*cx*cy))^2/(1+lambda*(2*cx^2+(1-rhoyx^2)*cy^2-2*rhoyx*cx*cy)))
mse.muGMR2.th<-muy^2*((1-0.25*lambda*cx^2)-(1+lambda*((23/8)*cx^2
-2*rhoyx*cx*cy))^2/(1+lambda*(6*cx^2+(1-rhoyx^2)*cy^2-4*rhoyx*cx*cy)))
# The Theoretical Mean Square Errors for the non-RRT Mean Estimators
mse.th.tb<-cbind(ybar=mse.ybar,muR.t=mse.muR.th, muReg.t=mse.muReg.th,
muER.t=mse.muER.th, muSRP.t=mse.muSRP.th, muRReg.t=mse.muRReg.th,
muGK=mse.muGK.th,muGM1=mse.muGM1.th,muGM2=mse.muGM2.th,
muGMR1=mse.muGMR1.th,muGMR2=mse.muGMR2.th)
```

```
#The Theort. Per. Rel. Eff. for the non-RRT Mean Estimators
pre.th.table<-100*(mse.ybar/mse.th.tb)

# R A N D O M I Z E D   R E S P O N S E   T E C H N I Q U E  #

population.data<-read.table("population.data",header=TRUE)
scrambled.data<-read.table("scrambled.data",header=TRUE)
X<-population.data$X; Y<-population.data$Y; S<-scrambled.data$S
Z=Y+S
N<-nrow(population.data); rhozx=cor(X,Z); rhoyx=cor(X,Y)
mux<-mean(X); muy<-mean(Y); muz<-mean(Z); varx<-var(X); vary<-var(Y);
vars<-var(S); stdevx<-sd(X); stdevy<-sd(Y); stdevz<-sd(Z);
cx<-stdevx/mux;  tcy<-stdevy/muy; cz<-stdevz/muz
beta1x<-skewness(X);beta2x<-kurtosis(X)

# # #  T H E  S I M U L A T I O N  P A R T   # # #

n<-100; b<-10000 # n=100, 200 and 500
f<-n/N; lambda<-(1-f)/n
xbar<-numeric(b); zbar<-numeric(b); rhozx.sample<-numeric(b);
var.samplex<-numeric(b);var.samplez<-numeric(b); sd.samplex<-numeric(b);
sd.samplez<-numeric(b);sd.samplezx<-numeric(b); cz.sample<-numeric(b);
beta.sample.zx<-numeric(b); muR<-numeric(b); muTR1<-numeric(b);
muTR2<-numeric(b); muTR3<-numeric(b); muTR4<-numeric(b);
muERR<-numeric(b); muReg<-numeric(b); muGRR<-numeric(b);
muGER<-numeric(b);muGRR1<-numeric(b);muGRR2<-numeric(b);
```

```
mse.zbar<-numeric(b);mse.muR<-numeric(b); mse.muTR1<-numeric(b);
mse.muTR2<-numeric(b);mse.muTR3<-numeric(b);mse.muTR4<-numeric(b);
mse.muERR<-numeric(b); mse.muReg<-numeric(b);mse.muGRR<-numeric(b);
mse.muGER<-numeric(b);mse.muGRR1<-numeric(b);mse.muGRR2<-numeric(b);
w1opt<-numeric(b); w2opt<-numeric(b);k1opt<-numeric(b);
k2opt<-numeric(b);d1optGRR1<-numeric(b); d2optGRR1<-numeric(b);
d1optGRR2<-numeric(b); d2optGRR2<-numeric(b);
for (k in 1:b){s<-sample(N,n,replace=TRUE); #print(s)
xbar[k]<-mean(X[s]); zbar[k]<-mean(Z[s]); var.samplez[k]<-var(Z[s])
var.samplex[k]<-var(X[s]); sd.samplez[k]<-sd(Z[s]);
sd.samplezx[k]<-cov(X[s],Z[s]);cz.sample[k]<-sd.samplez[k]/zbar[k];
beta.sample.zx[k]<-sd.samplezx[k]/var.samplex[k]
k1opt[k]<-(1-lambda*cx^2)/(1-lambda*(cx^2-cz.sample[k]^2*(1
-rhozx.sample[k]^2)))
k2opt[k]<-(zbar[k]/xbar[k])*(1+k1opt[k]*((rhozx.sample[k]
*(cz.sample[k]/cx))-2))
w1opt[k]<-(1-lambda*(1/8)*cx^2)/(1+lambda*cz.sample[k]^2
*(1-rhozx.sample[k]^2))
w2opt[k]<-(zbar[k]/xbar[k])*(0.5-w1opt[k]*(1-rhozx.sample[k]
*(cz.sample[k]/cx)))
d1optGRR1[k]<-(1+(7/8)*lambda*cx^2-lambda*rhozx.sample[k]*cx
*cz.sample[k])/(1+lambda*cz.sample[k]^2*(1-rhozx.sample[k]^2)
+2*lambda*cx^2-2*lambda*rhozx.sample[k]*cx*cz.sample[k])
d2optGRR1[k]<-(zbar[k]/xbar[k])*(0.5-d1optGRR1[k]*(2-rhozx.sample[k]
*(cz.sample[k]/cx)))
d1optGRR2[k]<-(1+(23/8)*lambda*cx^2-2*lambda*rhozx.sample[k]*cx
```

```
*cz.sample[k])/(1+lambda*cz.sample[k]^2*(1-rhozx.sample[k]^2)+
6*lambda*cx^2-4*lambda*rhozx.sample[k]*cx*cz.sample[k])
d2optGRR2[k]<-(zbar[k]/xbar[k])*(0.5-d1optGRR2[k]*(3-rhozx.sample[k]
*(cz.sample[k]/cx)))
# RRT Mean Estimators
muR[k]<-zbar[k]*(mux/xbar[k])
muTR1[k]<-zbar[k]*((mux+beta1x)/(xbar[k]+beta1x))
muTR2[k]<-zbar[k]*((mux+beta2x)/(xbar[k]+beta2x))
muTR3[k]<-zbar[k]*((beta1x*mux+beta2x)/(beta1x*xbar[k]+beta2x))
muTR4[k]<-zbar[k]*((beta2x*mux+beta1x)/(beta2x*xbar[k]+beta1x))
muReg[k]<-zbar[k]+beta.sample.zx[k]*(mux-xbar[k])
muERR[k]<-zbar[k]*exp((mux-xbar[k])/(mux+xbar[k]))
muGRR[k]<-(k1opt[k]*zbar[k]+k2opt[k]*(mux-xbar[k]))*(mux/xbar[k])
muGER[k]<-(w1opt[k]*zbar[k]+w2opt[k]*(mux-xbar[k]))*exp((mux-xbar[k])/
(mux+xbar[k]))
muGRR1[k]<-(d1optGRR1[k]*zbar[k]*(mux/xbar[k])+d2optGRR1[k]*(mux
-xbar[k]))*exp((mux-xbar[k])/(mux+xbar[k])) # Zatezalo et al.(2016)
muGRR2[k]<-(d1optGRR2[k]*zbar[k]*(mux/xbar[k])^2+d2optGRR2[k]*(mux
-xbar[k]))*exp((mux-xbar[k])/(mux+xbar[k]))
mse.zbar[k]<-(zbar[k]-muy)^2; mse.zbar.E<-mean(mse.zbar)
mse.muR[k]<-(muR[k]-muy)^2; mse.muR.E<-mean(mse.muR)
mse.muTR1[k]<-(muTR1[k]-muy)^2; mse.muTR1.E<-mean(mse.muTR1)
mse.muTR2[k]<-(muTR2[k]-muy)^2; mse.muTR2.E<-mean(mse.muTR2)
mse.muTR3[k]<-(muTR3[k]-muy)^2; mse.muTR3.E<-mean(mse.muTR3)
mse.muTR4[k]<-(muTR4[k]-muy)^2; mse.muTR4.E<-mean(mse.muTR4)
mse.muReg[k]<-(muReg[k]-muy)^2; mse.muReg.E<-mean(mse.muReg)
```

```
mse.muERR[k]<-(muER[k]-muy)^2; mse.muERR.E<-mean(mse.muERR)
mse.muGRR[k]<-(muGRR[k]-muy)^2;  mse.muGRR.E<-mean(mse.muGRR)
mse.muGER[k]<-(muGER[k]-muy)^2; mse.muGER.E<-mean(mse.muGER)
mse.muGRR1[k]<-(muGRR1[k]-muy)^2; mse.muGRR1.E<-mean(mse.muGRR1)
mse.muGRR2[k]<-(muGRR2[k]-muy)^2;mse.muGRR2.E<-mean(mse.muGRR2)
}
#The Empirical Mean Square Errors for the  RRT  Mean Estimators
mse.E<-cbind(zbar=mse.zbar.E,muR=mse.muR.E,muTR1=mse.muTR1.E,
muTR2=mse.muTR2.E,muTR3=mse.muTR3.E,muTR4=mse.muTR4.E,
muReg=mse.muReg.E,muER=mse.muER.E,muGRR=mse.muGRR.E,
muGER=mse.muGER.E,muGRR1=mse.muGRR1.E,muGRR2=mse.muGRR2.E)
# The Emp. Perc. Rel. Efficiency for the RRT Mean Estimators
pre.E<-100*(mse.zbar.E/mse.E)

#   T H E     T H E O R E T I C A L    P A R T      #

mse.zbar.th<-lambda*(vary+vars)
mse.muR.th<-lambda*muy^2*(cz^2+cx^2-2*rhozx*cz*cx)
mse.muTR1.th<-lambda*muy^2*(cz^2+(mux/(mux+beta1x))^2*cx^2
-2*(mux/(mux+beta1x))*rhozx*cz*cx)
mse.muTR2.th<-lambda*muy^2*(cz^2+(mux/(mux+beta2x))^2*cx^2
-2*(mux/(mux+beta2x))*rhozx*cz*cx)
mse.muTR3.th<-lambda*muy^2*(cz^2+((beta1x*mux)/(beta1x*mux+beta2x))^2
*cx^2-2*((beta1x*mux)/(beta1x*mux+beta2x))*rhozx*cz*cx)
mse.muTR4.th<-lambda*muy^2*(cz^2+((beta2x*mux)/(beta2x*mux+beta1x))^2
*cx^2-2*((beta2x*mux)/(beta2x*mux+beta1x))*rhozx*cz*cx)
```

```
mse.muReg.th<-lambda*muy^2*cz^2*(1-rhozx^2)
mse.muER.th<-lambda*muy^2*(cz^2-rhozx*cz*cx+0.25*cx^2)
mse.muGRR.th<-muy^2*(lambda*cz^2*(1-rhozx^2)*(1-lambda*cx^2))/((lambda
*cz^2*(1-rhozx^2)+(1-lambda*cx^2)))
mse.muGER.th<-muy^2*((1-0.25*lambda*cx^2)-((1-lambda*(1/8)*cx^2)^2/
(1+lambda*(1-rhozx^2)*cz^2)))
mse.muGRR1.th<-muy^2*((1-0.25*lambda*cx^2)-((1+lambda*((7/8)*cx^2
-rhozx*cz*cx))^2/(1+lambda*(2*cx^2+(1-rhozx^2)*cz^2-2*rhozx*cz*cx))))
mse.muGRR2.th<-muy^2*((1-0.25*lambda*cx^2)-((1+lambda*((23/8)*cx^2
-2*rhozx*cz*cx))^2/(1+lambda*(6*cx^2+(1-rhozx^2)*cz^2-4*rhozx*cz*cx))))
# the theoretical mean square for the RRT mean estimators
mse.th<-cbind(zbar.th=mse.zbar.th,muR.th=mse.muR.th,
muTR1.th=mse.muTR1.th,muTR2.th=mse.muTR2.th,muTR3.th=mse.muTR3.th,
muTR4.th=mse.muTR4.th,muReg.th=mse.muReg.th,muER.th=mse.muER.th,
muGRR.th=mse.muGRR.th,muGER.th=mse.muGER.th,muGRR1.th=mse.muGRR1.th,
muGRR2.th=mse.muGRR2.th)
# The Theor. Perc. Rel. Eff. for the RRT Mean Estimators
pre.th.table<-100*(mse.zbar.th/mse.th)
```