

ZHANG, QI, M.A. Ratio Estimation of the Mean under RRT Models. (2016)
Directed by Dr. Sat Gupta. 70 pp.

Ratio estimation is a parameter estimation technique that uses a known auxiliary variable that is correlated with the study variable. In many situations, the primary variable of interest may be sensitive and it cannot be observed directly. However, we can observe directly a non-sensitive variable that is highly correlated with the study variable. In these cases, we have to rely on some Randomized Response Technique (RRT) models to obtain information on the study variable.

In this thesis, we first review some RRT models, some general ratio and product estimation techniques, and two Kalucha et al. (2015) ratio estimators that are based on Gupta et al. (2010) additive optional RRT model. One of the Kalucha et al. (2015) estimators, the multiplicative ratio estimator, did not work efficiently and was abandoned. The main focus of this thesis is on fixing the Kalucha et al. (2015) abandoned multiplicative ratio estimator and reevaluating its performance. We discuss the Bias and the Mean Square Error (MSE) of our proposed multiplicative ratio estimator correct up to first order approximation, and present the comparisons with other estimators under the additive optional RRT model. A simulation study is also conducted to verify the theoretical result. Both the theoretical and the empirical results show that the proposed multiplicative ratio estimator is more efficient than the ordinary RRT estimator that does not utilize the auxiliary variable. It also compares well with the additive ratio estimator of Kalucha et al. (2015).

RATIO ESTIMATION OF THE MEAN UNDER RRT MODELS

by

Qi Zhang

A Thesis 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
Master of Arts

Greensboro
2016

Approved by

Committee Chair

To my parents, for their support and encouragement.

APPROVAL PAGE

This thesis written by Qi Zhang has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____
Sat Gupta

Committee Members _____
Greg Bell

Scott Richter

Haimeng Zhang

Date of Acceptance by Committee

Date of Final Oral Examination

ACKNOWLEDGMENTS

This thesis would not have been possible without the help of so many people in so many ways.

Most of all, I would like to thank my mentor, Dr. Sat Gupta, for motivating me to work on this topic, helping me to deeply understand and giving substantial advice regarding the topic of my research. I would also like to express my appreciations for his understanding, patience, encouragement and for pushing me further than I thought I could go. Dr. Gupta is an excellent mentor, he always inspired me with a desire to learn. He was influential in guiding me, while boosting my confidence as well. It was a pleasure working with him.

Next I would like to thank Dr. Haimeng Zhang, for helping me to improve my programming skills. He has always been very patient, kind, and knowledgeable about the topics at hand, which I will take with me in the future.

I would like to thank Dr. Scott Richter and Dr. Greg Bell for the assistance and direction they have provided me.

I would also like to express my gratitude to all my professors throughout my college career at UNCG who have assisted me in becoming a better statistician. Also, I am appreciative of the Department of Mathematics and Statistics which provided me an opportunity to obtain my Master's degree.

Last but not least, I would like to show my sincere appreciation to all my friends and family for helping me survive all the stress, and not letting me give up.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vii
CHAPTER	
I. INTRODUCTION	1
1.1. Randomized Response Technique (RRT) Models	1
1.2. Ordinary Ratio and Product Estimators	11
1.3. Improvement in RRT Mean Estimation using Auxiliary Information	14
1.4. Thesis Outline	16
II. RATIO ESTIMATORS OF MEAN WHEN THE STUDY VARIABLE IS SENSITIVE	18
2.1. Sousa et al. (2010): Ratio Estimator under RRT Models	18
2.2. Kalucha et al. (2015): Ratio Estimators under Optional RRT Models	20
2.3. Simulation Results	24
2.4. Concluding Chapter Remarks	31
III. IMPROVED MULTIPLICATIVE RATIO ESTIMATOR OF MEAN	35
3.1. Introduction	35
3.2. The Proposed Multiplicative Ratio Estimator	37
3.3. Efficiency Comparisons	40
3.4. Concluding Chapter Remarks	48
IV. SIMULATIONS	49
4.1. Simulation with Equal Sub-samples	50
4.2. Simulations with Unequal Sub-samples	54
V. DISCUSSION	65
5.1. General Discussion	65
5.2. Future Directions	67

REFERENCES 68

LIST OF TABLES

	Page
Table 1. Theoretical (bold) and Empirical MSE Comparisons with Estimates of W , μ_y and μ_{AR} for $\rho_{xy} = 0.3$ and $\rho_{xy} = 0.8$ using Equal Sample Split ($\theta_1 = 5, \theta_2 = 0.5, n_1 = n_2$)	27
Table 2. Theoretical (bold) and Empirical MSE Comparisons with Estimates of W , μ_y and μ_{AR} for $\rho_{xy} = 0.3$ and $\rho_{xy} = 0.8$ using Unequal Sample Split ($\theta_1 = 0.5, \theta_2 = 5, n_1 < n_2$)	28
Table 3. Theoretical (bold) and Empirical MSE Comparisons with Estimates of W , μ_y and μ_{AR} for $\rho_{xy} = 0.3$ and $\rho_{xy} = 0.8$ using Unequal Sample Split ($\theta_1 = -5, \theta_2 = -0.5, n_1 > n_2$)	29
Table 4. Theoretical (bold) and Empirical MSE Comparisons of μ_y, μ_{AR} and μ_{MR} for Correlations $\rho_{xy} = 0.2$ and $\rho_{xy} = 0.7$ using Equal Sample Split ($\theta_1 = 6, \theta_2 = 0.6, n_1 = n_2$)	32
Table 5. Theoretical (bold) and Empirical MSE Comparisons of μ_y, μ_{AR} and μ_{MR} for Correlations $\rho_{xy} = 0.2$ and $\rho_{xy} = 0.7$ using Unequal Sample Split ($\theta_1 = 0.6, \theta_2 = 6, n_1 < n_2$)	33
Table 6. Theoretical (bold) and Empirical MSE Comparisons of μ_y, μ_{AR} and μ_{MR} for $\rho_{xy} = 0.2$ and $\rho_{xy} = 0.7$ using Unequal Sample Split ($\theta_1 = -6, \theta_2 = -0.6, n_1 > n_2$)	34
Table 7. Point Estimates of $W, \mu_y, \mu_{AR}, \mu_{MR}$ and μ_{GMR} , for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using Equal Sample Split ($\theta_1 = 6, \theta_2 = 0.6, n_1 = n_2$)	52
Table 8. Theoretical (bold) and Empirical MSE Comparisons of $\mu_y, \mu_{AR}, \mu_{MR}$ and μ_{GMR} for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using Equal Sample Split ($\theta_1 = 6, \theta_2 = 0.6, n_1 = n_2$)	53
Table 9. Point Estimates of $W, \mu_y, \mu_{AR}, \mu_{MR}$ and μ_{GMR} , for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using Unequal Sample Split ($\theta_1 = 0.6, \theta_2 = 6, n_1 < n_2$)	57

Table 10. Theoretical (bold) and Empirical MSE Comparisons of μ_y , μ_{AR} , μ_{MR} and μ_{GMR} for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using Unequal Sample Split ($\theta_1 = 0.6$, $\theta_2 = 6$, $n_1 < n_2$).	58
Table 11. Point Estimates of W , μ_y , μ_{AR} , μ_{MR} and μ_{GMR} , for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using Unequal Sample Split ($\theta_1 = -6$, $\theta_2 = -0.6$, $n_1 > n_2$).	59
Table 12. Theoretical (bold) and Empirical MSE Comparisons of μ_y , μ_{AR} , μ_{MR} and μ_{GMR} for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using Unequal Sample Split ($\theta_1 = -6$, $\theta_2 = -0.6$, $n_1 > n_2$).	60
Table 13. Point Estimates of W , μ_y , μ_{AR} , μ_{MR} and μ_{GMR} , for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using Unequal Sample Split ($\theta_1 = -6$, $\theta_2 = 0.6$, $n_1 > n_2$).	61
Table 14. Theoretical (bold) and Empirical MSE Comparisons of μ_y , μ_{AR} , μ_{MR} and μ_{GMR} for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using Unequal Sample Split ($\theta_1 = -6$, $\theta_2 = 0.6$, $n_1 > n_2$).	62
Table 15. Point Estimates of W , μ_y , μ_{AR} , μ_{MR} , and μ_{GMR} for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using Unequal Sample Split ($\theta_1 = 6$, $\theta_2 = 0$, $n_1 > n_2$).	63
Table 16. Theoretical (bold) and Empirical MSE Comparisons of μ_y , μ_{AR} , μ_{MR} and μ_{GMR} for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using Unequal Sample Split ($\theta_1 = 6$, $\theta_2 = 0$, $n_1 > n_2$).	64

CHAPTER I

INTRODUCTION

Our work in this thesis will use two main survey techniques - Randomized Response Technique (RRT) models and Ratio Estimation. We describe these in the first two sections.

1.1 Randomized Response Technique (RRT) Models

In survey sampling, we try to use a random sample to estimate population parameters, but sometimes the sample does not represent the population accurately. In other words, a sample may under- or over-estimate a population parameter due to several sampling or non-sampling errors. Sampling error is an error caused by working with a part of the population and not the whole population. Most of the time it can be reduced by increasing the sample size. Non-sampling errors can be attributed to several problems including nonresponse, respondent mistakes, etc. Nonresponse is the result when individuals chosen for the sample are unable or unwilling to participate in the survey. This error usually happens in mail-in or phone surveys.

Personal interviews are a way to increase response rate but it is easy to cause another non-sampling error - social desirability response bias. For example, if a survey question asks "Have you ever used illegal drugs?" or "How many sexual partners have you had in the past 3 months?", most people like to present themselves in a socially desirable light, therefore their response may be biased toward what they feel is socially desirable. Social Desirability Bias (SDB) is one of the most common problems that affects survey results by under-reporting some bad behaviors or over-reporting good

behaviors. It usually happens when survey questions contain personal or sensitive questions, such as drug use, sexual behavior, etc. People have worked over the years on how to mitigate this effect. In 1985, Anton Nederhof [20] summarized a list of techniques that help researchers reduce the effect of SDB. The seven main strategies he mentioned are: forced-choice items, neutral questions, randomized response technique, self-administered questionnaires, selection interviewers, and proxy subjects. The application of these strategies is not limited to one strategy. Sometimes one can use a combination of several strategies. This is determined by the specific situation.

Randomized response technique (RRT) is an important strategy to prevent or reduce SDB and is widely used in survey interviews. The first RRT model was proposed by Warner in 1965 [27], it was modified later by many researchers including Greenberg et al. (1969) [9], Eichhorn (1983) [6], Gupta et al. (2002) [11], Gupta et al. (2010) [13], Sousa et al. (2010) [24], and Mehta et al. (2012) [18] etc. It allows respondents to answer sensitive questions more comfortably and provide more accurate responses. RRT models have been used in many field surveys, such as Kerkvliet et al. (1994) [16], Gill et al. (2013) [8], Chhabra et al. (2016) [2], Chen et al. (2014) [29], and Geng et al. (2016) [7]. Several Randomized Response Technique (RRT) models will be described in detail in this thesis, but an additive optional RRT model will be the main focus.

1.1.1 Warner's Binary RRT Model (1965)

In 1965, Warner [27] proposed the first Binary Randomized Response Technique model to estimate the prevalence of a particular sensitive characteristic in a population. It increases response rate, and also makes the respondent feel more comfortable

to answer survey questions truthfully and reduces SDB. Warner's Binary RRT model will be illustrated below by an example.

Suppose we are interested in estimating what proportion of college students have a sexually transmitted disease (STD). A randomization device, for instance, a deck of cards that contains two questions (or statements), will be used in this survey to divide the sample into two groups. The two statements we are using are:

- (1) I have been told by a healthcare professional that I have an STD.
- (2) I have never been told by a healthcare professional that I have an STD.

A known proportion p of the cards in the deck contain Statement 1, and the remaining cards contain Statement 2. A simple random sample of n people is drawn from a population, and each subject is asked to pick a card from the deck and provide "yes" or "no" response to the statement on the card. Among these n subjects, let there be n_1 respondents who answer "yes". A "yes" response does not mean this person has an STD; there is another possibility that the person picked the second statement. The same is true for a "no" response. In this case, the interviewer does not know which statement the respondent picked. And the respondent is more likely to give a true answer since their privacy is assured.

Let π be the true probability of a subject having an STD in the population, and p_y be the probability of "yes" responses. Therefore,

$$p_y = p\pi + (1 - p)(1 - \pi). \tag{1.1}$$

The estimate of π is then given as

$$\hat{\pi} = \frac{\hat{p}_y - (1 - p)}{2p - 1} = \frac{\frac{n_1}{n} - (1 - p)}{2p - 1}. \quad (1.2)$$

The variance of this estimator is given by

$$Var(\hat{\pi}) = \frac{\pi(1 - \pi)}{n} + \frac{p(1 - p)}{n(2p - 1)^2}. \quad (1.3)$$

In order to minimize the variance, a large sample size n should be chosen and the proportion (p) of Statement 1 should be close to 0 or 1.

1.1.2 Warner's Quantitative RRT Model (1971)

The estimator proposed in 1965 can be used to estimate binary variables, but many times the question of interest is a quantitative one. Warner [28] modified the RRT model for quantitative cases in 1971. We use another example to illustrate this model.

Suppose we are interested in estimating how many sexual partners a college student had in the last 3 months. Instead of creating cards with two questions in the deck, we make cards with random numbers which are from a pre-assigned distribution. The respondents are asked to pick a card, and add the number on the card to their true answer. Then they report a number, which is the sum of the true answer and the random number they picked. Let X be the sensitive variable with unknown mean μ_x and unknown variance σ_x^2 , and S be the scrambling variable (independent of X) with known mean θ and known variance σ_s^2 . Also let Z be the reported response.

Then

$$Z = X + S. \quad (1.4)$$

The expected response is given by

$$E(Z) = E(X) + E(S). \quad (1.5)$$

This leads to an unbiased estimator of the mean of the sensitive variable X given by

$$\hat{\mu}_x = \bar{Z} - \theta. \quad (1.6)$$

The variance of $\hat{\mu}_x$ is given by

$$Var(\hat{\mu}_x) = Var(\bar{Z}) = \frac{Var(Z)}{n} = \frac{\sigma_x^2 + \sigma_s^2}{n} = \frac{\sigma_x^2}{n} + \frac{\sigma_s^2}{n}, \quad (1.7)$$

where $\frac{\sigma_s^2}{n}$ is the penalty for using the RRT model.

1.1.3 Eichhorn and Hayre's Multiplicative RRT Model (1983)

Eichhorn and Hayre [6] introduced a multiplicative RRT model. Instead of adding a random number to the true response, the respondent needs to multiply the true response by a randomly selected number from a known distribution and divided by the mean of the scrambling variable.

The reported response is given by $Z = XS/\theta$, where $\theta = E(S)$. This leads to the unbiased estimator

$$\hat{\mu}_x = \bar{Z}, \quad (1.8)$$

where \bar{Z} is the sample mean of the reported responses. The variance of $\hat{\mu}_x$ is given by

$$Var(\hat{\mu}_x) = \frac{1}{n} \left[\sigma_x^2 + \frac{\sigma_S^2(\sigma_x^2 + \mu_x^2)}{\theta^2} \right]. \quad (1.9)$$

1.1.4 Gupta et al. Optional RRT Model (2002)

The shortcoming of the Eichhorn and Hayre (1983) [6] model is that some respondents may not like to multiply or may not know how to multiply the scrambling variable. The respondents still provide untruthful response. Singh et al. (1996) [23] showed that this case is more dangerous than using the scrambled response. Also, researchers realized that a question may be sensitive to one respondent, but not sensitive to another. In order to make the survey result more accurate, Gupta et al. (2002) [11] modified the Eichhorn and Hayre (1983) [6] multiplicative scrambling RRT model, and introduced an Optional RRT model that allows researchers to estimate not only the variable of interest, but also the sensitivity level W (the proportion of subjects who consider the question sensitive).

In this model, if respondents feel the question is sensitive, they will provide a scrambled response XS . Otherwise, the respondents will answer the sensitive survey question directly and provide the true response X . In this model, we assume that both

X and S are positive random variables and that the mean of the scrambling variable $\theta_s = 1$ and the variance σ_s^2 . Under this model, the reported response Z is given by

$$Z = \begin{cases} X & \text{with probability } 1-W \\ XS & \text{with probability } W. \end{cases} \quad (1.10)$$

The expected value of Z is given by

$$E(Z) = E(X)(1 - W) + E(XS)W = \mu_x(1 - W) + \mu_x\theta_s W = \mu_x, \quad (1.11)$$

and the variance of this unbiased estimator of the population mean μ is given by

$$Var(\hat{\mu}_x) = \frac{1}{n}[\sigma_x^2 + W\sigma_s^2(\sigma_x^2 + \mu_x^2)]. \quad (1.12)$$

Note that $Var(\hat{\mu}_x)$ increases with W, and hence there is gain in efficiency compared to the non-optional model when W=1. The estimator for the sensitivity level W is given by

$$\hat{W} = \frac{\frac{1}{n} \sum_{i=1}^n \log(Z_i) - \log(\frac{1}{n} \sum_{i=1}^n Z_i)}{\delta}, \quad (1.13)$$

where $\delta = E[\log(S)]$. This estimator uses first order approximation.

1.1.5 Gupta et al. Optional Split-sample Approach RRT Model (2004)

Gupta et al. (2004) [12] proposed an optional split-sample approach RRT model that allows simultaneous estimation of both the sensitive variable and the sensitivity level, and is not based on any approximation.

The split-sample approach means that we split the sample into two subgroups. One group of respondents uses a scrambling variable S_1 , and the other group uses a different scrambling variable S_2 . Thus, under this model, the reported responses Z_i ($i=1,2$) are given by

$$Z_i = \begin{cases} X & \text{with probability } 1-W \\ XS_i & \text{with probability } W \end{cases} \quad \text{where } i = 1, 2. \quad (1.14)$$

The expected value of Z_i is given by

$$E(Z_i) = \mu_x(1 - W) + \mu_x\theta_iW, \quad \text{where } E(S_i) = \theta_i \quad (i = 1, 2). \quad (1.15)$$

The two estimators $\hat{\mu}_x$ and \hat{W} are given by:

$$\hat{\mu}_x = \frac{\bar{Z}_1(\theta_2 - 1) - \bar{Z}_2(\theta_1 - 1)}{\theta_2 - \theta_1}, \quad (1.16)$$

$$\hat{W} = \frac{\bar{Z}_1 - \bar{Z}_2}{[\bar{Z}_2(\theta_1 - 1) - \bar{Z}_1(\theta_2 - 1)]}, \quad \theta_1 \neq \theta_2. \quad (1.17)$$

1.1.6 Gupta et al. Optional Additive RRT Model (2010)

The multiplicative scrambling compromises respondent anonymity. For example, if the respondent's true response is zero, no matter what scrambling number s/he chooses, the reported response will be zero. In this case, a non-zero response means the respondent has some degree of sensitive characteristic.

In order to deal with the problem and to improve accuracy, Gupta et al. (2010) [13] made two adjustments. The new model uses partial scrambling which means a randomly selected proportion (T) of respondents are asked to provide true responses directly and the remaining respondents in the sample are asked to use the optional scrambling method; since the multiplicative scrambling method compromises respondent anonymity, the scrambling method used is additive scrambling.

Again, let X be sensitive variable with mean μ_x , $S_i (i = 1, 2)$ be scrambling variable (independent of X) with mean $\theta_i (i = 1, 2)$ and variance $\sigma_{S_i}^2 (i = 1, 2)$, and $Z_i (i = 1, 2)$ be the reported response. If a proportion W of respondents feel the survey question is sensitive, then the proportion of respondents who use the scrambling method is $(1 - T)W$, and the proportion of respondents who provide the true response is $T + (1 - T)(1 - W)$. Thus, under this model, the reported response Z_i in the i^{th} sub-sample is given by

$$Z_i = \begin{cases} X & \text{with probability } T + (1 - T)(1 - W) \\ X + S_i & \text{with probability } (1 - T)W \end{cases} \quad \text{where } i = 1, 2. \quad (1.18)$$

The expected value and variance of Z_i are given by

$$E(Z_i) = \mu_x + \theta_i W(1 - T), \text{ where } \theta_i = E(S_i) \text{ (} i = 1, 2\text{)}, \quad (1.19)$$

$$Var(Z_i) = \sigma_x^2 + \sigma_{S_i}^2 [(1 - T)W] + \theta_i^2 [(1 - T)W] \{1 - [(1 - T)W]\}. \quad (1.20)$$

The unbiased estimators $\hat{\mu}_x$ and \hat{W} and their corresponding variance are given by

$$\hat{\mu}_x = \frac{\theta_1 \bar{Z}_2 - \theta_2 \bar{Z}_1}{\theta_1 - \theta_2}, \quad (1.21)$$

$$\hat{W} = \frac{\bar{Z}_1 - \bar{Z}_2}{(\theta_1 - \theta_2)(1 - T)} \quad \theta_1 \neq \theta_2, \quad (1.22)$$

$$Var(\hat{\mu}_x) = \frac{1}{(\theta_2 - \theta_1)^2} \left(\theta_2^2 \frac{\sigma_{z_1}^2}{n_1} + \theta_1^2 \frac{\sigma_{z_2}^2}{n_2} \right), \quad (1.23)$$

and

$$Var(\hat{W}) = \frac{1}{(\theta_2 - \theta_1)^2 (1 - T)^2} \left(\frac{\sigma_{z_1}^2}{n_1} + \frac{\sigma_{z_2}^2}{n_2} \right). \quad (1.24)$$

In this section, we introduced several RRT models and presented estimators for the sensitive variable mean, as well as the estimator for the sensitivity level. Among them, the Gupta et al. (2010) model [13] is used in our research.

1.2 Ordinary Ratio and Product Estimators

1.2.1 Ordinary Ratio Estimation

Finite population sampling is used for drawing inferences about a finite population based on a sample taken from the population. Let Y be a variable of interest with a value of y_i on the i^{th} unit. The purpose is to estimate population mean μ_y and population variance σ_y^2 , where $\mu_y = \frac{1}{N} \sum_{i=1}^N y_i$ and $\sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \mu)^2$. With a simple random sample without replacement (SRSWOR), an unbiased ordinary mean estimator of the population mean μ_y is the sample mean \bar{y} , and an unbiased estimator of the finite-population variance σ^2 is s^2 , where

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad \text{and} \quad s_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \mu)^2. \quad (1.25)$$

Many times there are one or more auxiliary variables, associated with the variable of interest, are available. Auxiliary information is usually used in the sampling design or in estimation. The use of samples from both Y and X can produce more precise estimates than the estimator obtained from sample from Y alone. Cochran [4] proposed the ratio estimator for the population mean μ_r , where it is assumed that the population mean μ_x of an auxiliary variable X is known. For example, if the variable of interest Y is the quantity of fruits produced in a plot, an auxiliary variable X could be the area of the plot. The area of the plot and quantity of fruits produced have

a positive linear relationship through the origin. Suppose the plot area x_i is known for the whole population, and the mean quantity of fruit for the population μ_y is the parameter to be estimated. For a simple random sample without replacement with sample size n , the sample ratio r is given by

$$r = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} = \frac{\bar{y}}{\bar{x}}. \quad (1.26)$$

The ratio estimator of the population mean μ_y is given by

$$\hat{\mu}_r = \bar{y} \left(\frac{\mu_x}{\bar{x}} \right) = r \mu_x. \quad (1.27)$$

The expected value of $\hat{\mu}_r$, based on the first order approximation, is given by

$$E(\hat{\mu}_r) \approx \mu_x R = \mu_x \left(\frac{\mu_y}{\mu_x} \right) = \mu_y. \quad (1.28)$$

Let C_x and C_y respectively be coefficients of variation of the study variable and the auxiliary variable, and ρ be the correlation coefficient between these two variables.

The approximate MSE of the ratio estimator $\hat{\mu}_r$ is given by

$$MSE(\hat{\mu}_r) \approx \left(\frac{N-n}{Nn} \right) \mu_y^2 (C_x^2 - 2\rho C_x C_y + C_y^2). \quad (1.29)$$

1.2.2 Product Estimation

When the study variable Y and an auxiliary variable X are negatively correlated, the product estimator of population mean μ , introduced by Robson (1957) [21] and later revised by Murthy (1964) [19], is used. Again, let μ_y and μ_x be population mean of the study variable and an auxiliary variable, \bar{y} and \bar{x} be the sample means based on the sample drawn. The product estimator of the population mean μ_y is given by

$$\hat{\mu}_p = \bar{y} \frac{\bar{x}}{\mu_x}. \quad (1.30)$$

The expected value of $\hat{\mu}_p$, based on the first order approximation, is given by

$$E(\hat{\mu}_p) \approx \mu_y \frac{\mu_x}{\mu_x} = \mu_y. \quad (1.31)$$

The approximate MSE of the product estimator is given by

$$MSE(\hat{\mu}_p) \approx \left(\frac{N-n}{Nn}\right) \mu_y^2 (C_x^2 + 2\rho C_x C_y + C_y^2). \quad (1.32)$$

1.2.3 Efficiency Comparisons

Based on the first order approximation, the expected values of $\hat{\mu}_r$ and $\hat{\mu}_p$ are given by

$$E(\hat{\mu}_r) \approx \mu_x R = \mu_x \left(\frac{\mu_y}{\mu_x}\right) = \mu_y \quad \text{and} \quad E(\hat{\mu}_p) \approx \mu_y \frac{\mu_x}{\mu_x} = \mu_y. \quad (1.33)$$

Thus, the ratio estimator and the product estimator are approximately unbiased. Because the ordinary mean estimator is unbiased, the first order approximation to the mean square error is the same as its variance. Let us just compare the approximate MSEs of the ratio estimator and the product estimator with the variance of the ordinary mean estimator. The variance of the ordinary mean estimator with simple random sampling without replacement is given by

$$Var(\bar{y}) = \left(\frac{N-n}{Nn}\right)\mu_y^2 C_y^2. \quad (1.34)$$

By comparing the expressions of MSEs in equations (1.29) and (1.32) with the variance of the ordinary mean estimator, we note that $MSE(\hat{\mu}_r) < Var(\bar{y})$ if $\rho > \frac{C_x}{2C_y}$, and $MSE(\hat{\mu}_p) < Var(\bar{y})$ if $\rho < -\frac{C_x}{2C_y}$. This means when the coefficient of correlation ρ between the study variable and an auxiliary variable is greater than $\frac{1}{2}$ ($\rho > \frac{1}{2}$) and the coefficient of variation $C_x \approx C_y$, the ratio estimator is more precise than the ordinary estimator. Similarly for the product estimator, when $\rho < -\frac{1}{2}$ and the coefficient of variations $C_x \approx C_y$, the product estimator is more efficient than the ordinary mean estimator. Hence, it is better to use the ratio estimator $\hat{\mu}_r$ when the y_i and x_i have a roughly linear relationship through the origin and with high positive correlation coefficient ($\rho > \frac{1}{2}$), and use the product estimator when two variables are negatively correlated with correlation coefficient smaller than $-\frac{1}{2}$.

1.3 Improvement in RRT Mean Estimation using Auxiliary Information

In the previous sections, we discussed various types of quantitative RRT models, the ordinary ratio estimation and product estimation. RRT models are used for surveys containing sensitive issues, such as collecting data on drug use or sexual

behavior. These models protect the privacy of the respondents. The efficiency of RRT models have been examined both through simulations and some field tests. Chow et al. (1979) [3] used a RRT model to estimate the incidence of induced abortion among currently married women of childbearing age in a rural area (Nekempte) in Ethiopia; Chaloupka et al. (1985) [1] applied a RRT model to estimate noncompliance of collecting marine product permits; Cross et al. (2010) [5] used a RRT to obtain sensitive information on animal disease prevalence; Kwan et al.(2010) [17] had a study on software piracy based on RRT model, and their results showed that respondents responding using RRT were more willing to provide true response about the behaviors on software piracy; Chhabra et al.(2016) [2] used RRT models to estimate prevalence of sexual abuse of girls by an acquaintance.

In addition, ratio and product estimators provide more accurate results than the ordinary mean estimator when an auxiliary variable that is highly correlated with the study variable is used. These two estimators have been used by many researchers including Kadilar et al. (2006) [14] and Grover et al. (2011) [10].

Suppose we have a sensitive primary variable and a non-sensitive but strongly correlated auxiliary variable is available. In this situation, one can take advantage of both the RRT methodology and ratio estimation. Sousa et al. (2010) was the first to use used ratio estimators under RRT models. In this case, they are estimating the mean of a sensitive variable using an non-optional RRT model with the utilization of a non-sensitive auxiliary variable. Later, Kalucha et al. (2015) improved the Sousa et al. (2010) estimator [24] further by using an optional RRT model. Kalucha et al. (2015) [15] proposed two ratio estimators in the context of optional RRT models. They showed that one of them, the additive ratio estimator, is more efficient

than the estimator by Gupta et al. (2010) [13]. However, their second estimator, a multiplicative ratio estimator, did not perform well. This thesis is focused on improving the multiplicative ratio estimator proposed by Kalucha et al.

1.4 Thesis Outline

Chapter 1 provides a brief introduction to several RRT models and the basic idea about ratio and product estimators, as well as RRT models using auxiliary information.

Chapter 2 presents ratio estimators of a finite population mean in the presence of a non-sensitive auxiliary variable, including Sousa et al. (2010) and Kalucha et al. (2015) estimators. The efficiency of the two ratio estimators by Kalucha et al. (2015) are compared with the ordinary RRT estimator by Gupta et al. (2010) both theoretically and numerically.

Chapter 3 presents a new multiplicative ratio estimator under the additive optional RRT model using a non-sensitive auxiliary variable. This chapter will compare bias and mean square error (MSE), up to the first order of approximation, of the proposed ratio estimator and the two previous ratio estimators of Kalucha et al. (2015), as well as the ordinary mean estimator. The theoretical results shows that the proposed mean ratio estimator coincides with the additive ratio estimator but it is less biased than the additive ratio estimator. The proposed mean ratio estimator is more efficient than the multiplicative ratio estimator. Also, the result shows that the proposed mean ratio estimator is more efficient than the ordinary mean estimator when the variable of interest and the auxiliary variable are highly correlated.

Chapter 4 presents simulation result comparing the proposed estimator with other estimators.

Chapter 5 presents a general discussion on the research carried out in this thesis, as well as some future directions.

CHAPTER II
RATIO ESTIMATORS OF MEAN WHEN THE STUDY VARIABLE IS
SENSITIVE

2.1 Sousa et al. (2010): Ratio Estimator under RRT Models

In Chapter 1, we briefly described the improvement in RRT mean estimation using non-sensitive auxiliary variable X which is highly correlated with the sensitive study variable Y . Ratio estimators provide more accurate results than the ordinary mean estimators when the study variable is nonsensitive. Such estimators have been presented by many researchers, including Kadilar et al. (2006) [14], Turgut et al. (2008) [26] and Shabbir et al. (2010) [22], but sometimes the study variable is sensitive, and the respondents may not feel comfortable to provide true response. In such cases, RRT models play a role on improving the response rate because RRT models increase subject anonymity and the respondents are more willing to provide accurate response.

Sousa et al. (2010) [24] proposed a non-optional RRT estimator for the mean of sensitive variable improved by utilizing a non-sensitive auxiliary variable. Let a simple random sample of size n be drawn without replacement from a finite population. Under non-optional additive RRT model, all the respondents are asked to report a scrambled response for the sensitive variable Y given by $Z=Y+S$, where S is a scrambling variable (independent of X and Y), but are asked to report a true response for an auxiliary variable X .

We assume that μ_x is known and $E(S) = 0$. Thus, $E(Z) = E(Y)$. The ratio estimator under this RRT model is given by

$$\hat{\mu}_R = \bar{z} \left(\frac{\mu_x}{\bar{x}} \right), \quad (2.1)$$

where \bar{z} is the sample mean of reported responses, μ_x and \bar{x} are population mean and sample mean of an auxiliary variable.

A large sample size is assumed so that $|\delta_z| < 1$ and $|\delta_x| < 1$, where $\delta_z = \frac{\bar{z} - \mu_z}{\mu_z}$, $\delta_x = \frac{\bar{x} - \mu_x}{\mu_x}$. Thus, for large samples, the ratio estimator (2.1) can be written as

$$\hat{\mu}_R = \mu_z (1 + \delta_z) (1 + \delta_x)^{-1}. \quad (2.2)$$

Using Taylor's approximation and retaining terms of order up to 2, (2.2) can be rewritten as

$$\hat{\mu}_R - \mu_z \approx \mu_z (\delta_z + \delta_x^2 - \delta_x - \delta_z \delta_x). \quad (2.3)$$

Under the assumption of bivariate normality (Sukhatme et al., 1970): $E(\delta_x) = 0$; $E(\delta_z) = 0$; $E(\delta_z^2) = \frac{1-f}{n} C_z^2$; $E(\delta_x^2) = \frac{1-f}{n} C_x^2$; $E(\delta_x \delta_z) = \frac{(1-f)}{n} C_{zx}$; where $f = \frac{n}{N}$, $C_{zx} = \rho_{zx} C_z C_x$, and C_z and C_x are the coefficients of variation of Z and X respectively.

Recognizing that $\mu_Z = \mu_Y$, and using first order Taylor's approximation, the bias of the ratio estimator (2.1) is given by

$$Bias(\hat{\mu}_R) \approx \frac{1-f}{n} \mu_Y (C_x^2 - \rho_{zx} C_z C_x). \quad (2.4)$$

Similarly from (2.3) and (2.4), the mean square error (MSE) of $\hat{\mu}_R$, correct to first order of approximation, is given by

$$MSE(\hat{\mu}_R) \approx \frac{1-f}{n} \mu_Y^2 (C_x^2 + C_z^2 - 2\rho_{zx} C_z C_x). \quad (2.5)$$

Sousa et al. (2010) also compared this efficiency of this ratio estimator and the estimator where the auxiliary variable is not utilized. Their theoretical and empirical results show that this ratio estimator is more efficient.

2.2 Kalucha et al. (2015): Ratio Estimators under Optional RRT Models

Gupta et al. (2010) [13] showed that the additive optional RRT model performs better than the non-optional RRT models. Thus, Kalucha et al. (2015) [15] modified the Sousa et al. (2010) [24] estimators further by using an additive optional quantitative RRT model. Let us first recall the Gupta et al (2010) estimator with $T=0$.

In the Gupta et al. (2010) additive optional RRT model with $T=0$, a simple random sample of size n is chosen without replacement (SRSWOR) from of finite population of size N . The two sub-samples have sample of sizes n_1 and n_2 , respectively. Let Y be the sensitive variable and X be a non-sensitive auxiliary variable which has

high positive correlation with Y. Let S_i ($i=1,2$) be scrambling variables that are independent of X and Y, with mean θ_i and variance $\sigma_{S_i}^2$ ($i=1,2$); and Z_i ($i=1,2$) be the responses from the two sub-samples. Let $\mu_x = E(X)$, $\mu_y = E(Y)$, $\mu_{Z_1} = E(Z_1)$, $\mu_{Z_2} = E(Z_2)$ be the population means for X, Y, Z_1 , and Z_2 , respectively. The unbiased mean and sensitivity estimators $\hat{\mu}_y$ and \hat{W} are given by

$$\hat{\mu}_y = \frac{\theta_2 \bar{Z}_1 - \theta_1 \bar{Z}_2}{\theta_2 - \theta_1}, \quad \text{and} \quad \hat{W} = \frac{\bar{Z}_2 - \bar{Z}_1}{\theta_2 - \theta_1}, \quad (\theta_1 \neq \theta_2). \quad (2.6)$$

The MSE of $\hat{\mu}_y$ is given by

$$MSE(\hat{\mu}_y) = \frac{1}{(\theta_2 - \theta_1)^2} [\theta_2^2 (\frac{1-f_1}{n_1}) \sigma_{Z_1}^2 + \theta_1^2 (\frac{1-f_2}{n_2}) \sigma_{Z_2}^2], \quad (\theta_1 \neq \theta_2), \quad (2.7)$$

where $f_1 = \frac{n_1}{N}$, $f_2 = \frac{n_2}{N}$, $f = \frac{n}{N} = f_1 + f_2$, $\sigma_{Z_1}^2 = \frac{1}{N-1} \sum_{i=1}^N (Z_{1i} - \mu_z)^2$, and $\sigma_{Z_2}^2 = \frac{1}{N-1} \sum_{i=1}^N (Z_{2i} - \mu_z)^2$.

Based on this additive optional RRT model, Kalucha et al. (2015) proposed two mean ratio estimators. The details are presented below and the simulation results will be shown in Section 2.3.

2.2.1 Additive and Multiplicative Ratio Estimators

Kalucha et al. (2015) assume that μ_x is known and proposed two ratio estimators of finite population mean and called them additive ratio estimator and the multiplicative ratio estimator. These are given by

$$\hat{\mu}_{AR} = \left(\frac{\theta_2 \bar{z}_1 - \theta_1 \bar{z}_2}{\theta_2 - \theta_1} \right) \left(\frac{\mu_x}{\bar{x}_1} + \frac{\mu_x}{\bar{x}_2} \right) \left(\frac{1}{2} \right) \quad (2.8)$$

and

$$\hat{\mu}_{MR} = \left(\frac{\theta_2 \bar{z}_1 - \theta_1 \bar{z}_2}{\theta_2 - \theta_1} \right) \left(\frac{\mu_x}{\bar{x}_1} \right) \left(\frac{\mu_x}{\bar{x}_2} \right). \quad (2.9)$$

A large sample size is assumed so that $|\delta_{z_i}| < 1$ and $|\delta_{x_i}| < 1$ (i=1,2), where $\delta_{Z_1} = \frac{\bar{z}_1 - \mu_{Z_1}}{\mu_{Z_1}}$, $\delta_{Z_2} = \frac{\bar{z}_2 - \mu_{Z_2}}{\mu_{Z_2}}$, $\delta_{x_1} = \frac{\bar{x}_1 - \mu_X}{\mu_X}$, $\delta_{x_2} = \frac{\bar{x}_2 - \mu_X}{\mu_X}$. The two ratio estimators in (2.8) and (2.9) can now be written as

$$\hat{\mu}_{AR} = \frac{1}{2} \left[\mu_Y + \delta_{Z_1} \mu_{Z_1} \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) - \delta_{Z_2} \mu_{Z_2} \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) \right] \left[(1 + \delta_{x_1})^{-1} + (1 + \delta_{x_2})^{-1} \right] \quad (2.10)$$

and

$$\hat{\mu}_{MR} = \left[\mu_Y + \delta_{Z_1} \mu_{Z_1} \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) - \delta_{Z_2} \mu_{Z_2} \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) \right] (1 + \delta_{x_1})^{-1} (1 + \delta_{x_2})^{-1}. \quad (2.11)$$

Under the assumption of bivariate normality (Sukhatme et al.1970)[25]: $E(\delta_{x_1}) = 0$; $E(\delta_{x_2}) = 0$; $E(\delta_{x_1}^2) = \left(\frac{1-f_1}{n_1} \right) C_{x_1}^2$; $E(\delta_{x_2}^2) = \left(\frac{1-f_2}{n_2} \right) C_{x_2}^2$; $E(\delta_{x_1} \delta_{x_2}) = 0$; $E(\delta_{z_1}) = 0$; $E(\delta_{z_2}) = 0$; $E(\delta_{z_1} \delta_{x_2}) = 0$; $E(\delta_{z_2} \delta_{x_1}) = 0$; $E(\delta_{z_1} \delta_{x_1}) = \left(\frac{1-f_1}{n_1} \right) C_{z_1 x_1}$; $E(\delta_{z_2} \delta_{x_2}) = \left(\frac{1-f_2}{n_2} \right) C_{z_2 x_2}$, where $f_1 = \frac{n_1}{N}$, $f_2 = \frac{n_2}{N}$, $C_{z_1 x_1} = \rho_{z_1 x_1} C_{z_1} C_{x_1}$, and $C_{z_2 x_2} = \rho_{z_2 x_2} C_{z_2} C_{x_2}$.

Using first order Taylor's approximation, the bias of these estimators are given by

$$\begin{aligned} Bias(\hat{\mu}_{AR}) \approx & \frac{\mu_Y}{2} \left[\left(\frac{1-f_1}{n_1} \right) C_x^2 + \left(\frac{1-f_2}{n_2} \right) C_x^2 \right] - \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) \frac{1}{2} \left[\left(\frac{1-f_1}{n_1} \right) \rho_{yx} \sigma_Y C_x \right] \\ & + \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) \frac{1}{2} \left[\left(\frac{1-f_2}{n_2} \right) \rho_{yx} \sigma_Y C_x \right] \end{aligned} \quad (2.12)$$

and

$$\begin{aligned} Bias(\hat{\mu}_{MR}) \approx & \mu_Y \left[\left(\frac{1-f_1}{n_1} \right) C_x^2 + \left(\frac{1-f_2}{n_2} \right) C_x^2 \right] - \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) \left[\left(\frac{1-f_1}{n_1} \right) \rho_{yx} \sigma_Y C_x \right] \\ & + \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) \left[\left(\frac{1-f_2}{n_2} \right) \rho_{yx} \sigma_Y C_x \right]. \end{aligned} \quad (2.13)$$

Expressions in (2.12) and (2.13) use the facts: $E(\delta_{x_1}^2) = \left(\frac{1-f_1}{n_1} \right) C_{x_1}^2$; $E(\delta_{x_2}^2) = \left(\frac{1-f_2}{n_2} \right) C_{x_2}^2$; $E(\delta_{x_1} \delta_{x_2}) = 0$; $E(\delta_{z_1}^2) = \left(\frac{1-f_1}{n_1} \right) C_{z_1}^2$; $E(\delta_{z_2}^2) = \left(\frac{1-f_2}{n_2} \right) C_{z_2}^2$; $E(\delta_{z_1} \delta_{z_2}) = 0$; $E(\delta_{z_1} \delta_{x_1}) = \left(\frac{1-f_1}{n_1} \right) \rho_{z_1 x_1} C_{z_1} C_{x_1}$; $E(\delta_{z_2} \delta_{x_2}) = \left(\frac{1-f_2}{n_2} \right) \rho_{z_2 x_2} C_{z_2} C_{x_2}$; $E(\delta_{z_2} \delta_{x_1}) = 0$; $E(\delta_{z_1} \delta_{x_2}) = 0$. Since the two sub-samples come from the same population, the coefficients of variation $C_{x_1} = C_{x_2} = C_x$, and the correlations $\rho_{z_1 x_1} = \rho_{z_1 x}$, and $\rho_{z_2 x_2} = \rho_{z_2 x}$. Thus, the expressions for MSE of $\hat{\mu}_{MR}$ and $\hat{\mu}_{AR}$, up to the first order of approximation, are given by

$$\begin{aligned} MSE(\hat{\mu}_{MR}) = & E(\mu_{MR} - \mu_Y)^2 \\ \approx & \left(\frac{1-f_1}{n_1} \right) \left[\left(\frac{\theta_2}{\theta_2 - \theta_1} \right)^2 \sigma_{z_1}^2 + \mu_Y^2 C_x^2 - 2\mu_Y \rho_{yx} \sigma_Y \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) C_x \right] \\ & + \left(\frac{1-f_2}{n_2} \right) \left[\left(\frac{\theta_1}{\theta_2 - \theta_1} \right)^2 \sigma_{z_2}^2 + \mu_Y^2 C_x^2 + 2\mu_Y \rho_{yx} \sigma_Y \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) C_x \right] \end{aligned} \quad (2.14)$$

and

$$\begin{aligned} MSE(\hat{\mu}_{AR}) = & E(\mu_{AR} - \mu_Y)^2 \\ \approx & \left(\frac{1-f_1}{n_1} \right) \left[\left(\frac{\theta_2}{\theta_2 - \theta_1} \right)^2 \sigma_{z_1}^2 + \frac{1}{4} \mu_Y^2 C_x^2 - \mu_Y \rho_{yx} \sigma_Y \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) C_x \right] \\ & + \left(\frac{1-f_2}{n_2} \right) \left[\left(\frac{\theta_1}{\theta_2 - \theta_1} \right)^2 \sigma_{z_2}^2 + \frac{1}{4} \mu_Y^2 C_x^2 + \mu_Y \rho_{yx} \sigma_Y \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) C_x \right]. \end{aligned} \quad (2.15)$$

Kalucha et al. (2015) also compared the efficiency of these proposed mean ratio estimators in (2.8) (2.9) and the ordinary ratio estimator in (2.6). Their results showed that in both cases of unequal and equal sample split, the multiplicative mean ratio estimator was not as efficient as the ordinary mean estimator under additive optional RRT model. However, the additive mean ratio estimator was more efficient than the ordinary mean estimator when the coefficient of correlation ρ_{xy} is greater than $\frac{1}{2}$. Also the additive mean ratio estimator was always more efficient than the multiplicative mean ratio estimator.

In the unequal sample split case, parameters of the scrambling variables should satisfy the following conditions:

If the scrambling variable means are positive, then a smaller sub-sample is used for the smaller scrambling variable mean; if the scrambling variable means are negative, then a smaller sub-sample is used for the larger scrambling variable mean; if the scrambling variable means have opposite signs, then a smaller sub-sample is used for the larger absolute scrambling variable mean; and if one of the scrambling variable means is zero, then a smaller sub-sample is used to this scrambling variable.

2.3 Simulation Results

We present below two simulation studies to show the performance of the Kalucha et al. (2015) mean ratio estimators as compared with the ordinary mean estimator under the optional RRT model.

2.3.1 Kalucha et al. Simulation

First we use the same parameters as Kalucha et al. (2015), and consider two finite populations with size $N=5000$ from a bivariate normal distribution with correlation

coefficients $\rho_{xy} = 0.3$ and $\rho_{xy} = 0.8$, respectively. The non-sensitive auxiliary variable X in each population has a mean of 4 and a variance of 4. And the sensitive variable Y has a mean of 6 and a variance of 9. The variance of the scrambling variables are $\sigma_1^2 = 2$ and $\sigma_2^2 = 1$. For each population, samples with different sizes (n=500, 1000, 1500, 2000) will be considered. We also consider both cases of equal sample split and unequal sample split. The empirical MSE is estimated by using 1000 samples selected from the population.

Table 1 presents the MSE comparisons for the ordinary RRT estimator μ_y and the additive ratio estimator μ_{AR} for different correlations ρ_{xy} using equal sample split when the mean of the scrambling variables are $\theta_1 = 5$ and $\theta_2 = 0.5$.

Table 2 and Table 3 show the theoretical and empirical MSEs of the ordinary RRT estimator and the additive ratio estimator when using unequal sample split. In Table 2, we use the two positive scrambling variable means $\theta_1 = 0.5$ and $\theta_2 = 5$, and the scrambling variable with the larger mean associated with a larger sub-sample; in Table 3, we use two negative scrambling variable means $\theta_1 = -5$ and $\theta_2 = -0.5$, and the scrambling variable with the smaller mean associated with a larger sub-sample.

In all cases, the results shows that the additive ratio estimator is more efficient than the ordinary RRT estimator when the correlation $\rho_{xy} > 0.5$ ($\rho_{xy} = 0.8$); but it is not as efficient as the ordinary RRT estimator for correlation $\rho_{xy} < 0.5$ ($\rho_{xy} = 0.3$). We also observe that as the sensitivity level W increases, the MSE of the two estimators increase as well. This indicates that the efficiency of an optional RRT model is greater than that of a non-optional RRT where $W=1$.

The theoretical MSEs of the ordinary RRT estimator and the additive ratio estimator we calculated are very close to the results in Kalucha et al. (2015). Even though

the empirical MSEs are slightly different, they are all close to the theoretical MSEs. The reason for the differences is that we simulated the populations using different seeds. In order to make sure the parameters we used here are not a clever choice, we conduct another simulation using different parameters. The simulation of Kalucha et al. (2015) concentrated only on the additive ratio estimator since their multiplicative ratio estimator was not as efficient as the other two estimators. In the simulation with new parameters below, we include the comparisons of their multiplicative ratio estimator as well.

Table 1. Theoretical (**bold**) and Empirical MSE Comparisons with Estimates of W , μ_y and μ_{AR} for $\rho_{xy} = 0.3$ and $\rho_{xy} = 0.8$ using **Equal Sample Split** ($\theta_1 = 5$, $\theta_2 = 0.5$, $n_1 = n_2$).

n	n_1	n_2	W	$p = 0.8$						$p = 0.3$					
				\hat{W}	MSE		\hat{W}	MSE		\hat{W}	MSE				
					$\hat{\mu}_y$	$\hat{\mu}_{AR}$		$\hat{\mu}_y$	$\hat{\mu}_{AR}$		$\hat{\mu}_y$	$\hat{\mu}_{AR}$			
500	250	250	0.3	0.2939	0.0441	0.0341	0.2969	0.0441	0.0512	0.0440	0.0352	0.2969	0.0440	0.0519	
			0.5	0.5060	0.0451	0.0352	0.4953	0.0451	0.0522	0.0450	0.0357	0.4953	0.0450	0.0510	
			0.7	0.7028	0.0460	0.0361	0.7022	0.0460	0.0531	0.0469	0.0369	0.7022	0.0469	0.0540	
			0.8	0.8110	0.0464	0.0364	0.8006	0.0464	0.0535	0.0465	0.0369	0.8006	0.0465	0.0538	
1000	500	500	0.3	0.3017	0.0209	0.0162	0.2993	0.0209	0.0242	0.0208	0.0169	0.2993	0.0208	0.0240	
			0.5	0.5011	0.0214	0.0167	0.5007	0.0214	0.0247	0.0213	0.0169	0.5007	0.0213	0.0259	
			0.7	0.6991	0.0218	0.0171	0.7010	0.0218	0.0252	0.0225	0.0182	0.7010	0.0225	0.0250	
			0.8	0.7982	0.0220	0.0173	0.8013	0.0220	0.0253	0.0229	0.0179	0.8013	0.0229	0.0263	
1500	750	750	0.3	0.3014	0.0131	0.0102	0.3005	0.0131	0.0153	0.0135	0.0106	0.3005	0.0135	0.0150	
			0.5	0.4986	0.0135	0.0105	0.5011	0.0135	0.0156	0.0130	0.0113	0.5011	0.0130	0.0160	
			0.7	0.6991	0.0137	0.0108	0.6987	0.0137	0.0158	0.0141	0.0110	0.6987	0.0141	0.0163	
			0.8	0.7988	0.0138	0.0109	0.7959	0.0138	0.0160	0.0144	0.0116	0.7959	0.0144	0.0170	
2000	1000	1000	0.3	0.2986	0.0093	0.0072	0.2993	0.0093	0.0108	0.0086	0.0076	0.2993	0.0086	0.0115	
			0.5	0.4899	0.0095	0.0074	0.4989	0.0095	0.0110	0.0089	0.0083	0.4989	0.0089	0.0116	
			0.7	0.6985	0.0097	0.0076	0.6996	0.0097	0.0112	0.0090	0.0077	0.6996	0.0090	0.0119	
			0.8	0.8013	0.0098	0.0077	0.8008	0.0098	0.0113	0.0109	0.0080	0.8008	0.0109	0.0116	

Table 2. Theoretical (**bold**) and Empirical MSE Comparisons with Estimates of W , μ_y and μ_{AR} for $\rho_{xy} = 0.3$ and $\rho_{xy} = 0.8$ using **Unequal Sample Split** ($\theta_1 = 0.5$, $\theta_2 = 5$, $n_1 < n_2$).

n	n_1	n_2	W	$p = 0.8$				$p = 0.3$			
				\hat{W}	MSE		\hat{W}	MSE			
					$\hat{\mu}_y$	$\hat{\mu}_{AR}$		$\hat{\mu}_y$	$\hat{\mu}_{AR}$		
500	200	300	0.3	0.2974	0.0570 0.0536	0.0393 0.0420	0.2991	0.0570 0.0536	0.0616 0.0630		
			0.5	0.4968	0.0594 0.0585	0.0418 0.0430	0.5009	0.0594 0.0585	0.0640 0.0685		
			0.7	0.6999	0.0617 0.0596	0.0440 0.0468	0.7007	0.0617 0.0596	0.0662 0.0691		
			0.8	0.8022	0.0627 0.0656	0.0450 0.0459	0.7981	0.0627 0.0656	0.0673 0.0693		
1000	450	550	0.3	0.3021	0.0245 0.0233	0.0187 0.0183	0.3016	0.0245 0.0233	0.0269 0.0303		
			0.5	0.4966	0.0249 0.0238	0.0189 0.0206	0.5016	0.0249 0.0238	0.0279 0.0333		
			0.7	0.6991	0.0263 0.0272	0.0198 0.0203	0.7004	0.0263 0.0272	0.0288 0.0330		
			0.8	0.7981	0.0274 0.0289	0.0205 0.0219	0.7995	0.0274 0.0289	0.0298 0.0350		
1500	650	850	0.3	0.2993	0.0163 0.0153	0.0117 0.0119	0.3017	0.0163 0.0153	0.0173 0.0179		
			0.5	0.4968	0.0172 0.0170	0.0120 0.0120	0.5016	0.0172 0.0170	0.0185 0.0191		
			0.7	0.6992	0.0176 0.0169	0.0130 0.0137	0.6990	0.0176 0.0169	0.0191 0.0203		
			0.8	0.8023	0.0178 0.0196	0.0129 0.0129	0.8018	0.0178 0.0196	0.0195 0.0210		
2000	950	1050	0.3	0.3006	0.0108 0.0123	0.0076 0.0068	0.3017	0.0108 0.0123	0.0126 0.0119		
			0.5	0.5016	0.0110 0.0113	0.0080 0.0103	0.5031	0.0110 0.0113	0.0126 0.0132		
			0.7	0.7001	0.0115 0.0129	0.0081 0.0099	0.6985	0.0115 0.0129	0.0131 0.0129		
			0.8	0.7982	0.0120 0.0135	0.0083 0.0106	0.7896	0.0120 0.0135	0.0132 0.0136		

Table 3. Theoretical (**bold**) and Empirical MSE Comparisons with Estimates of W , μ_y and μ_{AR} for $\rho_{xy} = 0.3$ and $\rho_{xy} = 0.8$ using **Unequal Sample Split** ($\theta_1 = -5$, $\theta_2 = -0.5$, $n_1 > n_2$).

n	n_1	n_2	W	$p = 0.8$				$p = 0.3$			
				\hat{W}	MSE		\hat{W}	MSE			
					$\hat{\mu}_y$	$\hat{\mu}_{AR}$		$\hat{\mu}_y$	$\hat{\mu}_{AR}$		
500	300	200	0.3	0.3014	0.0560	0.0382	0.3010	0.0560	0.0605		
					0.0546	0.0386		0.0546	0.0609		
			0.5	0.5016	0.0573	0.0396	0.4962	0.0573	0.0619		
					0.0596	0.0402		0.0596	0.0629		
			0.7	0.7002	0.0584	0.0407	0.7022	0.0584	0.0630		
					0.0602	0.0409		0.0602	0.0637		
			0.8	0.7983	0.0589	0.0412	0.8009	0.0589	0.0635		
					0.0600	0.0419		0.0600	0.0659		
1000	550	450	0.3	0.3017	0.0242	0.0170	0.3012	0.0242	0.0265		
					0.0220	0.0200		0.0220	0.0269		
			0.5	0.5008	0.0247	0.0178	0.5006	0.0247	0.0273		
					0.0231	0.0186		0.0231	0.0287		
			0.7	0.6999	0.0252	0.0181	0.7003	0.0252	0.0275		
					0.0256	0.0189		0.0256	0.0290		
			0.8	0.7807	0.0253	0.0183	0.8023	0.0253	0.0278		
					0.0268	0.0192		0.0268	0.0302		
1500	850	650	0.3	0.2963	0.0162	0.0112	0.2869	0.0162	0.0169		
					0.0143	0.0115		0.0143	0.0179		
			0.5	0.4852	0.0165	0.0113	0.4926	0.0165	0.0171		
					0.0158	0.0119		0.0158	0.0186		
			0.7	0.6983	0.0166	0.0115	0.7016	0.0166	0.0172		
					0.0160	0.0115		0.0160	0.0198		
			0.8	0.8004	0.0169	0.0116	0.8026	0.0169	0.0183		
					0.0170	0.0120		0.0170	0.0197		
2000	1050	950	0.3	0.2963	0.0102	0.0073	0.2981	0.0102	0.0115		
					0.0097	0.0065		0.0097	0.0119		
			0.5	0.4985	0.0106	0.0079	0.5007	0.0106	0.0117		
					0.0109	0.0059		0.0109	0.0118		
			0.7	0.6983	0.0109	0.0080	0.7010	0.0109	0.0121		
					0.0116	0.0083		0.0116	0.0126		
			0.8	0.7853	0.0111	0.0083	0.7953	0.0111	0.0125		
					0.0113	0.0092		0.0113	0.0136		

2.3.2 New Simulation

We consider two finite populations with size $N=5000$ from a bivariate normal distribution with correlation coefficients $\rho_{xy} = 0.2$ and $\rho_{xy} = 0.7$ respectively. The non-sensitive auxiliary variable X in each population has a mean of 5 and a variance of 5. And the sensitive variable Y has a mean of 9 and a variance of 16. The variances of the scrambling variables are $\sigma_1^2 = 2$ and $\sigma_2^2 = 1$. For each population, samples with different sizes ($n=500, 1000, 1500, 2000$) are considered. We also consider both cases of equal sample split and unequal sample split. The empirical MSE is estimated by using 1000 samples selected from the population.

Table 4 presents the MSEs for the ordinary RRT estimator, the additive ratio estimator and the multiplicative ratio estimator using equal sample split when the scrambling variable means are $\theta_1 = 6$ and $\theta_2 = 0.6$. Table 5 and Table 6 show the same comparisons but using unequal sample split and different scrambling variable means. In Table 5, two scrambling variable means have positive sign and the smaller scrambling variable mean is associated with a smaller sample size. Two scrambling variable means are $\theta_1 = 0.6$ and $\theta_2 = 6$, respectively. In Table 6, two scrambling variable means used ($\theta_1 = -6$ and $\theta_2 = -0.6$) have negative sign and the smaller scrambling variable mean with a larger sample size.

All three tables show that the additive ratio estimator $\hat{\mu}_{AR}$ is more efficient than the ordinary mean estimator $\hat{\mu}_y$ when the correlation coefficient $\rho_{xy} = 0.7$, but it is not as efficient as the ordinary mean estimator when $\rho_{xy} = 0.2$. The multiplicative ratio estimator is not as efficient as the ordinary estimator or the additive ratio estimator under any of the sensitivity levels and correlations. We also observe that the MSEs

increases as the sensitivity level W increases. Thus, the new simulation study also gives results on expected lines for both equal and unequal sub-sample size cases.

2.4 Concluding Chapter Remarks

This chapter presented the ratio estimator by Sousa et al. (2010) under a non-optional RRT model, and also the two ratio estimators by Kalucha et al. (2015) under the Gupta et al. additive optional RRT model using a non-sensitive auxiliary variable. Kalucha et al. (2015) results showed that the additive ratio estimator is more efficient than the ordinary RRT estimator when the sensitive study variable Y and the auxiliary variable X have a correlation greater than $\frac{1}{2}$ for both equal and unequal sample split. Their results also showed that the multiplicative ratio estimator is not as good as the ordinary RRT estimator and the additive ratio estimator. Two simulation studies with different parameters showed this in the previous section. We also notice that the MSE of the estimators increase as the sensitivity level W increases, which indicates that the optional RRT ratio estimators are more efficient than the non-optional RRT ratio estimators.

In general, ratio estimator should be more efficient than the ordinary ratio estimator. But the multiplicative ratio estimator does not work efficiently, may be because the ratio in the estimator is the ordinary product of the ratios in the two sub-sample. If we take the geometric mean of these ratios, the efficiency of the multiplicative estimator may change. With this motivation, we will modify the multiplicative ratio estimator in the next chapter and improve its efficiency.

Table 4. Theoretical (**bold**) and Empirical MSE Comparisons of μ_y , μ_{AR} and μ_{MR} for Correlations $\rho_{xy} = 0.2$ and $\rho_{xy} = 0.7$ using **Equal Sample Split** ($\theta_1 = 6$, $\theta_2 = 0.6$, $n_1 = n_2$).

n	n_1	n_2	W	MSE								
				$p = 0.7$			$p = 0.2$					
				$\hat{\mu}_y$	$\hat{\mu}_{AR}$	$\hat{\mu}_{MR}$	$\hat{\mu}_y$	$\hat{\mu}_{AR}$	$\hat{\mu}_{MR}$			
500	250	250	0.3	0.0760	0.0637	0.1123	0.0760	0.0943	0.1731			
				0.0792	0.0632	0.1162	0.0792	0.0976	0.1715			
			0.5	0.0771	0.0648	0.1134	0.0771	0.0954	0.1742			
				0.0806	0.0685	0.1170	0.0806	0.0988	0.1781			
			0.7	0.0780	0.0657	0.1143	0.0780	0.0963	0.1751			
				0.0795	0.0692	0.1179	0.0795	0.0999	0.1796			
			0.8	0.0784	0.0661	0.1147	0.0784	0.0967	0.1755			
				0.0812	0.0699	0.1174	0.0812	0.1003	0.1790			
			1000	500	500	0.3	0.0360	0.0302	0.0532	0.0360	0.0447	0.0825
							0.0349	0.0326	0.0556	0.0349	0.0461	0.0820
0.5	0.0365	0.0307				0.0537	0.0365	0.0456	0.0830			
	0.0373	0.0324				0.0572	0.0373	0.0472	0.0846			
0.7	0.0370	0.0311				0.0541	0.0370	0.0466	0.0846			
	0.0389	0.0336				0.0569	0.0389	0.4790	0.0843			
0.8	0.0371	0.0313				0.0543	0.0371	0.0458	0.0831			
	0.0396	0.0340				0.0576	0.0396	0.0486	0.0859			
1500	750	750				0.3	0.0227	0.0190	0.0335	0.0227	0.0281	0.0516
							0.0216	0.0186	0.0349	0.0216	0.0292	0.0538
			0.5	0.0230	0.0193	0.0338	0.0230	0.0285	0.0520			
				0.0245	0.0219	0.0356	0.0245	0.0296	0.0547			
			0.7	0.0233	0.0196	0.0341	0.0238	0.0287	0.0522			
				0.0249	0.0200	0.0368	0.0249	0.0300	0.0555			
			0.8	0.0234	0.0197	0.0342	0.0234	0.0288	0.0523			
				0.0243	0.0209	0.0359	0.0243	0.0305	0.0546			
			2000	1000	1000	0.3	0.0160	0.0134	0.0236	0.0160	0.0198	0.0372
							0.0138	0.0149	0.0249	0.0138	0.0209	0.0365
0.5	0.0162	0.0136				0.0239	0.0162	0.0201	0.0365			
	0.0176	0.0157				0.0256	0.0176	0.0210	0.0383			
0.7	0.0164	0.0138				0.0241	0.0164	0.0203	0.0367			
	0.0182	0.0152				0.0260	0.0182	0.0219	0.0390			
0.8	0.0165	0.0139				0.0241	0.0165	0.0203	0.0369			
	0.0186	0.0164				0.0257	0.0186	0.0219	0.0392			

Table 5. Theoretical (**bold**) and Empirical MSE Comparisons of μ_y , μ_{AR} and μ_{MR} for Correlations $\rho_{xy} = 0.2$ and $\rho_{xy} = 0.7$ using **Unequal Sample Split** ($\theta_1 = 0.6$, $\theta_2 = 6$, $n_1 < n_2$).

n	n_1	n_2	W	MSE								
				$p = 0.7$			$p = 0.2$					
				$\hat{\mu}_y$	$\hat{\mu}_{AR}$	$\hat{\mu}_{MR}$	$\hat{\mu}_y$	$\hat{\mu}_{AR}$	$\hat{\mu}_{MR}$			
500	200	300	0.3	0.0980	0.0738	0.1132	0.0980	0.1139	0.1931			
				0.0972	0.0751	0.1159	0.0972	0.1152	0.2002			
			0.5	0.1004	0.0762	0.1156	0.1003	0.1163	0.1954			
				0.0982	0.0768	0.1168	0.0982	0.1164	0.2015			
			0.7	0.1025	0.0783	0.1177	0.1025	0.1184	0.1975			
				0.0989	0.0790	0.1166	0.0989	0.1179	0.2010			
			0.8	0.1035	0.0793	0.1187	0.1034	0.1194	0.1985			
				0.1015	0.0803	0.1189	0.1015	0.1175	0.2019			
			1000	400	600	0.3	0.0470	0.0352	0.0536	0.0469	0.0544	0.0919
							0.0452	0.0359	0.0549	0.0452	0.0564	0.0931
						0.5	0.0481	0.0363	0.0548	0.0481	0.0556	0.0931
							0.0480	0.0367	0.0560	0.0480	0.0548	0.0948
0.7	0.0491	0.0373				0.0558	0.0491	0.0566	0.0941			
	0.0493	0.0366				0.0541	0.0493	0.0550	0.0955			
0.8	0.0496	0.0378				0.0562	0.0496	0.0570	0.0945			
	0.0500	0.0380				0.0571	0.0500	0.0571	0.0951			
1500	600	900				0.3	0.0299	0.0223	0.0337	0.0299	0.0346	0.0582
							0.0308	0.0231	0.0349	0.0308	0.0357	0.0587
						0.5	0.0307	0.0230	0.0345	0.0306	0.0353	0.0589
							0.0297	0.0239	0.0356	0.0297	0.0368	0.0602
			0.7	0.0313	0.0237	0.0351	0.0313	0.0360	0.0596			
				0.0318	0.0252	0.0362	0.0318	0.0369	0.0601			
			0.8	0.0316	0.0240	0.0354	0.0316	0.0363	0.0599			
				0.0320	0.0241	0.0361	0.0320	0.0377	0.0615			
			2000	800	1200	0.3	0.0214	0.0159	0.0238	0.0214	0.0247	0.0414
							0.0220	0.0163	0.0256	0.0220	0.0256	0.0425
						0.5	0.0219	0.0164	0.0243	0.0219	0.0252	0.0419
							0.0228	0.0178	0.0268	0.0228	0.0253	0.0434
0.7	0.0224	0.0168				0.0248	0.0224	0.0257	0.0423			
	0.0215	0.0192				0.0272	0.0215	0.0271	0.0436			
0.8	0.0226	0.0171				0.0250	0.0226	0.0259	0.0425			
	0.0231	0.0171				0.0267	0.0231	0.0283	0.0433			

Table 6. Theoretical (**bold**) and Empirical MSE Comparisons of μ_y , μ_{AR} and μ_{MR} for $\rho_{xy} = 0.2$ and $\rho_{xy} = 0.7$ using **Unequal Sample Split** ($\theta_1 = -6$, $\theta_2 = -0.6$, $n_1 > n_2$).

n	n_1	n_2	W	MSE					
				$p = 0.7$			$p = 0.2$		
				$\hat{\mu}_y$	$\hat{\mu}_{AR}$	$\hat{\mu}_{MR}$	$\hat{\mu}_y$	$\hat{\mu}_{AR}$	$\hat{\mu}_{MR}$
500	300	200	0.3	0.1009	0.0763	0.1153	0.1009	0.1167	0.1957
				0.0962	0.0758	0.1153	0.0962	0.1158	0.1960
			0.5	0.1024	0.0778	0.1163	0.1023	0.1182	0.1972
				0.0982	0.0769	0.1159	0.0982	0.1162	0.1982
			0.7	0.1036	0.0790	0.1180	0.1035	0.1193	0.1984
				0.1013	0.0777	0.1158	0.1013	0.1157	0.1976
			0.8	0.1040	0.0794	0.1184	0.1040	0.1198	0.1988
				0.1007	0.0790	0.1181	0.1007	0.1182	0.2003
1000	600	400	0.3	0.0484	0.0364	0.0546	0.0483	0.0557	0.0932
				0.0478	0.0365	0.0561	0.0478	0.0560	0.0942
			0.5	0.0491	0.0371	0.0553	0.0490	0.0565	0.0939
				0.0492	0.0368	0.0570	0.0492	0.0565	0.0971
			0.7	0.0492	0.0376	0.0559	0.0492	0.0570	0.0945
				0.0482	0.0376	0.0583	0.0482	0.0571	0.0946
			0.8	0.0498	0.0379	0.0561	0.0498	0.0572	0.0947
				0.0490	0.0382	0.0575	0.0490	0.0569	0.0900
1500	900	600	0.3	0.0308	0.0231	0.0344	0.0308	0.0354	0.0590
				0.0309	0.0240	0.0361	0.0309	0.0359	0.0600
			0.5	0.0313	0.0235	0.0348	0.0313	0.0359	0.0595
				0.0283	0.0249	0.0355	0.0283	0.0369	0.0621
			0.7	0.0316	0.0239	0.0352	0.0316	0.0362	0.0598
				0.0316	0.0242	0.0368	0.0316	0.0372	0.0631
			0.8	0.0318	0.0240	0.0353	0.0317	0.0364	0.0600
				0.0325	0.0251	0.0371	0.0325	0.0381	0.0626
2000	1200	800	0.3	0.0221	0.0164	0.0242	0.0220	0.0253	0.0419
				0.0220	0.0168	0.0255	0.0220	0.0259	0.0435
			0.5	0.0224	0.0167	0.0246	0.0224	0.0256	0.0423
				0.0199	0.0173	0.0267	0.0199	0.0292	0.0445
			0.7	0.0226	0.0170	0.0248	0.0226	0.0259	0.0425
				0.0210	0.0186	0.0259	0.0210	0.0283	0.0451
			0.8	0.0227	0.0171	0.0249	0.0227	0.0260	0.0426
				0.0246	0.0176	0.0283	0.0246	0.0296	0.0463

CHAPTER III

IMPROVED MULTIPLICATIVE RATIO ESTIMATOR OF MEAN

3.1 Introduction

In Chapter 2, we presented the ordinary RRT estimator under Gupta et al. (2010) [13] additive optional RRT model. The mean estimator and the corresponding mean square error (MSE) are given below:

$$\hat{\mu}_y = \frac{\theta_2 \bar{Z}_1 - \theta_1 \bar{Z}_2}{\theta_2 - \theta_1}, \quad (\theta_1 \neq \theta_2), \quad (3.1)$$

and

$$MSE(\hat{\mu}_y) = \frac{1}{(\theta_2 - \theta_1)^2} [\theta_2^2 (\frac{1-f_1}{n_1}) \sigma_{Z_1}^2 + \theta_1^2 (\frac{1-f_2}{n_2}) \sigma_{Z_2}^2], \quad \theta_1 \neq \theta_2. \quad (3.2)$$

We also presented two ratio estimators by Kalucha et al. (2015) [15]. Kalucha et al. (2015) were motivated by a ratio estimator under a non-optional RRT model by Sousa et al. (2010) [24]. The two ratio estimators were the additive ratio estimator and the multiplicative ratio estimator respectively. These estimators with associated MSE's are given below:

$$\hat{\mu}_{AR} = \left(\frac{\theta_2 \bar{z}_1 - \theta_1 \bar{z}_2}{\theta_2 - \theta_1} \right) \left(\frac{\mu_x}{\bar{x}_1} + \frac{\mu_x}{\bar{x}_2} \right) \left(\frac{1}{2} \right), \quad (3.3)$$

$$\hat{\mu}_{MR} = \left(\frac{\theta_2 \bar{z}_1 - \theta_1 \bar{z}_2}{\theta_2 - \theta_1} \right) \left(\frac{\mu_x}{\bar{x}_1} \right) \left(\frac{\mu_x}{\bar{x}_2} \right), \quad (3.4)$$

$$\begin{aligned} MSE(\hat{\mu}_{MR}) &= E(\mu_{MR} - \mu_Y)^2 \\ &\approx \left(\frac{1-f_1}{n_1} \right) \left[\left(\frac{\theta_2}{\theta_2 - \theta_1} \right)^2 \sigma_{z_1}^2 + \mu_Y^2 C_x^2 - 2\mu_Y \rho_{yx} \sigma_Y \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) C_x \right] \\ &\quad + \left(\frac{1-f_2}{n_2} \right) \left[\left(\frac{\theta_1}{\theta_2 - \theta_1} \right)^2 \sigma_{z_2}^2 + \mu_Y^2 C_x^2 + 2\mu_Y \rho_{yx} \sigma_Y \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) C_x \right], \end{aligned} \quad (3.5)$$

and

$$\begin{aligned} MSE(\hat{\mu}_{AR}) &= E(\mu_{AR} - \mu_Y)^2 \\ &\approx \left(\frac{1-f_1}{n_1} \right) \left[\left(\frac{\theta_2}{\theta_2 - \theta_1} \right)^2 \sigma_{z_1}^2 + \frac{1}{4} \mu_Y^2 C_x^2 - \mu_Y \rho_{yx} \sigma_Y \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) C_x \right] \\ &\quad + \left(\frac{1-f_2}{n_2} \right) \left[\left(\frac{\theta_1}{\theta_2 - \theta_1} \right)^2 \sigma_{z_2}^2 + \frac{1}{4} \mu_Y^2 C_x^2 + \mu_Y \rho_{yx} \sigma_Y \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) C_x \right]. \end{aligned} \quad (3.6)$$

Kalucha et al. (2015) showed that the additive ratio estimator $\hat{\mu}_{AR}$ is more efficient than the ordinary RRT estimator when the correlation between the study variable and the auxiliary variable is greater than $\frac{1}{2}$. But the multiplicative ratio estimator is not as efficient as the ordinary RRT estimator and the additive ratio estimator. One can note that their additive ratio estimator uses the arithmetic mean of the two ratios in the sub-samples. But the multiplicative ratio estimator simply multiplies the two ratios of the sub-sample. It is reasonable to think that if we take the geometric mean of the two ratios, the efficiency of the estimator will be improved. Based on this, we propose a new ratio estimator. The basic idea of the geometric mean ratio estimator and the corresponding derivations of bias and MSE are provided in

Section 3.2. Theoretical efficiency comparison is presented in Section 3.3, However, the empirical comparisons are deferred until Chapter 4.

3.2 The Proposed Multiplicative Ratio Estimator

Suppose the sensitive study variable Y has mean μ_y , the non-sensitive auxiliary variable X has mean μ_x , and the correlation coefficient between X and Y is ρ_{xy} . Let the two scrambling variable means in the two sub-samples be θ_1 and θ_2 , and variances be σ_1^2 and σ_2^2 , respectively. Let the reported responses Z_i ($i=1,2$) have sample means \bar{z}_1 and \bar{z}_2 in the two sub-samples. We assume μ_x is known. The proposed mean ratio estimator is

$$\hat{\mu}_{GMR} = \left(\frac{\theta_2 \bar{z}_1 - \theta_1 \bar{z}_2}{\theta_2 - \theta_1} \right) \sqrt{\left(\frac{\mu_x}{\bar{x}_1} \right) \left(\frac{\mu_x}{\bar{x}_2} \right)} \quad (3.7)$$

A large sample size is assumed so that $|\delta_{z_i}| < 1$ and $|\delta_{x_i}| < 1$ ($i=1,2$), where $\delta_{Z_1} = \frac{\bar{z}_1 - \mu_{Z_1}}{\mu_{Z_1}}$, $\delta_{Z_2} = \frac{\bar{z}_2 - \mu_{Z_2}}{\mu_{Z_2}}$, $\delta_{x_1} = \frac{\bar{x}_1 - \mu_X}{\mu_X}$, $\delta_{x_2} = \frac{\bar{x}_2 - \mu_X}{\mu_X}$. The proposed ratio estimators (3.7) can be written as

$$\hat{\mu}_{GMR} = \left[\mu_y + \mu_{z_1} \delta_{z_1} \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) - \mu_{z_2} \delta_{z_2} \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) \right] (1 + \delta_{x_1})^{-1} (1 + \delta_{x_2})^{-1}. \quad (3.8)$$

Using first order Taylor's approximation and retaining terms of order up to 2, the difference between the geometric mean ratio estimator and the true mean can be

written as

$$\begin{aligned} \hat{\mu}_{GMR} - \mu_y = & \mu_y \left(-\frac{1}{2}\delta_{x_2} + \frac{3}{8}\delta_{x_2}^2 - \frac{1}{2}\delta_{x_1} + \frac{1}{4}\delta_{x_1}\delta_{x_2} + \frac{3}{8}\delta_{x_1}^2 \right) + \mu_{z_1} \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) \\ & \left(\delta_{z_1} - \frac{1}{2}\delta_{z_1}\delta_{x_2} - \frac{1}{2}\delta_{z_1}\delta_{x_1} \right) - \mu_{z_2} \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) \left(\delta_{z_2} - \frac{1}{2}\delta_{z_2}\delta_{x_2} - \frac{1}{2}\delta_{z_2}\delta_{x_1} \right). \end{aligned} \quad (3.9)$$

Under the assumption of bivariate normality (Sukhatme et al. 1970): $E(\delta_{x_1}) = 0$; $E(\delta_{x_2}) = 0$; $E(\delta_{x_1}^2) = (\frac{1-f_1}{n_1})C_{x_1}^2$; $E(\delta_{x_2}^2) = (\frac{1-f_2}{n_2})C_{x_2}^2$; $E(\delta_{x_1}\delta_{x_2}) = 0$; $E(\delta_{z_1}) = 0$; $E(\delta_{z_2}) = 0$; $E(\delta_{z_1}\delta_{x_2}) = 0$; $E(\delta_{z_2}\delta_{x_1}) = 0$; $E(\delta_{z_1}\delta_{x_1}) = (\frac{1-f_1}{n_1})C_{z_1x_1}$; $E(\delta_{z_2}\delta_{x_2}) = (\frac{1-f_2}{n_2})C_{z_2x_2}$, where $f_1 = \frac{n_1}{N}$, $f_2 = \frac{n_2}{N}$, $C_{z_1x_1} = \rho_{z_1x_1}C_{z_1}C_{x_1}$, and $C_{z_2x_2} = \rho_{z_2x_2}C_{z_2}C_{x_2}$. Note that the correlations $\rho_{z_ix_i}$ (i=1,2) are given by

$$\rho_{z_ix_i} = \frac{Cov(z_i, x_i)}{\sqrt{Var(z_i)}\sqrt{Var(x_i)}} = \frac{\sigma_{z_ix_i}}{\sigma_{z_i}\sigma_{x_i}} = \frac{\sigma_{x_iy}}{\sigma_{z_i}\sigma_{x_i}} = \frac{\rho_{x_iy}\sigma_y\sigma_{x_i}}{\mu_{z_i}C_{z_i}\sigma_{x_i}} = \frac{\rho_{xy}\sigma_y}{\mu_{z_i}C_{z_i}}, \quad (3.10)$$

also, $C_{x_1} = C_{x_2} = C_x$, $\rho_{z_1x_1} = \rho_{z_1x}$, and $\rho_{z_2x_2} = \rho_{z_2x}$.

Taking expectation on both sides of (3.9), the bias of the geometric mean ratio estimator $\hat{\mu}_{GMR}$, correct to first order of approximation, is given by

$$\begin{aligned} Bias(\hat{\mu}_{GMR}) \approx & \frac{3}{8}\mu_y \left[\left(\frac{1-f_1}{n_1} \right) C_{x_1}^2 + \left(\frac{1-f_2}{n_2} \right) C_{x_2}^2 \right] - \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) \frac{1}{2} \left(\frac{1-f_1}{n_1} \right) \rho_{z_1x_1} C_{z_1} C_{x_1} \\ & + \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) \frac{1}{2} \left(\frac{1-f_2}{n_2} \right) \rho_{z_2x_2} C_{z_2} C_{x_2} \\ = & \frac{3}{8}\mu_y \left[\left(\frac{1-f_1}{n_1} \right) C_x^2 + \left(\frac{1-f_2}{n_2} \right) C_x^2 \right] - \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) \frac{1}{2} \left(\frac{1-f_1}{n_1} \right) \rho_{xy}\sigma_y C_x \\ & + \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) \frac{1}{2} \left(\frac{1-f_2}{n_2} \right) \rho_{xy}\sigma_y C_x. \end{aligned} \quad (3.11)$$

To determine the expression for MSE of the geometric mean ratio estimator $\hat{\mu}_{GMR}$, we use Taylor's approximation retaining terms of order up to 2. With this

$$\begin{aligned}
(\hat{\mu}_{GMR} - \mu_y)^2 &\approx \mu_y^2 \left(\frac{1}{4} \delta_{x_2}^2 + \frac{1}{4} \delta_{x_1}^2 + \frac{1}{4} \delta_{x_1} \delta_{x_2} \right) + \mu_{z_1}^2 \left(\frac{\theta_2}{\theta_2 - \theta_1} \right)^2 \delta_{z_1}^2 + \mu_{z_2}^2 \left(\frac{\theta_1}{\theta_2 - \theta_1} \right)^2 \delta_{z_2}^2 \\
&\quad + 2\mu_y \mu_{z_1} \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) \left(-\frac{1}{2} \delta_{x_2} \delta_{z_1} - \frac{1}{2} \delta_{x_1} \delta_{z_1} \right) - 2\mu_y \mu_{z_2} \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) \\
&\quad \left(-\frac{1}{2} \delta_{x_2} \delta_{z_2} - \frac{1}{2} \delta_{x_1} \delta_{z_2} \right) - \mu_{z_1} \mu_{z_2} \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) (\delta_{z_1} \delta_{z_2}).
\end{aligned} \tag{3.12}$$

Then using $E(\delta_{x_1}^2) = \left(\frac{1-f_1}{n_1}\right)C_{x_1}^2$; $E(\delta_{x_2}^2) = \left(\frac{1-f_2}{n_2}\right)C_{x_2}^2$; $E(\delta_{x_1}\delta_{x_2}) = 0$; $E(\delta_{z_1}^2) = \left(\frac{1-f_1}{n_1}\right)C_{z_1}^2$; $E(\delta_{z_2}^2) = \left(\frac{1-f_2}{n_2}\right)C_{z_2}^2$; $E(\delta_{z_1}\delta_{z_2}) = 0$; $E(\delta_{z_1}\delta_{x_1}) = \left(\frac{1-f_1}{n_1}\right)\rho_{z_1x_1}C_{z_1}C_{x_1}$; $E(\delta_{z_2}\delta_{x_2}) = \left(\frac{1-f_2}{n_2}\right)\rho_{z_2x_2}C_{z_2}C_{x_2}$; $E(\delta_{z_2}\delta_{x_1}) = 0$; $E(\delta_{z_1}\delta_{x_2}) = 0$. Taking the expected value on both sides of (3.12), the expression for MSE of $\hat{\mu}_{GMR}$, correct to first order of approximation, is given by

$$\begin{aligned}
MSE(\hat{\mu}_{GMR}) &\approx \frac{1}{4} \mu_y^2 \left[\left(\frac{1-f_1}{n_1} \right) C_{x_1}^2 + \left(\frac{1-f_2}{n_2} \right) C_{x_2}^2 \right] + \mu_{z_1}^2 \left(\frac{\theta_2}{\theta_2 - \theta_1} \right)^2 \left(\frac{1-f_1}{n_1} \right) C_{z_1}^2 \\
&\quad + \mu_{z_2}^2 \left(\frac{\theta_1}{\theta_2 - \theta_1} \right)^2 \left(\frac{1-f_2}{n_2} \right) C_{z_2}^2 + 2\mu_y \mu_{z_1} \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) \left[-\frac{1}{2} \left(\frac{1-f_1}{n_1} \right) \rho_{z_1x_1} C_{z_1} C_{x_1} \right] \\
&\quad - 2\mu_y \mu_{z_2} \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) \left[-\frac{1}{2} \left(\frac{1-f_1}{n_2} \right) \rho_{z_2x_2} C_{z_2} C_{x_2} \right] \\
&= \left(\frac{1-f_1}{n_1} \right) \left[\left(\frac{\theta_2}{\theta_2 - \theta_1} \right)^2 \mu_{z_1}^2 C_{z_1}^2 + \frac{1}{4} \mu_y^2 C_{x_1}^2 - \mu_y \mu_{z_1} \left(\frac{\theta_2}{\theta_2 - \theta_1} \right) \rho_{z_1x_1} C_{z_1} C_{x_1} \right] \\
&\quad + \left(\frac{1-f_2}{n_2} \right) \left[\left(\frac{\theta_1}{\theta_2 - \theta_1} \right)^2 \mu_{z_2}^2 C_{z_2}^2 + \frac{1}{4} \mu_y^2 C_{x_2}^2 - \mu_y \mu_{z_2} \left(\frac{\theta_1}{\theta_2 - \theta_1} \right) \rho_{z_2x_2} C_{z_2} C_{x_2} \right].
\end{aligned} \tag{3.13}$$

This can be rewritten as

$$\begin{aligned}
MSE(\hat{\mu}_{GMR}) &= E[(\hat{\mu}_{GMR} - \mu_Y)^2] \\
&\approx \left(\frac{1-f_1}{n_1}\right) \left[\left(\frac{\theta_2}{\theta_2 - \theta_1}\right)^2 \sigma_{z_1}^2 + \frac{1}{4} \mu_Y^2 C_x^2 - \mu_Y \rho_{yx} \sigma_Y \left(\frac{\theta_2}{\theta_2 - \theta_1}\right) C_x \right] \\
&\quad + \left(\frac{1-f_2}{n_2}\right) \left[\left(\frac{\theta_1}{\theta_2 - \theta_1}\right)^2 \sigma_{z_2}^2 + \frac{1}{4} \mu_Y^2 C_x^2 + \mu_Y \rho_{yx} \sigma_Y \left(\frac{\theta_1}{\theta_2 - \theta_1}\right) C_x \right].
\end{aligned} \tag{3.14}$$

3.3 Efficiency Comparisons

3.3.1 MSE Comparisons

- (1) The difference between the approximate MSE of the geometric mean ratio estimator (3.14) and the multiplicative ratio estimator (3.5) is given by

$$\begin{aligned}
MSE^{(1)}(\hat{\mu}_{GMR}) - MSE^{(1)}(\hat{\mu}_{MR}) &= -\frac{3}{4} \left(\frac{1-f_1}{n_1}\right) \mu_y^2 C_x^2 - \frac{3}{4} \left(\frac{1-f_2}{n_2}\right) \mu_y^2 C_x^2 + \mu_y C_x \\
&\quad \rho_{xy} \sigma_y \left[\left(\frac{1-f_1}{n_1}\right) \left(\frac{\theta_2}{\theta_2 - \theta_1}\right) - \left(\frac{1-f_2}{n_2}\right) \left(\frac{\theta_1}{\theta_2 - \theta_1}\right) \right].
\end{aligned} \tag{3.15}$$

Thus, $MSE^{(1)}(\hat{\mu}_{GMR}) < MSE^{(1)}(\hat{\mu}_{MR})$ if

$$\begin{aligned}
&-\frac{3}{4} \left[\left(\frac{1-f_1}{n_1}\right) + \left(\frac{1-f_2}{n_2}\right) \right] + \\
&\frac{C_y}{C_x} \rho_{xy} \left[\left(\frac{1-f_1}{n_1}\right) \left(\frac{\theta_2}{\theta_2 - \theta_1}\right) - \left(\frac{1-f_2}{n_2}\right) \left(\frac{\theta_1}{\theta_2 - \theta_1}\right) \right] < 0.
\end{aligned} \tag{3.16}$$

- (2) The difference between approximate MSE of the geometric mean ratio estimator (3.14) and the ordinary RRT estimator (3.1) is given by

$$\begin{aligned}
MSE^{(1)}(\hat{\mu}_{GMR}) - MSE^{(1)}(\hat{\mu}_y) = & \left(\frac{1-f_1}{n_1}\right) \left[\frac{1}{4}\mu_y^2 C_x^2 - \mu_y \rho_{xy} \sigma_y C_x \left(\frac{\theta_2}{\theta_2 - \theta_1}\right)\right] \\
& + \left(\frac{1-f_2}{n_2}\right) \left[\frac{1}{4}\mu_y^2 C_x^2 - \mu_y \rho_{xy} \sigma_y C_x \left(\frac{\theta_1}{\theta_2 - \theta_1}\right)\right].
\end{aligned} \tag{3.17}$$

Thus, $MSE^{(1)}(\hat{\mu}_{GMR}) < MSE^{(1)}(\hat{\mu}_y)$ if

$$\left(\frac{1-f_1}{n_1}\right) \left[\frac{1}{4} - \rho_{xy} \frac{C_y}{C_x} \left(\frac{\theta_2}{\theta_2 - \theta_1}\right)\right] + \left(\frac{1-f_2}{n_2}\right) \left[\frac{1}{4} + \rho_{xy} \frac{C_y}{C_x} \left(\frac{\theta_1}{\theta_2 - \theta_1}\right)\right] < 0, \tag{3.18}$$

which can be rewritten as

$$\begin{aligned}
& \frac{1}{4} \left[\left(\frac{1-f_1}{n_1}\right) + \left(\frac{1-f_2}{n_2}\right) \right] - \rho_{xy} \frac{C_y}{C_x} \left[\left(\frac{1-f_1}{n_1}\right) \left(\frac{\theta_2}{\theta_2 - \theta_1}\right) - \right. \\
& \left. \left(\frac{1-f_2}{n_2}\right) \left(\frac{\theta_1}{\theta_2 - \theta_1}\right) \right] < 0.
\end{aligned} \tag{3.19}$$

- (3) We can also note that the approximate MSE's of the geometric mean ratio estimator (3.14) and the additive ratio estimator (3.6) are equal. That is,

$$MSE^{(1)}(\hat{\mu}_{GMR}) - MSE^{(1)}(\hat{\mu}_{AR}) = 0. \tag{3.20}$$

Since the approximate MSE of the geometric estimator, up to first order of approximation, is the same as the MSE of the additive ratio estimator, we will only compare

it with the multiplicative ratio estimator and the ordinary RRT estimator. In order to make the above expressions simpler, we let

$$\alpha = \left(\frac{1-f_1}{n_1}\right) + \left(\frac{1-f_2}{n_2}\right) \quad (3.21)$$

and

$$\beta = \left(\frac{1-f_1}{n_1}\right)\left(\frac{\theta_2}{\theta_2-\theta_1}\right) - \left(\frac{1-f_2}{n_2}\right)\left(\frac{\theta_1}{\theta_2-\theta_1}\right). \quad (3.22)$$

Also, we assume the coefficient of variations of X and Y are the same, that is, we assume

$$C_x = C_y. \quad (3.23)$$

Thus, the comparisons (3.16) and (3.19) can be written as:

$$(1) \text{ } MSE^{(1)}(\hat{\mu}_{GMR}) < MSE^{(1)}(\hat{\mu}_{MR}) \text{ if } -\frac{3}{4}\alpha + \rho_{xy}\beta < 0$$

$$(2) \text{ } MSE^{(1)}(\hat{\mu}_{GMR}) < MSE^{(1)}(\hat{\mu}_y) \text{ if } \frac{1}{4}\alpha - \rho_{xy}\beta < 0$$

Since f_i ($i = 1, 2$) are less than one, $\alpha = \left(\frac{1-f_1}{n_1}\right) + \left(\frac{1-f_2}{n_2}\right)$ is always positive. Thus, whether or not the cases (1) and (2) can be established depends on the value of β . If $\beta < 0$, Case (1) is always true, but Case (2) will not be true. In other words, when $\beta < 0$, the geometric mean ratio estimator is more efficient than the multiplicative ratio estimator, but not be as efficient as the ordinary RRT estimator. The situation when $\beta > 0$ is considered below.

In order to make the expression of β more clear, we let

$$a = \left(\frac{1-f_1}{n_1}\right) \quad \text{and} \quad b = \left(\frac{1-f_2}{n_2}\right) \quad (3.24)$$

Thus,

$$\beta = a\left(\frac{\theta_2}{\theta_2-\theta_1}\right) - b\left(\frac{\theta_1}{\theta_2-\theta_1}\right). \quad (3.25)$$

It is easy to see that both a and b are greater than zero. Also, if $a > b$, then $n_1 < n_2$; and if $b > a$, then $n_1 > n_2$.

Various situations when $\beta > 0$ are presented below:

- $0 < \theta_1 < \theta_2$, $n_1 < n_2$. In this case, $a > b$ and $0 < \left(\frac{\theta_1}{\theta_2-\theta_1}\right) < \left(\frac{\theta_2}{\theta_2-\theta_1}\right)$. Thus, $\beta = a\left(\frac{\theta_2}{\theta_2-\theta_1}\right) - b\left(\frac{\theta_1}{\theta_2-\theta_1}\right) > 0$.
- $0 < \theta_2 < \theta_1$, $n_2 < n_1$. In this case, $a < b$ and $\left(\frac{\theta_1}{\theta_2-\theta_1}\right) < \left(\frac{\theta_2}{\theta_2-\theta_1}\right) < 0$. Thus, $\beta = a\left(\frac{\theta_2}{\theta_2-\theta_1}\right) - b\left(\frac{\theta_1}{\theta_2-\theta_1}\right) > 0$.
- $\theta_2 < \theta_1 < 0$, $n_1 < n_2$. In this case, $a > b$ and $0 < \left(\frac{\theta_1}{\theta_2-\theta_1}\right) < \left(\frac{\theta_2}{\theta_2-\theta_1}\right)$. Thus, $\beta = a\left(\frac{\theta_2}{\theta_2-\theta_1}\right) - b\left(\frac{\theta_1}{\theta_2-\theta_1}\right) > 0$.
- $\theta_1 < \theta_2 < 0$, $n_2 < n_1$. In this case, $a < b$ and $\left(\frac{\theta_1}{\theta_2-\theta_1}\right) < \left(\frac{\theta_2}{\theta_2-\theta_1}\right) < 0$. Thus, $\beta = a\left(\frac{\theta_2}{\theta_2-\theta_1}\right) - b\left(\frac{\theta_1}{\theta_2-\theta_1}\right) > 0$.
- $\theta_1 = 0$. In this case, $\beta = b > 0$.
- $\theta_2 = 0$. In this case, $\beta = a > 0$.

- $\theta_1 < 0 < \theta_2$. In this case, $a > 0$, $b > 0$, and $(\frac{\theta_1}{\theta_2 - \theta_1}) < 0 < (\frac{\theta_2}{\theta_2 - \theta_1})$ Thus, $\beta = a(\frac{\theta_2}{\theta_2 - \theta_1}) - b(\frac{\theta_1}{\theta_2 - \theta_1}) > 0$.
- $\theta_2 < 0 < \theta_1$. In this case, $a > 0$, $b > 0$, and $(\frac{\theta_1}{\theta_2 - \theta_1}) < 0 < (\frac{\theta_2}{\theta_2 - \theta_1})$ Thus, $\beta = a(\frac{\theta_2}{\theta_2 - \theta_1}) - b(\frac{\theta_1}{\theta_2 - \theta_1}) > 0$.

The situations when $\beta > 0$ can be summarized as follows:

- If two scrambling variable means are positive, then the scrambling variable with the smaller mean should associate with the smaller sub-sample size.
- If two scrambling variable means are negative, then the scrambling variable with the smaller mean should associate with the larger sub-sample size.
- If one of the scrambling variable means is 0 or two scrambling variable means have opposite signs, then there is no specific requirement for sub-sample size.

Therefore, we can rewrite (1) and (2), when $\alpha > 0$ and $\beta > 0$, as:

$$MSE^{(1)}(\hat{\mu}_{GMR}) < MSE^{(1)}(\hat{\mu}_{MR}) \quad \text{if} \quad \rho_{xy} < \frac{3\alpha}{4\beta} \quad (3.26)$$

$$MSE^{(1)}(\hat{\mu}_{GMR}) < MSE^{(1)}(\hat{\mu}_y) \quad \text{if} \quad \rho_{xy} > \frac{\alpha}{4\beta} \quad (3.27)$$

We explore these conditions under both cases of equal sample split and unequal sample split:

Equal Sample Split Case

If $n_1 = n_2$, then

$$\alpha = \left(\frac{1-f_1}{n_1}\right) + \left(\frac{1-f_2}{n_2}\right) = 2\left(\frac{1-f_1}{n_1}\right) > 0$$

and

$$\beta = \left(\frac{1-f_1}{n_1}\right)\left(\frac{\theta_2}{\theta_2-\theta_1}\right) - \left(\frac{1-f_2}{n_2}\right)\left(\frac{\theta_1}{\theta_2-\theta_1}\right) = \left(\frac{1-f_1}{n_1}\right) > 0$$

Therefore, $\frac{3\alpha}{4\beta} = \frac{3}{2}$ and $\frac{\alpha}{4\beta} = \frac{1}{2}$. Based on these, we can conclude the following for equal sample split:

- The condition $\rho_{xy} < \frac{3}{2}$ is always true, hence $MSE^{(1)}(\hat{\mu}_{GMR}) < MSE^{(1)}(\hat{\mu}_{MR})$. This means that the geometric mean ratio estimator is always more efficient than the multiplicative ratio estimator.
- If $\rho_{xy} > \frac{1}{2}$, $MSE^{(1)}(\hat{\mu}_{GMR}) < MSE^{(1)}(\hat{\mu}_y)$. This means that the geometric mean ratio estimator is more efficient than the ordinary RRT estimator when the correlation coefficient between X and Y is greater than $\frac{1}{2}$.

Unequal Sample Split Case

If $n_1 \neq n_2$, then $\frac{\alpha}{4\beta} < \frac{1}{2}$. Since $\alpha = a + b$ and $\beta = a\left(\frac{\theta_2}{\theta_2-\theta_1}\right) + b\left(\frac{\theta_1}{\theta_2-\theta_1}\right)$ ($\beta > 0$), it can be written as:

$$(a + b) < 2\left(\frac{a\theta_2 - b\theta_1}{\theta_2 - \theta_1}\right), \quad \beta > 0. \quad (3.28)$$

The conditions that satisfy the above inequality (3.28) and $\beta > 0$ are:

- $0 \leq \theta_1 < \theta_2, n_1 < n_2$
- $\theta_1 < 0 < \theta_2, |\theta_1| < \theta_2, n_1 < n_2$
- $\theta_1 < \theta_2 \geq 0, n_2 < n_1$
- $\theta_1 < 0 < \theta_2, |\theta_1| > \theta_2, n_2 < n_1$
- $0 \leq \theta_2 < \theta_1, n_2 < n_1$
- $\theta_2 < 0 < \theta_1, \theta_1 > |\theta_2|, n_2 < n_1$
- $\theta_2 < \theta_1 \leq 0, n_1 < n_2$
- $\theta_2 < 0 < \theta_1, \theta_1 < |\theta_2|, n_1 < n_2$

These conditions can be summarized as follows:

- If the two scrambling variable means are positive, then the scrambling variable with the smaller mean should associate with the smaller sub-sample size.
- If the two scrambling variable means are negative, then the scrambling variable with the smaller mean should associate with the larger sub-sample size.
- If the two scrambling variable means have opposite signs, then the scrambling variable with the larger absolute mean should associate with the larger sub-sample size.
- If one of the two scrambling variable means is zero, then the scrambling variable with the mean of zero should associate with the smaller sub-sample size.

If the parameters satisfy the above conditions for unequal sample split, the geometric mean ratio estimator is more efficient than the ordinary RRT estimator when the correlation coefficient ρ_{xy} is greater than $\frac{1}{2}$. Also note that $\frac{\alpha}{4\beta} < \frac{1}{2} \Leftrightarrow \frac{3\alpha}{4\beta} < \frac{3}{2}$. Therefore, we can conclude in the unequal sample split case that

$$(1) \text{ } MSE^{(1)}(\hat{\mu}_{GMR}) < MSE^{(1)}(\hat{\mu}_{MR}) \text{ if } \rho_{xy} < \frac{3}{2} \text{ (which is always true)}$$

$$(2) \text{ } MSE^{(1)}(\hat{\mu}_{GMR}) < MSE^{(1)}(\hat{\mu}_y) \text{ if } \rho_{xy} > \frac{1}{2}$$

In other words,

(1) If $\rho_{xy} < \frac{3}{2}$, which is always true, $MSE^{(1)}(\hat{\mu}_{GMR}) < MSE^{(1)}(\hat{\mu}_{MR})$. This means the geometric mean ratio estimator is always more efficient than the multiplicative ratio estimator.

(2) If $\rho_{xy} > \frac{1}{2}$, $MSE^{(1)}(\hat{\mu}_{GMR}) < MSE^{(1)}(\hat{\mu}_y)$. This means the geometric mean ratio estimator is more efficient than the ordinary RRT estimator when the correlation coefficient between X and Y is greater than $\frac{1}{2}$.

3.3.2 Bias Comparisons

Since approximate MSE, up to the first order of approximation, of the geometric mean ratio estimator is the same as that of the additive ratio estimator, the biases of these two estimators (3.11) (2.12) are compared.

$$Bias^{(1)}(\hat{\mu}_{GMR}) - Bias^{(1)}(\hat{\mu}_{AR}) = -\frac{1}{8}\mu_y C_x^2 \left[\left(\frac{1-f_1}{n_1} \right) + \left(\frac{1-f_2}{n_2} \right) \right] < 0 \quad (3.29)$$

Thus, the geometric mean ratio estimator has smaller bias than the additive ratio estimator.

3.4 Concluding Chapter Remarks

In this chapter, a geometric mean ratio estimator is proposed and its bias and MSE are derived, up to first order approximation. Also, its efficiency is compared with the multiplicative ratio estimator, the ordinary RRT estimator and the additive ratio estimator. The results showed that for both equal and unequal sample split cases, the geometric mean ratio estimator is more efficient than the multiplicative ratio estimator; it is more efficient than the ordinary RRT ratio estimator when the correlation coefficient between the sensitive variable Y and the auxiliary variable X is greater than $\frac{1}{2}$; the efficiency of the geometric mean ratio estimator coincides with the additive ratio estimator but it has smaller bias as compared to the additive ratio estimator. All the derivations are based on the first order approximation.

CHAPTER IV
SIMULATIONS

In this chapter, we conduct a simulation study focusing on comparisons of the theoretical and the empirical MSEs of our proposed geometric mean ratio estimator $\hat{\mu}_{GMR}$ with other RRT estimators. We consider two bivariate normal distributions for $[X, Y]$ with the same mean and variances but different correlation coefficients ($\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$). The two covariance matrices Σ are as given below:

$$\text{Population 1 } \Sigma = \begin{bmatrix} 5 & 6.26 \\ 6.26 & 16 \end{bmatrix}$$

$$\text{Population 2 } \Sigma = \begin{bmatrix} 5 & 1.79 \\ 1.79 & 16 \end{bmatrix}$$

To explain the simulation process further, we started with two samples of size 5000 each from two bivariate normal populations with parameters

$$\begin{aligned} \mu_{x_1} = 5, \sigma_{x_1}^2 = 5, \mu_{y_1} = 9, \sigma_{y_1}^2 = 16, \rho_{xy_1} = 0.7 \text{ and} \\ \mu_{x_2} = 5, \sigma_{x_2}^2 = 5, \mu_{y_2} = 9, \sigma_{y_2}^2 = 16, \rho_{xy_2} = 0.2 \end{aligned} \tag{A}$$

Then we used these two sets of 5000 data points as our finite populations. The means, variances and correlation coefficients for these 5000 data points given below in (B) are very close to the parameter values we used to generate these data but not exactly same.

$$\begin{aligned} \mu_{x_1} = 4.9941, \sigma_{x_1}^2 = 4.9284, \mu_{y_1} = 9.0095, \sigma_{y_1}^2 = 16.0236, \rho_{xy_1} = 0.7029 \text{ and} \\ \mu_{x_2} = 4.9873, \sigma_{x_2}^2 = 4.8852, \mu_{y_2} = 9.0079, \sigma_{y_2}^2 = 16.0141, \rho_{xy_2} = 0.1974 \end{aligned} \quad (\text{B})$$

For the simulation study, we used parameter values in (B) and not those in (A).

The scrambling variables S_1 and S_2 are taken from two normal distributions with variances $\sigma_{S_1}^2 = 2$ and $\sigma_{S_2}^2 = 1$, respectively. The means of the scrambling variables will be described later as per the guidelines in Section 3.3, depending on how the sample is split (equal or unequal). We used sample sizes $n=500, 1000, 1500$, and 2000 . The results are averaged over 1000 trials.

4.1 Simulation with Equal Sub-samples

For the equal sample split case, the two scrambling variable means used are $\theta_1 = 6$ and $\theta_2 = 0.6$. Table 7 presents the sensitive level estimator \hat{W} , the ordinary mean estimator $\hat{\mu}_y$, the additive ratio estimator $\hat{\mu}_{AR}$ and the geometric mean ratio estimator $\hat{\mu}_{GMR}$ under the optional RRT model for different correlation coefficients ($\rho_{xy} = 0.7, 0.2$) and different sensitive levels ($W = 0.2, 0.5, 0.8$). All the estimators are fairly accurate compared with true parameter values.

Table 8 presents the MSE comparisons for the four mean estimators. The results show that the geometric mean ratio estimator is more efficient than the ordinary RRT estimator when the correlation coefficient $\rho_{xy} = 0.7$ (which is greater than $\frac{1}{2}$),

but not as efficient as the ordinary RRT estimator when $\rho_{xy} = 0.2 < \frac{1}{2}$. Also, the MSE values of the geometric mean ratio estimator are very similar to the additive ratio estimator MSE, but are all smaller than the MSE of the multiplicative ratio estimator. Thus, the simulation results for the equal sub-samples case confirm our idea that using the geometric mean ratio estimator will produce better results. In addition, as the sensitivity level W increases, the MSE of each estimator increases. This indicates that the optional RRT models are more efficient than the non-optional RRT models.

Table 7. Point Estimates of W , μ_y , μ_{AR} , μ_{MR} and μ_{GMR} , for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using **Equal Sample Split** ($\theta_1 = 6$, $\theta_2 = 0.6$, $n_1 = n_2$).

n	n_1	n_2		$W = 0.2$		$W = 0.5$		$W = 0.8$	
				\hat{W}	Estimate	\hat{W}	Estimate	\hat{W}	Estimate
$\rho_{xy} = 0.7$									
500	300	200	μ_y	0.2011	9.0050	0.4997	9.0049	0.7983	9.0049
			μ_{AR}	0.2003	9.0038	0.5006	9.0021	0.7998	9.0021
			μ_{MR}	0.1996	9.0011	0.5011	9.0016	0.8976	9.0016
			μ_{GMR}	0.1976	8.9869	0.5006	9.0063	0.7999	9.0038
1000	600	400	μ_y	0.2004	8.9999	0.5009	9.0050	0.7998	9.0016
			μ_{AR}	0.2013	8.9898	0.5026	8.9956	0.8976	9.0049
			μ_{MR}	0.1987	9.0127	0.4985	9.0038	0.7999	9.0021
			μ_{GMR}	0.2004	9.0023	0.4972	9.0023	0.7993	8.9788
1500	900	600	μ_y	0.2013	9.0061	0.5031	8.9955	0.7998	9.0021
			μ_{AR}	0.1987	9.0003	0.5029	8.9749	0.7982	9.0016
			μ_{MR}	0.2009	9.1008	0.5002	9.0020	0.8016	8.9875
			μ_{GMR}	0.1987	9.0050	0.4962	9.0032	0.8020	9.0072
2000	1200	800	μ_y	0.2011	9.0038	0.5013	9.0041	0.8015	9.0047
			μ_{AR}	0.1996	8.9955	0.5009	9.0022	0.7984	9.0017
			μ_{MR}	0.2002	8.9994	0.5013	8.8998	0.7998	9.0038
			μ_{GMR}	0.2016	9.0050	0.4983	9.0013	0.8001	8.9981
$\rho_{xy} = 0.2$									
$\rho_{xy} = 0.2$									
500	300	200	μ_y	0.2011	9.0050	0.4997	9.0049	0.7983	9.0049
			μ_{AR}	0.2014	8.9768	0.5013	9.0032	0.7993	9.0056
			μ_{MR}	0.2011	8.9749	0.5007	9.1008	0.7995	8.9955
			μ_{GMR}	0.1987	9.0021	0.5003	9.0061	0.7984	8.9994
1000	600	400	μ_y	0.2004	8.9999	0.5009	9.0050	0.7998	9.0016
			μ_{AR}	0.1996	9.0049	0.5007	9.0020	0.8006	9.0016
			μ_{MR}	0.1986	9.0055	0.4974	9.0016	0.7982	9.0025
			μ_{GMR}	0.2012	8.9749	0.5007	9.0049	0.7995	9.0061
1500	900	600	μ_y	0.2013	9.0061	0.5031	8.9955	0.7998	9.0021
			μ_{AR}	0.2004	8.9955	0.4974	9.0050	0.8019	9.0050
			μ_{MR}	0.2013	8.9994	0.5007	9.0038	0.7982	9.0028
			μ_{GMR}	0.1987	9.0050	0.5013	9.0011	0.8004	9.0060
2000	1200	800	μ_y	0.2011	9.0038	0.5013	9.0041	0.8015	9.0047
			μ_{AR}	0.2009	9.1008	0.5013	8.9999	0.7999	8.9955
			μ_{MR}	0.1987	9.0061	0.4987	9.0043	0.7993	9.0043
			μ_{GMR}	0.2011	9.0016	0.5003	9.0050	0.8005	9.0016

Table 8. Theoretical (**bold**) and Empirical MSE Comparisons of μ_y , μ_{AR} , μ_{MR} and μ_{GMR} for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using **Equal Sample Split** ($\theta_1 = 6$, $\theta_2 = 0.6$, $n_1 = n_2$).

n	n_1	n_2		MSE					
				$W = 0.2$		$W = 0.5$		$W = 0.8$	
				$\rho_{xy} = 0.7$					
500	250	250	$\hat{\mu}_y$	0.0745	0.0754	0.0806	0.0771	0.0812	0.0784
			$\hat{\mu}_{AR}$	0.0623	0.0631	0.0685	0.0648	0.0699	0.0661
			$\hat{\mu}_{MR}$	0.1152	0.1117	0.1170	0.1134	0.1174	0.1147
			$\hat{\mu}_{GMR}$	0.0629	0.0631	0.0687	0.0648	0.0688	0.0661
1000	500	500	$\hat{\mu}_y$	0.0368	0.0357	0.0373	0.0365	0.0396	0.0371
			$\hat{\mu}_{AR}$	0.0311	0.0299	0.0324	0.0307	0.0340	0.0313
			$\hat{\mu}_{MR}$	0.0547	0.0529	0.0572	0.0537	0.0576	0.0543
			$\hat{\mu}_{GMR}$	0.0315	0.0299	0.0321	0.0307	0.0324	0.0313
1500	750	750	$\hat{\mu}_y$	0.0232	0.0225	0.0245	0.0230	0.0243	0.0234
			$\hat{\mu}_{AR}$	0.0201	0.0188	0.0219	0.0193	0.0209	0.0197
			$\hat{\mu}_{MR}$	0.0345	0.0333	0.0356	0.0338	0.0359	0.0342
			$\hat{\mu}_{GMR}$	0.0200	0.0188	0.0205	0.0193	0.0209	0.0197
2000	1000	1000	$\hat{\mu}_y$	0.0170	0.0159	0.0176	0.0162	0.0186	0.0165
			$\hat{\mu}_{AR}$	0.0129	0.0133	0.0157	0.0136	0.0164	0.0139
			$\hat{\mu}_{MR}$	0.0249	0.0235	0.0256	0.0239	0.0257	0.0241
			$\hat{\mu}_{GMR}$	0.0140	0.0133	0.0145	0.0136	0.0145	0.0139
				$\rho_{xy} = 0.2$					
500	250	250	$\hat{\mu}_y$	0.0745	0.0754	0.0806	0.0771	0.0812	0.0784
			$\hat{\mu}_{AR}$	0.0962	0.0937	0.0988	0.0954	0.1003	0.0967
			$\hat{\mu}_{MR}$	0.1765	0.1725	0.1781	0.1742	0.1790	0.1755
			$\hat{\mu}_{GMR}$	0.0965	0.0937	0.0985	0.0954	0.0992	0.0967
1000	500	500	$\hat{\mu}_y$	0.3868	0.0357	0.0373	0.0365	0.0396	0.0371
			$\hat{\mu}_{AR}$	0.0459	0.0444	0.0472	0.0452	0.0486	0.0458
			$\hat{\mu}_{MR}$	0.0835	0.0817	0.0846	0.0825	0.0859	0.0831
			$\hat{\mu}_{GMR}$	0.0459	0.0444	0.0468	0.0452	0.0473	0.0458
1500	750	750	$\hat{\mu}_y$	0.0232	0.0225	0.0245	0.0230	0.0243	0.0234
			$\hat{\mu}_{AR}$	0.0292	0.0279	0.0296	0.0284	0.0305	0.0288
			$\hat{\mu}_{MR}$	0.0530	0.0514	0.0547	0.0520	0.0546	0.0523
			$\hat{\mu}_{GMR}$	0.0295	0.0279	0.0306	0.0284	0.0309	0.0288
2000	1000	1000	$\hat{\mu}_y$	0.0170	0.0159	0.0176	0.0162	0.0186	0.0165
			$\hat{\mu}_{AR}$	0.0208	0.0197	0.0210	0.0201	0.0219	0.0203
			$\hat{\mu}_{MR}$	0.0375	0.0363	0.0383	0.0367	0.0392	0.0369
			$\hat{\mu}_{GMR}$	0.0209	0.0197	0.0213	0.0201	0.0214	0.0203

4.2 Simulations with Unequal Sub-samples

For the unequal sample split, we consider four cases that we mentioned in Chapter 3. They are

- (1) When both scrambling variable means are positive, the scrambling variable with the larger mean is associated with the larger sub-sample size.
- (2) When both scrambling variable means are negative, the scrambling variable with the smaller mean is associated with the larger sub-sample size.
- (3) When the two scrambling variable means have opposite signs, the scrambling variable with larger absolute mean value is associated with the larger sub-sample size.
- (4) When one of the scrambling variable means is zero, the scrambling variable with the mean of zero is associated with the smaller sub-sample size.

Two scrambling variables are used in Case (1) with means $\theta_1 = 0.6$ and $\theta_2 = 6$. The first sub-sample is associated with the smaller sub-sample size since $\theta_1 = 0.6$ is smaller. Table 9 presents the sensitivity level estimator \hat{W} , the ordinary RRT estimator μ_y , the additive ratio estimator μ_{AR} , the multiplicative ratio estimator μ_{MR} and the geometric mean ratio estimator μ_{GMR} . All the point estimates are fairly accurate under different correlation coefficients ρ_{xy} and different sensitivity levels W . Tables 10 shows the theoretical and empirical MSE comparisons for the four estimators when $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$.

Two scrambling variables are used in Case (2) with means $\theta_1 = -6$ and $\theta_2 = -0.6$. The second sub-sample is associated with the smaller sub-sample size since

the absolute value of $\theta_2 = 0.6$ is smaller. Table 11 presents the sensitivity level estimator \hat{W} , the ordinary RRT estimator μ_y , the additive ratio estimator μ_{AR} , the multiplicative ratio estimator μ_{MR} and the geometric mean ratio estimator μ_{GMR} . All the point estimates are fairly accurate under different correlation coefficients ρ_{xy} and different sensitivity levels W . Tables 12 shows the theoretical and empirical MSE comparisons for the four estimators when $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$.

Two scrambling variables are used in Case (3) with means $\theta_1 = -6$ and $\theta_2 = 0.6$. The second sub-sample is associated with a smaller sample size since the absolute value of $\theta_2 = 0.6$ is smaller. Table 13 presents the sensitivity level estimator \hat{W} , the ordinary RRT estimator μ_y , the additive ratio estimator μ_{AR} , the multiplicative ratio estimator μ_{MR} and the geometric mean ratio estimator μ_{GMR} . All the point estimates are fairly accurate under different correlation coefficients ρ_{xy} and different sensitivity levels W . Tables 14 shows the theoretical and empirical MSE comparisons for the four estimators when $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$.

Two scrambling variables are used in Case (4) with means $\theta_1 = 6$ and $\theta_2 = 0$. The second sub-sample is associated with a smaller sample size since $\theta_2 = 0$. Table 15 presents the sensitivity level estimator \hat{W} , the ordinary RRT estimator μ_y , the additive ratio estimator μ_{AR} , the multiplicative ratio estimator μ_{MR} and the geometric mean ratio estimator μ_{GMR} . All the point estimates are fairly accurate under different correlation coefficients ρ_{xy} and different sensitivity levels W . Tables 16 shows the theoretical and empirical MSE comparisons for four estimators when $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$.

For all four cases, the geometric mean ratio estimator is more efficient than the ordinary RRT estimator when $\rho_{xy} = 0.7 > 0.5$, but less efficient when $\rho_{xy} = 0.2 < 0.5$.

In addition, compared with the additive ratio estimator and the multiplicative ratio estimator, the geometric mean ratio estimator has similar MSE compared to the additive ratio estimator but has smaller MSE than the multiplicative ratio estimator for all correlations. Thus, the simulation results are consistent with the theoretical conclusions made in Chapter 3. Again, the MSE of each estimator increases as the sensitivity level W increases, which shows that an optional RRT model is more efficient than a non-optional RRT model.

In addition, for both equal and unequal sample split, most of the empirical MSEs are slightly higher than the theoretical MSEs since the expressions for the theoretical MSEs are approximate, only up to the first order.

Table 9. Point Estimates of W , μ_y , μ_{AR} , μ_{MR} and μ_{GMR} , for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using **Unequal Sample Split** ($\theta_1 = 0.6$, $\theta_2 = 6$, $n_1 < n_2$).

n	n_1	n_2		$W = 0.2$		$W = 0.5$		$W = 0.8$	
				\hat{W}	Estimate	\hat{W}	Estimate	\hat{W}	Estimate
$\rho_{xy} = 0.7$									
500	200	300	μ_y	0.1971	9.0043	0.4985	9.0035	0.7989	9.0025
			μ_{AR}	0.2011	9.0051	0.4969	9.0063	0.7987	9.0006
			μ_{MR}	0.2016	8.9960	0.5007	9.0050	0.8001	9.0001
			μ_{GMR}	0.2017	9.0006	0.5016	8.9956	0.8006	9.0049
1000	400	600	μ_y	0.1979	9.0200	0.4975	9.0033	0.7982	8.9973
			μ_{AR}	0.2013	9.0043	0.5008	8.9994	0.8004	9.0012
			μ_{MR}	0.2012	9.0023	0.5013	9.0050	0.7993	8.9975
			μ_{GMR}	0.2007	9.0032	0.5007	9.0038	0.8005	8.9970
1500	600	900	μ_y	0.2012	9.1008	0.4696	9.0033	0.7997	8.9898
			μ_{AR}	0.2014	9.0061	0.4989	9.0129	0.8015	9.0127
			μ_{MR}	0.1977	9.0126	0.4997	9.0021	0.8019	9.0050
			μ_{GMR}	0.1989	9.0056	0.5006	9.0016	0.7982	9.0049
2000	800	1200	μ_y	0.2012	9.0122	0.5011	9.0165	0.8020	9.0003
			μ_{AR}	0.1987	8.9955	0.4987	9.0129	0.8016	9.0056
			μ_{MR}	0.2003	9.0062	0.5003	8.9898	0.7984	9.0042
			μ_{GMR}	0.1998	9.0035	0.5006	9.0126	0.7998	9.0006
$\rho_{xy} = 0.2$									
500	200	300	μ_y	0.1971	9.0043	0.4985	9.0035	0.7989	9.0025
			μ_{AR}	0.2003	9.0045	0.5007	9.0122	0.8004	9.0023
			μ_{MR}	0.2009	9.0134	0.5013	8.9955	0.8004	9.0032
			μ_{GMR}	0.1987	9.0122	0.5007	8.9994	0.7993	9.1008
1000	400	600	μ_y	0.1979	9.0200	0.4975	9.0033	0.7982	8.9973
			μ_{AR}	0.1996	9.0038	0.5007	9.0038	0.7993	9.0050
			μ_{MR}	0.1976	8.9994	0.5012	9.0023	0.7982	9.0061
			μ_{GMR}	0.2012	9.0050	0.5008	8.9955	0.7997	9.0126
1500	600	900	μ_y	0.2012	9.1008	0.4696	9.0033	0.7997	8.9898
			μ_{AR}	0.2011	9.0011	0.4970	9.0050	0.8015	9.0049
			μ_{MR}	0.2004	8.9869	0.4983	8.9970	0.7984	9.0003
			μ_{GMR}	0.2013	8.9999	0.4993	8.9898	0.7998	9.0050
2000	800	1200	μ_y	0.2012	9.0122	0.5011	9.0165	0.8020	9.0003z
			μ_{AR}	0.1986	9.0001	0.5003	8.9994	0.8019	8.9956
			μ_{MR}	0.2006	9.0007	0.5014	9.0050	0.8021	8.9948
			μ_{GMR}	0.2001	8.9867	0.5017	8.9968	0.7979	9.0037

Table 10. Theoretical (**bold**) and Empirical MSE Comparisons of μ_y , μ_{AR} , μ_{MR} and μ_{GMR} for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using **Unequal Sample Split** ($\theta_1 = 0.6$, $\theta_2 = 6$, $n_1 < n_2$).

n	n_1	n_2		MSE					
				$W = 0.2$	$W = 0.5$	$W = 0.8$			
				$\rho_{xy} = 0.7$					
500	200	300	$\hat{\mu}_y$	0.0979	0.0968	0.0982	0.1004	0.1015	0.1035
			$\hat{\mu}_{AR}$	0.0736	0.0726	0.0768	0.0762	0.0803	0.0793
			$\hat{\mu}_{MR}$	0.1138	0.1120	0.1168	0.1156	0.1189	0.1187
			$\hat{\mu}_{GMR}$	0.0730	0.0726	0.0757	0.0762	0.0772	0.0793
1000	400	600	$\hat{\mu}_y$	0.0469	0.0464	0.0480	0.0481	0.0500	0.0496
			$\hat{\mu}_{AR}$	0.0352	0.0346	0.0367	0.0363	0.0380	0.0378
			$\hat{\mu}_{MR}$	0.0544	0.0530	0.0560	0.0548	0.0571	0.0562
			$\hat{\mu}_{GMR}$	0.0351	0.0346	0.0363	0.0363	0.0368	0.0378
1500	600	900	$\hat{\mu}_y$	0.0302	0.0296	0.0297	0.0307	0.0320	0.0316
			$\hat{\mu}_{AR}$	0.0226	0.0219	0.0239	0.0230	0.0241	0.0240
			$\hat{\mu}_{MR}$	0.0335	0.0334	0.0356	0.0345	0.0361	0.0354
			$\hat{\mu}_{GMR}$	0.0233	0.0219	0.0232	0.0230	0.0234	0.0240
2000	800	1200	$\hat{\mu}_y$	0.0214	0.0212	0.0228	0.0219	0.0231	0.0226
			$\hat{\mu}_{AR}$	0.0159	0.0156	0.0178	0.0164	0.0171	0.0171
			$\hat{\mu}_{MR}$	0.0241	0.0235	0.0268	0.0243	0.0267	0.0250
			$\hat{\mu}_{GMR}$	0.0159	0.0156	0.0165	0.0164	0.0167	0.0171
				$\rho_{xy} = 0.2$					
500	200	300	$\hat{\mu}_y$	0.0979	0.0968	0.0982	0.1004	0.1015	0.1035
			$\hat{\mu}_{AR}$	0.1139	0.1126	0.1164	0.1163	0.1175	0.1194
			$\hat{\mu}_{MR}$	0.1939	0.1918	0.2015	0.1954	0.2019	0.1985
			$\hat{\mu}_{GMR}$	0.1138	0.1126	0.1164	0.1163	0.1174	0.1194
1000	400	600	$\hat{\mu}_y$	0.0469	0.0463	0.0480	0.0481	0.0500	0.0496
			$\hat{\mu}_{AR}$	0.0548	0.0538	0.0548	0.0556	0.0571	0.0570
			$\hat{\mu}_{MR}$	0.0926	0.0913	0.0948	0.0931	0.0951	0.0945
			$\hat{\mu}_{GMR}$	0.0546	0.0538	0.0557	0.0556	0.0563	0.0570
1500	600	900	$\hat{\mu}_y$	0.0302	0.0295	0.0297	0.0306	0.0320	0.0316
			$\hat{\mu}_{AR}$	0.0346	0.0342	0.0368	0.0353	0.0377	0.0363
			$\hat{\mu}_{MR}$	0.0587	0.0578	0.0602	0.0589	0.0615	0.0599
			$\hat{\mu}_{GMR}$	0.0347	0.0342	0.0353	0.0353	0.0357	0.0363
2000	800	1200	$\hat{\mu}_y$	0.0214	0.0211	0.0228	0.0219	0.0231	0.0226
			$\hat{\mu}_{AR}$	0.0248	0.0244	0.0253	0.0252	0.0283	0.0259
			$\hat{\mu}_{MR}$	0.0416	0.0411	0.0434	0.0419	0.0433	0.0425
			$\hat{\mu}_{GMR}$	0.0247	0.0244	0.0252	0.0252	0.0254	0.0259

Table 11. Point Estimates of W , μ_y , μ_{AR} , μ_{MR} and μ_{GMR} , for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using **Unequal Sample Split** ($\theta_1 = -6$, $\theta_2 = -0.6$, $n_1 > n_2$).

n	n_1	n_2		$W = 0.2$		$W = 0.5$		$W = 0.8$	
				\hat{W}	Estimate	\hat{W}	Estimate	\hat{W}	Estimate
$\rho_{xy} = 0.7$									
500	300	200	μ_y	0.2012	8.9975	0.4974	9.0061	0.8020	9.0033
			μ_{AR}	0.1987	8.9970	0.5007	9.0126	0.8014	8.9994
			μ_{MR}	0.2003	8.9898	0.5013	9.0043	0.7982	9.0050
			μ_{GMR}	0.1998	9.0127	0.5009	8.9994	0.7995	9.0038
1000	600	400	μ_y	0.1996	9.0056	0.4696	9.0050	0.7979	9.0023
			μ_{AR}	0.2013	9.0122	0.4989	9.0038	0.7998	9.0032
			μ_{MR}	0.2011	8.9955	0.4997	9.0056	0.8009	9.1008
			μ_{GMR}	0.2008	9.0033	0.5006	9.0122	0.7997	9.0061
1500	900	600	μ_y	0.1979	8.9994	0.5013	8.9955	0.7996	8.9975
			μ_{AR}	0.2013	9.0050	0.5007	9.0050	0.7989	8.9970
			μ_{MR}	0.2012	9.0061	0.4696	9.0049	0.8019	8.9898
			μ_{GMR}	0.2007	9.0126	0.4989	9.0021	0.7982	9.0127
2000	1200	800	μ_y	0.2015	9.0050	0.4972	9.0016	0.8020	9.0033
			μ_{AR}	0.2006	9.0049	0.5005	8.9994	0.8015	8.9994
			μ_{MR}	0.1986	9.0003	0.5013	9.0050	0.7978	9.0050
			μ_{GMR}	0.1973	9.0021	0.5006	9.0010	0.7993	9.0023
$\rho_{xy} = 0.2$									
$\rho_{xy} = 0.2$									
500	300	200	μ_y	0.2012	8.9975	0.4974	9.0061	0.8020	9.0033
			μ_{AR}	0.2034	9.0126	0.4970	9.0032	0.8004	9.0049
			μ_{MR}	0.2021	9.0021	0.4990	9.1008	0.7993	9.0021
			μ_{GMR}	0.2015	9.0016	0.5012	9.0061	0.8005	9.0016
1000	600	400	μ_y	0.1996	9.00456	0.4696	9.0050	0.7979	9.0023
			μ_{AR}	0.2012	9.0021	0.4997	9.1008	0.7984	9.0021
			μ_{MR}	0.2007	9.0125	0.5006	9.0061	0.7998	8.9788
			μ_{GMR}	0.2012	9.0050	0.5011	9.0056	0.8006	9.0011
1500	900	600	μ_y	0.1979	8.9994	0.5013	8.9955	0.7996	8.9975
			μ_{AR}	0.2009	9.0108	0.5003	8.9955	0.7999	8.9749
			μ_{MR}	0.1987	9.0028	0.5002	8.9994	0.8014	9.0020
			μ_{GMR}	0.2011	8.9988	0.5016	9.0050	0.8019	8.9956
2000	1200	800	μ_y	0.2015	9.0050	0.4972	9.0016	0.8020	9.0033
			μ_{AR}	0.1976	9.0060	0.5037	9.0050	0.8029	8.9825
			μ_{MR}	0.1983	9.0061	0.5014	9.0049	0.7998	8.9980
			μ_{GMR}	0.1976	9.0126	0.4983	9.0003	0.8001	8.9996

Table 12. Theoretical (**bold**) and Empirical MSE Comparisons of μ_y , μ_{AR} , μ_{MR} and μ_{GMR} for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using **Unequal Sample Split** ($\theta_1 = -6$, $\theta_2 = -0.6$, $n_1 > n_2$).

n	n_1	n_2		MSE					
				$W = 0.2$	$W = 0.5$	$W = 0.8$			
				$\rho_{xy} = 0.7$					
500	300	200	$\hat{\mu}_y$	0.0980	0.1000	0.0982	0.1024	0.1007	0.1040
			$\hat{\mu}_{AR}$	0.0734	0.0754	0.0769	0.0778	0.0790	0.0794
			$\hat{\mu}_{MR}$	0.1136	0.1144	0.1159	0.1168	0.1181	0.1184
			$\hat{\mu}_{GMR}$	0.0736	0.0754	0.0759	0.0773	0.0770	0.0794
1000	900	600	$\hat{\mu}_y$	0.0469	0.0479	0.0492	0.0491	0.0490	0.0498
			$\hat{\mu}_{AR}$	0.0352	0.0359	0.0368	0.0370	0.0382	0.0379
			$\hat{\mu}_{MR}$	0.0542	0.0542	0.0570	0.0553	0.0575	0.0561
			$\hat{\mu}_{GMR}$	0.0353	0.0359	0.0362	0.0371	0.0369	0.0379
1500	900	600	$\hat{\mu}_y$	0.0303	0.0306	0.0283	0.0313	0.0325	0.0318
			$\hat{\mu}_{AR}$	0.0226	0.0228	0.0249	0.0235	0.0251	0.0240
			$\hat{\mu}_{MR}$	0.0339	0.0341	0.0355	0.0348	0.0371	0.0353
			$\hat{\mu}_{GMR}$	0.0226	0.0228	0.0232	0.0235	0.0237	0.0240
2000	1200	800	$\hat{\mu}_y$	0.0214	0.0219	0.0199	0.0224	0.0209	0.0227
			$\hat{\mu}_{AR}$	0.0158	0.0162	0.0173	0.0167	0.0176	0.0171
			$\hat{\mu}_{MR}$	0.0239	0.0240	0.0267	0.0246	0.0283	0.0249
			$\hat{\mu}_{GMR}$	0.0159	0.0162	0.0165	0.0167	0.0166	0.0171
				$\rho_{xy} = 0.2$					
500	300	200	$\hat{\mu}_y$	0.0980	0.1000	0.0982	0.1024	0.1007	0.1040
			$\hat{\mu}_{AR}$	0.1136	0.1158	0.1162	0.1182	0.1182	0.1198
			$\hat{\mu}_{MR}$	0.1941	0.1948	0.1982	0.1972	0.2003	0.1988
			$\hat{\mu}_{GMR}$	0.1135	0.1158	0.1158	0.1182	0.1179	0.1198
1000	600	400	$\hat{\mu}_y$	0.0469	0.0479	0.0492	0.0491	0.0490	0.0498
			$\hat{\mu}_{AR}$	0.0547	0.0553	0.0565	0.0565	0.0569	0.0572
			$\hat{\mu}_{MR}$	0.0928	0.0928	0.0971	0.0939	0.0900	0.0947
			$\hat{\mu}_{GMR}$	0.0548	0.0553	0.0558	0.0565	0.0563	0.0572
1500	900	600	$\hat{\mu}_y$	0.0303	0.0306	0.0283	0.0313	0.0325	0.0318
			$\hat{\mu}_{AR}$	0.0347	0.0352	0.0369	0.0359	0.0381	0.0364
			$\hat{\mu}_{MR}$	0.0591	0.0587	0.0621	0.0595	0.0626	0.0600
			$\hat{\mu}_{GMR}$	0.0346	0.0352	0.0353	0.0359	0.0357	0.0364
2000	1200	800	$\hat{\mu}_y$	0.0214	0.0219	0.0199	0.0224	0.0209	0.0227
			$\hat{\mu}_{AR}$	0.0247	0.0251	0.0292	0.0256	0.0296	0.0260
			$\hat{\mu}_{MR}$	0.0416	0.0417	0.0445	0.0423	0.0463	0.0426
			$\hat{\mu}_{GMR}$	0.0248	0.0251	0.0251	0.0256	0.0272	0.0260

Table 13. Point Estimates of W , μ_y , μ_{AR} , μ_{MR} and μ_{GMR} , for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using **Unequal Sample Split** ($\theta_1 = -6$, $\theta_2 = 0.6$, $n_1 > n_2$).

n	n_1	n_2		$W = 0.2$		$W = 0.5$		$W = 0.8$	
				\hat{W}	Estimate	\hat{W}	Estimate	\hat{W}	Estimate
$\rho_{xy} = 0.7$									
500	300	200	μ_y	0.2005	9.0043	0.5007	9.0010	0.8020	9.0061
			μ_{AR}	0.2032	9.0021	0.4896	9.0038	0.8016	9.0126
			μ_{MR}	0.2018	9.0016	0.4991	9.0033	0.7984	8.9994
			μ_{GMR}	0.2009	9.0043	0.5003	8.9898	0.7998	9.0050
1000	600	400	μ_y	0.1987	9.0033	0.5013	9.0127	0.7982	9.0033
			μ_{AR}	0.2011	8.9994	0.5007	9.0003	0.8004	8.9994
			μ_{MR}	0.1996	9.0050	0.4696	9.0052	0.7993	9.0050
			μ_{GMR}	0.1976	9.0038	0.4989	9.0010	0.8005	9.0050
1500	900	600	μ_y	0.2011	9.0061	0.5013	9.0033	0.7998	9.0049
			μ_{AR}	0.2004	9.0126	0.5007	8.9994	0.7982	9.0021
			μ_{MR}	0.2013	9.0043	0.4696	9.0050	0.8016	9.0016
			μ_{GMR}	0.1987	9.0021	0.4986	9.0038	0.8009	9.0038
2000	1200	800	μ_y	0.1986	9.0016	0.4999	9.0043	0.8001	9.0033
			μ_{AR}	0.2013	8.9875	0.4973	9.0131	0.7982	9.0020
			μ_{MR}	0.2009	9.0038	0.5002	9.0074	0.8019	8.9956
			μ_{GMR}	0.1987	9.0023	0.5014	9.0056	0.7982	8.9948
$\rho_{xy} = 0.2$									
$\rho_{xy} = 0.2$									
500	300	200	μ_y	0.2005	9.0043	0.5007	9.0010	0.8020	9.0061
			μ_{AR}	0.1981	9.0033	0.5006	9.0057	0.8005	8.9955
			μ_{MR}	0.2012	8.9898	0.5001	9.0059	0.7983	9.0043
			μ_{GMR}	0.2013	9.0127	0.4982	8.9943	0.7998	9.0016
1000	600	400	μ_y	0.1987	9.0033	0.5013	9.0127	0.7982	9.0033
			μ_{AR}	0.2012	9.0032	0.4997	9.0127	0.7999	9.0050
			μ_{MR}	0.2014	9.1008	0.5006	8.9716	0.7993	9.0038
			μ_{GMR}	0.2011	9.0061	0.5011	8.9795	0.7995	8.9994
1500	900	600	μ_y	0.2011	9.0061	0.5013	9.0033	0.7998	9.0049
			μ_{AR}	0.1996	9.0126	0.5003	9.0069	0.8019	9.0050
			μ_{MR}	0.1976	9.0043	0.4986	9.0072	0.7982	9.0033
			μ_{GMR}	0.2001	9.0050	0.4999	9.0047	0.8004	8.9994
2000	1200	800	μ_y	0.1986	9.0016	0.4999	9.0043	0.8001	9.0033
			μ_{AR}	0.2005	8.9757	0.5007	9.0045	0.7999	9.0049
			μ_{MR}	0.1983	9.0018	0.5007	8.9855	0.7998	9.0021
			μ_{GMR}	0.1981	9.0057	0.5016	8.9895	0.7982	9.0012

Table 14. Theoretical (**bold**) and Empirical MSE Comparisons of μ_y , μ_{AR} , μ_{MR} and μ_{GMR} for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using **Unequal Sample Split** ($\theta_1 = -6$, $\theta_2 = 0.6$, $n_1 > n_2$).

n	n_1	n_2		MSE					
				$W = 0.2$	$W = 0.5$	$W = 0.8$			
				$\rho_{xy} = 0.7$					
500	300	200	$\hat{\mu}_y$	0.0682	0.0665	0.0642	0.0680	0.0701	0.0691
			$\hat{\mu}_{AR}$	0.0482	0.0457	0.0487	0.0472	0.0499	0.0484
			$\hat{\mu}_{MR}$	0.0912	0.0886	0.0939	0.0901	0.0953	0.0913
			$\hat{\mu}_{GMR}$	0.0469	0.0457	0.0498	0.0472	0.0503	0.0484
1000	600	400	$\hat{\mu}_y$	0.0301	0.0319	0.0339	0.0326	0.0352	0.0331
			$\hat{\mu}_{AR}$	0.0239	0.0218	0.0251	0.0225	0.0263	0.0231
			$\hat{\mu}_{MR}$	0.0426	0.0420	0.0461	0.0427	0.0463	0.0432
			$\hat{\mu}_{GMR}$	0.0246	0.0218	0.0245	0.0225	0.0241	0.0231
1500	900	600	$\hat{\mu}_y$	0.0172	0.0203	0.0181	0.0208	0.0231	0.0211
			$\hat{\mu}_{AR}$	0.0152	0.0138	0.0168	0.0143	0.0168	0.0146
			$\hat{\mu}_{MR}$	0.0289	0.0264	0.0306	0.0269	0.0304	0.0272
			$\hat{\mu}_{GMR}$	0.0153	0.0138	0.0162	0.0143	0.0159	0.0146
2000	1200	800	$\hat{\mu}_y$	0.0155	0.0145	0.0153	0.0149	0.0181	0.0151
			$\hat{\mu}_{AR}$	0.0109	0.0098	0.0125	0.0102	0.0121	0.0104
			$\hat{\mu}_{MR}$	0.0199	0.0186	0.0209	0.0190	0.0221	0.0192
			$\hat{\mu}_{GMR}$	0.0116	0.0098	0.0130	0.0102	0.0131	0.0104
				$\rho_{xy} = 0.2$					
500	300	200	$\hat{\mu}_y$	0.0682	0.0665	0.0642	0.0680	0.0701	0.0691
			$\hat{\mu}_{AR}$	0.0839	0.0834	0.0862	0.0848	0.0859	0.0860
			$\hat{\mu}_{MR}$	0.1662	0.1635	0.1681	0.1650	0.1693	0.1661
			$\hat{\mu}_{GMR}$	0.0862	0.0834	0.0856	0.0848	0.0869	0.0860
1000	600	400	$\hat{\mu}_y$	0.0301	0.0319	0.0339	0.0326	0.0352	0.0331
			$\hat{\mu}_{AR}$	0.0415	0.0398	0.0423	0.0405	0.0452	0.0411
			$\hat{\mu}_{MR}$	0.0789	0.0778	0.0799	0.0785	0.0796	0.0791
			$\hat{\mu}_{GMR}$	0.0419	0.0398	0.0431	0.0405	0.0462	0.0411
1500	900	600	$\hat{\mu}_y$	0.0172	0.0203	0.0181	0.0208	0.0231	0.0211
			$\hat{\mu}_{AR}$	0.0260	0.0253	0.0290	0.0257	0.0289	0.0261
			$\hat{\mu}_{MR}$	0.0519	0.0492	0.0532	0.0497	0.0542	0.0500
			$\hat{\mu}_{GMR}$	0.0268	0.0253	0.0287	0.0257	0.0278	0.0261
2000	1200	800	$\hat{\mu}_y$	0.0155	0.0145	0.0153	0.0149	0.0181	0.0151
			$\hat{\mu}_{AR}$	0.0201	0.0180	0.0203	0.0184	0.0199	0.0186
			$\hat{\mu}_{MR}$	0.0379	0.0349	0.0392	0.0353	0.0385	0.0355
			$\hat{\mu}_{GMR}$	0.0196	0.0180	0.0198	0.0184	0.0190	0.0186

Table 15. Point Estimates of W , μ_y , μ_{AR} , μ_{MR} , and μ_{GMR} for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using **Unequal Sample Split** ($\theta_1 = 6, \theta_2 = 0, n_1 > n_2$).

n	n_1	n_2		$W = 0.2$		$W = 0.5$		$W = 0.8$	
				\hat{W}	Estimate	\hat{W}	Estimate	\hat{W}	Estimate
$\rho_{xy} = 0.7$									
500	300	200	μ_y	0.2034	9.0033	0.5019	9.0043	0.7983	8.9994
			μ_{AR}	0.1997	8.9994	0.4971	9.0050	0.8004	9.0050
			μ_{MR}	0.1983	9.0050	0.4991	9.0049	0.7993	9.0038
			μ_{GMR}	0.1981	9.0038	0.5003	9.0003	0.8005	8.9994
1000	600	400	μ_y	0.2012	9.0033	0.5006	9.0021	0.7998	9.0050
			μ_{AR}	0.2013	8.9975	0.4985	9.0016	0.7982	9.0033
			μ_{MR}	0.2007	8.9970	0.4986	9.0049	0.8004	8.9994
			μ_{GMR}	0.2012	8.9898	0.4999	9.0021	0.8016	9.0050
1500	900	600	μ_y	0.2012	9.0127	0.5013	8.9898	0.8019	9.0033
			μ_{AR}	0.1987	9.0049	0.5007	9.0127	0.7983	8.9994
			μ_{MR}	0.2003	9.0021	0.4996	9.0049	0.7995	9.0050
			μ_{GMR}	0.1998	9.0043	0.4983	9.0021	0.7983	9.0038
2000	1200	800	μ_y	0.1996	8.9994	0.4696	9.0016	0.8019	9.0033
			μ_{AR}	0.1976	9.0050	0.4989	8.9875	0.7982	8.9994
			μ_{MR}	0.1994	9.0038	0.4997	9.0012	0.7983	9.0050
			μ_{GMR}	0.1987	9.0012	0.5006	9.0030	0.8003	9.0038
$\rho_{xy} = 0.2$									
500	300	200	μ_y	0.2034	9.0033	0.5019	9.0043	0.7983	8.9994
			μ_{AR}	0.2003	9.0038	0.5006	9.0032	0.8014	9.0050
			μ_{MR}	0.2013	9.0021	0.5003	9.1008	0.7982	9.0038
			μ_{GMR}	0.2009	9.0016	0.5013	9.0061	0.7993	8.9975
1000	600	400	μ_y	0.2012	9.0033	0.5006	9.0021	0.7998	9.0050
			μ_{AR}	0.2011	9.0033	0.4997	8.9898	0.7998	8.9898
			μ_{MR}	0.1996	9.0020	0.5006	9.0127	0.8976	9.0127
			μ_{GMR}	0.1976	8.9956	0.5011	9.0022	0.7999	8.9994
1500	900	600	μ_y	0.2012	9.0127	0.5013	8.9898	0.8019	9.0033
			μ_{AR}	0.2004	9.0049	0.5007	9.0050	0.8005	9.0061
			μ_{MR}	0.2013	9.0021	0.4696	9.0038	0.8016	9.0126
			μ_{GMR}	0.1987	9.0016	0.4987	9.0003	0.7984	9.0050
2000	1200	800	μ_y	0.1996	8.9994	0.4696	9.0016	0.8019	9.0033
			μ_{AR}	0.2012	8.9898	0.4696	9.0021	0.8013	9.0003
			μ_{MR}	0.2013	9.0127	0.4989	9.0138	0.8004	9.0013
			μ_{GMR}	0.2011	9.0011	0.5013	8.9990	0.7993	9.0001

Table 16. Theoretical (**bold**) and Empirical MSE Comparisons of μ_y , μ_{AR} , μ_{MR} and μ_{GMR} for $\rho_{xy} = 0.7$ and $\rho_{xy} = 0.2$ using **Unequal Sample Split** ($\theta_1 = 6$, $\theta_2 = 0$, $n_1 > n_2$).

n	n_1	n_2		MSE					
				$W = 0.2$	$W = 0.5$	$W = 0.8$			
				$\rho_{xy} = 0.7$					
500	300	200	$\hat{\mu}_y$	0.0761	0.0762	0.0772	0.0777	0.0853	0.0792
			$\hat{\mu}_{AR}$	0.0581	0.0541	0.0598	0.0556	0.0613	0.0572
			$\hat{\mu}_{MR}$	0.1006	0.0957	0.1021	0.0972	0.1041	0.0987
			$\hat{\mu}_{GMR}$	0.0579	0.0541	0.0577	0.0556	0.0601	0.0572
1000	600	400	$\hat{\mu}_y$	0.0398	0.0365	0.0362	0.0372	0.0409	0.0380
			$\hat{\mu}_{AR}$	0.0298	0.0258	0.0306	0.0265	0.0331	0.0273
			$\hat{\mu}_{MR}$	0.0497	0.0453	0.0521	0.0460	0.0498	0.0468
			$\hat{\mu}_{GMR}$	0.0303	0.0258	0.0312	0.0265	0.0331	0.0273
1500	900	600	$\hat{\mu}_y$	0.0267	0.0233	0.0272	0.0237	0.0261	0.0242
			$\hat{\mu}_{AR}$	0.0198	0.0164	0.0206	0.0168	0.0214	0.0173
			$\hat{\mu}_{MR}$	0.0316	0.0285	0.0332	0.0290	0.0341	0.0294
			$\hat{\mu}_{GMR}$	0.0179	0.0164	0.0193	0.0168	0.0201	0.0173
2000	1200	800	$\hat{\mu}_y$	0.0197	0.0167	0.0183	0.0170	0.2000	0.0173
			$\hat{\mu}_{AR}$	0.0143	0.0116	0.0153	0.0120	0.0141	0.0123
			$\hat{\mu}_{MR}$	0.0226	0.0201	0.0231	0.0204	0.0240	0.0208
			$\hat{\mu}_{GMR}$	0.0132	0.0116	0.0159	0.0120	0.0143	0.0123
				$\rho_{xy} = 0.2$					
500	300	200	$\hat{\mu}_y$	0.0761	0.0762	0.0772	0.0777	0.0853	0.0792
			$\hat{\mu}_{AR}$	0.0969	0.0927	0.0971	0.0942	0.0992	0.0957
			$\hat{\mu}_{MR}$	0.1770	0.1724	0.1799	0.1739	0.1816	0.1754
			$\hat{\mu}_{GMR}$	0.0959	0.0927	0.0978	0.0942	0.0996	0.0957
1000	600	400	$\hat{\mu}_y$	0.0398	0.0365	0.0362	0.0372	0.0409	0.0379
			$\hat{\mu}_{AR}$	0.0469	0.0443	0.0476	0.0450	0.0479	0.0457
			$\hat{\mu}_{MR}$	0.0869	0.0821	0.0865	0.0828	0.0880	0.0835
			$\hat{\mu}_{GMR}$	0.0482	0.0443	0.0474	0.0450	0.0497	0.0457
1500	900	600	$\hat{\mu}_y$	0.0267	0.0233	0.0272	0.0237	0.0261	0.0242
			$\hat{\mu}_{AR}$	0.0299	0.0281	0.0309	0.0286	0.0331	0.0291
			$\hat{\mu}_{MR}$	0.0562	0.0519	0.0571	0.0524	0.0568	0.0529
			$\hat{\mu}_{GMR}$	0.0300	0.0281	0.0293	0.0286	0.0325	0.0291
2000	1200	800	$\hat{\mu}_y$	0.0197	0.0167	0.0183	0.0170	0.0200	0.0173
			$\hat{\mu}_{AR}$	0.0216	0.0201	0.0219	0.0204	0.0215	0.0207
			$\hat{\mu}_{MR}$	0.0398	0.0369	0.0388	0.0372	0.0400	0.0376
			$\hat{\mu}_{GMR}$	0.0223	0.0201	0.0221	0.0204	0.0216	0.0207

CHAPTER V

DISCUSSION

5.1 General Discussion

In this thesis, we first introduced several RRT models, and the basic ratio and product estimators. Ratio estimators improve efficiency by utilizing an auxiliary variable which is positively correlated with the study variable. Product estimators are used when the correlation is negative. In the situations where the study variable is sensitive and we cannot obtain it directly, we use RRT models to reduce Social Desirability Bias and also to increase response rate.

Sousa et al. (2010) first introduced the ratio estimator under the non-optional RRT models. Instead of using the sensitive variable alone, they also use a nonsensitive auxiliary variable which is positively correlated with the variable of interest. Their theoretical and simulation results showed that their ratio estimator is more efficient than the estimator without using an auxiliary variable under the non-optional RRT model.

Optional RRT models provide greater efficiency than non-optional RRT models since some respondents do not consider survey questions sensitive and are willing to provide a truthful response. But this adds another complexity to the model by introducing an additional parameter (W) to the model. Instead of the non-optional RRT model, Kalucha et al. (2015) proposed two ratio estimators, the additive ratio estimator and the multiplicative ratio estimator, under the optional RRT model of Gupta et al. (2010) with $T=0$ using a simple random sampling without replacement

from a finite population. Their results showed that the additive ratio estimator is more efficient than the ordinary RRT estimator when the study variable and the auxiliary variable has a correlation coefficient greater than $\frac{1}{2}$. However, the multiplicative ratio estimator was not as efficient as the other two. The reason for failure of the multiplicative ratio estimator was that they did not use the geometric mean of the two sub-sample ratios.

We improved the Kalucha et al. (2015) multiplicative ratio estimator by taking geometric mean of two sub-sample ratios. In Chapter 3, we derived the expressions for the bias and the MSE of the proposed geometric mean ratio estimator, correct up to first order approximation, and also compared it with the ordinary RRT estimator, as well as two previous ratio estimators proposed by Kalucha et al. (2015). The results showed that the geometric mean ratio estimator is more efficient than the ordinary RRT estimator when the correlation coefficient between the variable of interest and the auxiliary variable is greater than $\frac{1}{2}$. It also has efficiency similar to the additive ratio estimator but is less biased. For any correlations, the geometric mean ratio estimator is more efficient than the ordinary multiplicative ratio estimator of Kalucha et al. (2015).

In Chapter 4, we carried out an extensive empirical study to verify our theoretical results. The results are in agreement with the corresponding theoretical conclusions. The empirical MSEs are also compared with the theoretical MSEs. Since the expressions for the theoretical MSE are correct up to first order approximation, the empirical results showed that they are slightly greater than the theoretical MSEs.

5.2 Future Directions

For future study, we can derive the bias and MSE expressions up to the second order of approximation although the higher order properties will be more complicated. Also, we can study product estimators under RRT models if the auxiliary variable is negatively correlated with the study variable. Even though this thesis successfully demonstrated that the geometric mean ratio estimator is more efficient than the ordinary RRT estimator when $\rho_{xy} > \frac{1}{2}$, a real application is still needed to evaluate the performance of this estimator. But that will require an extensive field work. In addition, one can study the geometric mean ratio estimator under different sampling designs, such as stratified RRT models.

REFERENCES

- [1] M. Y. Chaloupka. Application of the randomized response technique to marine park management: an assessment of permit compliance. *Environment Management*, 9(5):393–398, 1985.
- [2] A. Chhabra, B. K. Dass, and S. Gupta. Estimating prevalence of sexual abuse by an acquaintance with an optional unrelated question rrt model. *North Carolina Journal of Mathematics and Statistics.*, 2:1–9, 2016.
- [3] L. P. Chow, W. Gruhn, and W. P. Chang. Feasibility of the randomized response technique in rural ethiopia. *American Journal of Public Health*, 69:273–276, 1979.
- [4] W. G. Cochran. *Sampling Techniques*. John Wiley and Sons, New-York, 1997.
- [5] P. Cross, G. Edwards-Jones, H. Omed, and A. Williams. Use of a randomized response technique to obtain sensitive information on animal disease prevalence. *Preventive Veterinary Medicine*, 96:252–262., 2010.
- [6] B. Eichhorn and L. Hayre. Scrambled randomized response methods for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference.*, 7:307–316, 1983.
- [7] G. Geng, G. Gao, Y.-H. Ruan, M.-R. Yu, and Y.-H. Zhou. Behavioral risk profile of men who have sex with men in beijing, china: results from a cross-sectional survey with randomized response techniques. *Chinese Medical Journal*, 129(5): 523–529., 2016.
- [8] T. Gill, A. Tuck, S. Gupta, M. Crowe, and J. Figueroa. A field test of optional unrelated question randomized response models: estimates of risky sexual behaviors. *Springer Proceedings in Mathematics and Statistics Series*, 64:135–146, 2013.
- [9] B. G. Greenberg, A.-L. A. Abul-Ela, W. R. Simmons, and D. G. Horvitz. The unrelated question randomized response model: theoretical framework. *Journal of the American Statistical Association*, 64(326)(520-539), 1969.
- [10] L. K. Grover and P. Kaur. An improved estimator of the finite population mean in simple random sampling. *Model Assisted Statistics and Applications*, 6(1): 47–55, 2011.

- [11] S. Gupta, B. Gupta, and S. Singh. Estimation of sensitivity level of personal interview survey questions. *Journal of Statistical Planning and inference*, 100: 239–247, 2002.
- [12] S. Gupta and J. Shabbir. Sensitivity estimation for personal interview survey questions. *Statistica*, 64:643–653, 2004.
- [13] S. Gupta, J. Shabbir, and S. J. Sehra. Mean and sensitivity estimation in optional randomized response models. *Journal of Statistical Planning and Inference*, 140 (10):2870–2874, 2010.
- [14] C. Kadilar and H. Cingi. Improvement in estimating the population mean in simple random sampling. *Applied Mathematics Letters*, 19(1):75–79, 2006.
- [15] G. Kalucha, S. Gupta, and B. Dass. Ratio estimation of finite population mean using optional randomized response models. *Journal of Statistical Theory and Practice*, 9(3):633–645, 2015.
- [16] J. Kerkvliet. Estimating a logit model with randomized data: the case of cocaine use. *Austral. J. Statist*, 36:9–20, 1994.
- [17] S. S. K. Kwan, M. K. P. So, and K. Y. Tam. Applying the randomized response technique to elicit truthful responses to sensitive questions in is research: The case of software privacy behavior. *Information Systems Research*, 21:941–959, 2010.
- [18] S. Mehta, B. Dass, J. Shabbir, and S. Gupta. A three stage optional randomized reponse model. *Journal of Statistical Theory and Practice*, 6(3):417–427, 2012.
- [19] M. N. Murthy. Product method of estimation. *Sankhya*, 26:69–74., 1963.
- [20] A. J. Nederhof. Methods of coping with social desirability bias: a review. *European Journal of Social Psychology*, 15(3):263–280, 1985.
- [21] D. S. Robson. Applications of multivariate polykays of the theory of unbiased ratio-type estimation. *Journal of the American Statistical Association*, 52:511–522, 1957.
- [22] J. Shabbir and S. Gupta. Estimation of the finite population mean in two-phase sampling when auxiliary variables are attribute. *Hacettepe Journal of Mathematics and Statistics*, 39(1):121–129, 2010.
- [23] S. Singh, A. H. Joarder, and M. King. Regression analysis using scrambled responses. *Austral. J. Statist*, 38(2):201–211, 1996.

- [24] R. Sousa, S. Shabbir, P. Corte-Real, and S. Gupta. Ratio estimation of the mean of a sensitive variable in the presence of auxiliary information. *Journal of Statistical Theory and Practice*, 4(3):495–507, 2010.
- [25] P. Sukhatme and B. Sukhatme. *Sampling theory of survey with applications*. Iowa State University Press., 1970.
- [26] Y. Turgut and H. Cingi. New generalized estimators for the population variance using auxiliary information. *Hacettepe Journal of Mathematics and Statistics*, 37(2):177–184, 2008.
- [27] S. L. Warner. Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.
- [28] S. L. Warner. The linear randomized response model. *Journal of the American Statistical Association*, 66(336):884–888, 1971.
- [29] C. X., D. Q., and J. Z. et al. The randomized response technique application in the survey of homosexual commercial sex among men in beijing. *Iran J Public Health*, 43 (4):416–422, 2014.