This study mainly consists of three important issues we face in survey sampling: social desirability bias, measurement errors, and non-response. In this dissertation, we study the mean estimation of a sensitive variable under measurement errors and non-response. We propose a generalized mean estimator, then discuss the bias and the mean square error (MSE) of this estimator and present the comparisons with other estimators under the measurement errors and non-response using optional RRT model (ORRT). We also study the performance of the proposed estimator under the same situations using stratified random sampling. Simulation studies are also conducted to verify the theoretical results. Both the theoretical and empirical results show that the generalized mean estimator is more efficient than the ordinary RRT estimator that does not utilize the auxiliary variable, and the ratio estimator which is one of the commonly used mean estimator.
MEAN ESTIMATION OF SENSITIVE VARIABLES UNDER MEASUREMENT ERRORS AND NON-RESPONSE

by

Qi Zhang

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

Greensboro
2020

Approved by

________________________
Committee Chair
This dissertation 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
Scott Richter
Jianping Sun
Somya Mohanty
Rakhi Singh

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 in influential in guiding me, while boosting my confidence as well. It was a pleasure working with him.

Next I would like to thank Dr. Scott Richter, Dr. Jianping Sun, Dr. Somya Mohanty, and Dr. Rakhi Singh for the assistance and direction they have provided me. I would also like to express my most sincere appreciation to Sadia Khalil of Lahore College for Women University also who is my co-author on several papers and who was always available for discussions and guidance.

Also, I would 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 the financial support and an opportunity to obtain my PhD degree.

Last but not least, I would like to express 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

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>vi</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION AND BACKGROUND</td>
<td>1</td>
</tr>
<tr>
<td>I.1. Randomized Response Technique Models</td>
<td>3</td>
</tr>
<tr>
<td>I.2. Hansen and Hurwitz Two-phase Sampling Techniques</td>
<td>11</td>
</tr>
<tr>
<td>I.3. Measurement Errors</td>
<td>13</td>
</tr>
<tr>
<td>I.4. Simple Random Sampling and Stratified Random Sampling</td>
<td>14</td>
</tr>
<tr>
<td>I.5. Outline of the Dissertation</td>
<td>15</td>
</tr>
<tr>
<td>II. LITERATURE REVIEW</td>
<td>17</td>
</tr>
<tr>
<td>II.1. Mean Estimation under RRT Models</td>
<td>17</td>
</tr>
<tr>
<td>II.2. Mean Estimation under RRT Models and Measurement Errors</td>
<td>22</td>
</tr>
<tr>
<td>II.3. Mean Estimation under Non-response</td>
<td>25</td>
</tr>
<tr>
<td>II.4. Mean Estimation under Stratified Random Sampling</td>
<td>30</td>
</tr>
<tr>
<td>III. MEAN ESTIMATION UNDER ORRT MODELS IN THE PRESENCE OF MEASUREMENT ERRORS</td>
<td>36</td>
</tr>
<tr>
<td>III.1. A General Scrambling Model</td>
<td>36</td>
</tr>
<tr>
<td>III.2. ORRT Version of the General Scrambling Model</td>
<td>39</td>
</tr>
<tr>
<td>III.3. Some Existing Mean Estimators under Measurement Errors</td>
<td>41</td>
</tr>
<tr>
<td>III.5. Simulation Study</td>
<td>49</td>
</tr>
<tr>
<td>III.6. Concluding Chapter Remarks</td>
<td>60</td>
</tr>
</tbody>
</table>
Table III.1. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators when $\sigma^2_v = \sigma^2_p = 1$ and $\sigma^2_s = 0.2*\sigma^2_x$. .................... 51

Table III.2. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators when $\sigma^2_v = \sigma^2_p = 1$ and $\sigma^2_s = 0.5*\sigma^2_x$. .................... 53

Table III.3. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators when $\sigma^2_v = \sigma^2_p = 1$ and $\sigma^2_s = 1*\sigma^2_x$. .................... 55

Table III.4. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under the Conditions of $\sigma^2_v = \sigma^2_p = 1, 5, 10$ when $W = 0.8$, $\sigma^2_T = 0.5$ and $\sigma^2_s = 0.5*\sigma^2_x$. .................... 58

Table III.5. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Generalized Mean Estimator under Different $\alpha$ and $\beta$ Values when $\sigma^2_v = \sigma^2_p = 1, W = 0.8$ and $\sigma^2_s = 0.5*\sigma^2_x$. .................... 59

Table IV.1. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators when Response Rate = 40% , $\sigma^2_v = \sigma^2_u = \sigma^2_p = 1$, $f = 2$ and $\sigma^2_s = 0.2*\sigma^2_x$. .................... 76

Table IV.2. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators when Response Rate = 40% , $\sigma^2_v = \sigma^2_u = \sigma^2_p = 1$, $f = 2$ and $\sigma^2_s = 0.5*\sigma^2_x$. .................... 78

Table IV.3. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators when Response Rate = 40% , $\sigma^2_v = \sigma^2_u = \sigma^2_p = 1$, $f = 2$ and $\sigma^2_s = 1*\sigma^2_x$. .................... 80

Table IV.4. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under the Conditions of $\sigma^2_v = \sigma^2_u = \sigma^2_p = 1, 5, 10$ when Response Rate = 40% , $\sigma^2_s = 0.5*\sigma^2_x$. .................... 83

Table IV.5. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under the Conditions of Response Rate (RR) = 20%, 40%, 60% when $\sigma^2_v = \sigma^2_u = \sigma^2_p = 1$, $W = 0.8$, and $\sigma^2_T = 0.5*\sigma^2_x$. .................... 85
Table V.1. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under Stratified Random Sampling when Response Rate = 40% , $\sigma_v^2 = \sigma_u^2 = \sigma_p^2 = 1$, $f = 2$ and $\sigma_s^2 = 0.2*\sigma_x^2$. .................................................... 99

Table V.2. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under Stratified Random Sampling when Response Rate = 40% , $\sigma_v^2 = \sigma_u^2 = \sigma_p^2 = 1$, $f = 2$ and $\sigma_s^2 = 0.5*\sigma_x^2$. .................................................... 101

Table V.3. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under Stratified Random Sampling when Response Rate = 40% , $\sigma_v^2 = \sigma_u^2 = \sigma_p^2 = 1$, $f = 2$ and $\sigma_s^2 = 1*\sigma_x^2$. .................................................... 103

Table V.4. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under the Conditions of $\sigma_v^2 = \sigma_u^2 = \sigma_p^2 = 1$, 5, 10 and Stratified Random Sampling when Response Rate = 40%, $W = 0.8$, and $\sigma_T^2 = 0.5*\sigma_x^2$. .................................................... 105

Table V.5. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under the Conditions of Response Rate = 20%, 40%, 60% and Stratified Random Sampling when $\sigma_v^2 = \sigma_u^2 = \sigma_p^2 = 1$, $W = 0.8$, and $\sigma_T^2 = 0.5*\sigma_x^2$. .................................................... 106
CHAPTER I
INTRODUCTION AND BACKGROUND

In statistical studies, one of the fundamental goals is to estimate the true value of population parameters. Unfortunately, collecting data from every member of the population would be too expensive or time-consuming if the population is large. Instead of conducting a census, we can collect data from a sample and use the sample statistics to make inferences about the target population. However, sometimes a sample cannot represent the population accurately 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 respondent mistakes, measurement errors and non-response, etc. Therefore, the inferences will be reasonable only if the sample truly represents the population and the responses collected from the sample are accurate. Otherwise the sample is biased and conclusion from the study are not trustworthy.

There are many sampling methods, such as the simple random sampling (SRS), cluster sampling, and stratified random sampling, etc. In order to have a representative sample, we could use different sampling methods depending on the situation. SRS is the basic sampling technique where each member of the population has an equal chance of being selected in the sample. But sometimes sub-populations within an overall population vary significantly, and it would be better to sample each sub-population independently. This refers to stratified random sampling.
We also have different types of survey methods that are often used, such as email surveys, phone surveys and personal interview surveys. Email and phone surveys are relatively cheaper but usually have high non-response rate. The non-response issue may cause some participation bias. For example, people who feel strongly about an issue may be more likely to participate, and their opinion may not represent the opinion of the whole population. People are unlikely to reject a personal interview survey compared to the other two methods but it costs more. Additionally, if the survey question is sensitive, a personal face-to-face interview may cause social desirability response bias. For example, if a survey question asks "What is your salary?" or "Have you ever used illegal drugs?", most people would want to present themselves in a socially desirable light, therefore their responses may be biased towards what they feel is socially desirable. In addition to participation bias and social desirability response bias, there are some other non-sampling errors, such as measurement errors occurring due to definition differences or misunderstandings which will also affect parameter estimation. Therefore, dealing with these problems is very important when we estimate the population parameters of a sensitive variable.

This dissertation will consist of three important issues we face in survey sampling: social desirability bias, measurement errors, and non-response. We will study mean estimation for a sensitive variable in the presence of such issues using both SRS and stratified random sampling. In Section I.1, we will introduce some RRT models which improve efficiency of the mean estimation if a survey involves sensitive questions. In Section I.2, we will demonstrate how the Hansen and Hurwitz two-phase sampling technique works if non-response exists. In Sections I.3 and I.4, the basic ideas of measurement errors and two common sampling methods will be briefly described. An outline of the dissertation will be presented in Section I.5.
I.1. Randomized Response Technique Models

Randomized response technique (RRT) is an important method to prevent or reduce social desirability response bias and is widely used in survey interviews. The first RRT model was proposed by Warner in 1965 [84]. It was modified later by many researchers including Greenberg et al. (1969) [15], Eichhorn (1983) [10], Gupta et al. (2002) [17], Gupta et al. (2010) [19], and Sousa et al. (2010) [77] etc. It allows respondents to answer sensitive questions more comfortably and provides more accurate estimates. RRT models have been used in many field surveys, such as Kerkvliet et al. (1994) [32], Gill et al. (2013) [14], Chhabra et al. (2016) [06], Chen et al. (2014) [05], and Geng et al. (2016) [13]. Several RRT models will be described in detail in this section, but the optional RRT model (ORRT) will be the main focus in this dissertation.

I.1.1 Warner’s Binary RRT Model (1965)

In 1965, Warner [84] proposed the first binary RRT model to estimate the prevalence of a sensitive characteristic in a population. It increases response rate, and also makes the respondent feel more comfortable in answering survey questions truthfully and reduces social desirability response bias. 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 may be:

1. I have been told by a healthcare professional that I have STD.

2. I have never been told by a healthcare professional that I have 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 $n$ respondents is drawn from the population, and each subject is asked to pick a card from the deck and provide a "yes" or a "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 STD; there is another possibility that the person may have 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 provide a true response since his/her privacy is assured.

Let $\pi$ be the true probability of a subject having an STD in the population, and $p_y$ be the probability of a "yes" responses. Then,

$$p_y = p\pi + (1 - p)(1 - \pi). \quad (I.1)$$

The estimate of $\pi$ is then given by

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

The variance of this estimator under simple random sampling with replacement is given by

$$Var(\hat{\pi}) = \frac{\pi(1 - \pi)}{n} + \frac{p(1 - p)}{n(2p - 1)^2}. \quad (I.3)$$

In order to minimize the variance, a large sample size $n$ should be chosen and the proportion ($p$) of Statement 1 should be closer to 0 or 1.
I.1.2 Warner’s (1971) and Pollock & Beck’s (1976) Quantitative RRT Model

The estimator proposed in 1965 can be used to estimate binary variables, but many times the question of interest is a quantitative one. Warner [85] modified the RRT model for quantitative cases in 1971. This was further expanded by Pollock and Bek (1971) [53]. 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 from a pre-assigned distribution, preferably with mean zero. 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 $Y$ be the sensitive variable with unknown mean $\mu_y$ and unknown variance $\sigma_y^2$, and $S$ be the scrambling variable (independent of $Y$) with known mean $\mu_s$ and known variance $\sigma_s^2$. Also let $Z$ be the reported response. Then

$$Z = Y + S.$$  \hfill (I.4)

The expected response is given by

$$E(Z) = E(Y) + E(S).$$  \hfill (I.5)

This leads to an unbiased estimator of the mean of the sensitive variable $Y$ is given by

$$\hat{\mu}_y = \bar{z} - \mu_s.$$  \hfill (I.6)

or simply $\bar{z}$ if $\mu_s=0$. 

5
The variance of $\hat{\mu}_y$ is given by

\[
\text{Var}(\hat{\mu}_y) = \text{Var}(\bar{Z}) = \frac{\text{Var}(Z)}{n} = \frac{\sigma^2_y + \sigma^2_s}{n} = \frac{\sigma^2_y}{n} + \frac{\sigma^2_s}{n},
\]

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

**I.1.3 Eichhorn and Hayre’s Multiplicative RRT Model (1983)**

Eichhorn and Hayre [10] 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 divide by the mean of the scrambling variable.

The reported response is given by $Z = YS/\mu_s$, where $\mu_s = E(S)$. Usually $\mu_s$ is chosen to be 1. This leads to the unbiased estimator

\[
\hat{\mu}_y = \bar{z},
\]

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

\[
\text{Var}(\hat{\mu}_y) = \frac{1}{n}[\sigma^2_y + \frac{\sigma^2_s(\sigma^2_y + \mu^2_y)}{\mu^2_s}].
\]

**I.1.4 Gutpa et al. Optional RRT Model (2002)**

In the above models, every respondent is forced to provide a scrambled response. However, researchers have realized that a question may be sensitive for one respondent, but not sensitive for another. In order to make the survey results more accurate, Gupta et al. (2002) [17] modified the Eichhorn and Hayre (1983) [10] multiplicative scrambling RRT model, and introduced an Optional RRT (ORRT) model that allows researchers to estimate not only the mean of the variable of interest, but also the...
sensitivity level $W$ (the proportion of subjects in the population who consider the question sensitive).

In this model, if respondents feel the question is sensitive, they will provide a scrambled response $Y_S$. Otherwise, the respondents will answer the sensitive survey question directly and provide the true response $Y$. In this model, we assume that both $Y$ and $S$ are positive valued random variables and that the mean of the scrambling variable $\mu_s = 1$ and the variance $\sigma_s^2$. Under this model, the reported response $Z$ is given by

$$ Z = \begin{cases} Y & \text{with probability } 1-W \\ YS & \text{with probability } W. \end{cases} \quad (I.10) $$

The expected value of $Z$ is given by

$$ E(Z) = E(Y)(1-W) + E(YS)W = \mu_y(1-W) + \mu_y \mu_s W = \mu_y, \quad (I.11) $$

and the variance of this unbiased estimator of the population mean $\mu$ is given by

$$ Var(\hat{\mu}_y) = \frac{1}{n} [\sigma_y^2 + W \sigma_s^2 (\sigma_y^2 + \mu_y^2)]. \quad (I.12) $$

Note that $Var(\hat{\mu}_y)$ increases with $W$, and hence there is gain in efficiency compared to the non-optional model where $W=1$. Gupta et al (2002) [17] gave an estimator for the sensitivity level $W$ which is given by

$$ \hat{W} = \frac{1}{n} \sum_{i=1}^{n} \log(Z_i) - \log(\frac{1}{n} \sum_{i=1}^{n} Z_i), \quad (I.13) $$

where $\delta = E[\log(S)]$. 

7
I.1.5 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 the sensitive characteristic. Another shortcoming of the multiplicative scrambling 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) [64] showed that this case is more dangerous than not using the scrambled response. In order to deal with these problems and also to estimate the sensitively level $W$ without using any approximations, Gupta et al. (2010) [19] proposed an additive ORRT model using a split-sample approach.

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 $S_2$. Since the multiplicative scrambling method compromises respondent anonymity, the scrambling method used is additive scrambling.

Again, let $Y$ be a sensitive variable with mean $\mu_y$, $S_i$ ($i = 1, 2$) be scrambling variable (independent of $Y$) with mean $\mu_{s_i}$ ($i = 1, 2$) and variance $\sigma_{s_i}^2$ ($i = 1, 2$), and $Z_i$ ($i = 1, 2$) be the reported response in sub-sample $i$ ($i = 1, 2$). Under this model, the reported response $Z_i$ in the $i^{th}$ sub-sample is given by

$$Z_i = \begin{cases} 
Y & \text{with probability } 1-W \\
Y + S_i & \text{with probability } W
\end{cases} \quad \text{where } i = 1, 2. \quad (I.14)$$

The expected value and variance of $Z_i$ are given by

$$E(Z_i) = \mu_y + \mu_{s_i}W \quad (I.15)$$
\[ Var(Z_i) = \sigma^2_y + \sigma^2_{s_1}W + \mu^2_{s_i}W(1 - W), \text{ where } \mu_{s_i} = E(S_i) \ (i = 1, 2). \] (I.16)

The unbiased estimators \( \hat{\mu}_y \) and \( \hat{W} \) and their corresponding variances are given by

\[ \hat{\mu}_y = \frac{\mu_{s_1} \bar{z}_2 - \mu_{s_2} \bar{z}_1}{\mu_{s_1} - \mu_{s_2}}, \] (I.17)

\[ \hat{W} = \frac{\bar{z}_1 - \bar{z}_2}{\mu_{s_1} - \mu_{s_1}}, \] (I.18)

\[ Var(\hat{\mu}_y) = \frac{1}{(\mu_{s_2} - \mu_{s_1})^2} \left( \mu_{s_2}^2 \frac{\sigma^2_{z_1}}{n_1} + \mu_{s_1}^2 \frac{\sigma^2_{z_2}}{n_2} \right), \] (I.19)

and

\[ Var(\hat{W}) = \frac{1}{(\mu_{s_2} - \mu_{s_1})^2} \left( \frac{\sigma^2_{z_1}}{n_1} + \frac{\sigma^2_{z_2}}{n_2} \right), \quad \mu_{s_1} \neq \mu_{s_2}. \] (I.20)

### I.1.6 Diana and Perri’s Linear Combination Model (2011)

The goal of a RRT model is to protect respondent privacy. Diana and Perri [08] believe that a combination of additive and multiplicative approaches can bring more confidence among the respondents about their privacy protection since two scrambling variables will be introduced to the model. Let \( T \) be a scrambling variable with mean \( \mu_T \) and variance \( \sigma^2_T; \) and \( S \) be another scrambling variable, independent of \( T, \) and with mean \( \mu_s \) and variance \( \sigma^2_s. \) Both \( T \) and \( S \) are independent of the study variable \( Y. \) They introduced a more general linear combination model given by

\[ Z = TY + S. \] (I.21)
It is common to assume $\mu_T = 1$ and $\mu_s = 0$. Then the expected value and variance of $Z$ are given by

$$E(Z) = \mu_y,$$  \hspace{1cm} (I.22)

and

$$Var(Z) = \sigma_s^2(\mu_y^2 + \sigma_y^2) + \sigma_y^2 + \sigma_T^2.$$  \hspace{1cm} (I.23)

If $\mu_T = 1$ and $\mu_s = 0$, the unbiased estimator $\hat{\mu}_y$ and its variance are given by

$$\hat{\mu}_y = (\bar{z} - \mu_s)/\mu_T = \bar{z},$$  \hspace{1cm} (I.24)

and

$$Var(\hat{\mu}_y) = \frac{1}{n} \{\sigma_s^2(\mu_y^2 + \sigma_y^2) + \sigma_y^2 + \sigma_T^2\}.  \hspace{1cm} (I.25)$$

In this section, we have introduced several RRT models and presented estimators for the sensitive variable mean, as well as the estimator for the sensitivity level. In Section I.2, we will talk about another technique that we often use for high non-response rate.
I.2. Hansen and Hurwitz Two-phase Sampling Techniques

As we mentioned earlier, non-response is widespread in email or phone surveys. Non-response refers to individuals who are chosen for the sample and are unwilling or unable to participate in the survey. Some of them may feel no obligation to complete a survey, or they do not care about the survey itself and refuse to do it; others may not be available at the time of the survey; or a person may not feel comfortable to provide the true answer for the survey question. Such cases reduce the precision of population estimates.

Since email or phone surveys are easier, cheaper and more convenient, nowadays many researches use these two survey methods to obtain information. However, the high non-response rate become an important concern in the study. According to Fan & Yan (2010) [11] and Miller & Dillman (2011)[49], a response rate of 40-50 percent is considered excellent. In reality, it is much smaller than this. Among all the sampling methods, personal face-to-face interview is the one that reduces non-response rate the most, but the cost is considerably higher than other methods. One may wonder if we could combine the strengths of different survey methods. Hansen and Hurwitz (1946)[25] were the first to suggest a procedure of taking a sub-sample of non-respondents after the first mail or phone attempt and then obtain information from the sub-sample by personal interview. The details are provided below.

Let $U = \{U_1, U_2, ..., U_N\}$ be a finite population of size $N$ and a random sample without replacement of size $n$ is taken. We assume that $n_1$ units provided response on the first call and therefore $n_2 = n - n_1$ units did not respond. Then a sub-sample of size $n_s = \frac{n_2}{f} \ (f > 1)$ is taken from the $n_2$ non-response units. Hansen and Hurwitz (1946) used mail survey at the first attempt and then used face-to-face interview at the second attempt. Let $\mu_y = \frac{\sum_{i=1}^{N} y_i}{N}$ and $\sigma_y^2 = \frac{\sum_{i=1}^{N} (y_i - \mu_y)^2}{N-1}$ respectively be the population
mean and variance of the study variable $y$. Let $\mu_{y1} = \frac{\sum_{i=1}^{N_1} y_i}{N_1}$ and $\sigma_{y1}^2 = \frac{\sum_{i=1}^{N_1} (y_i - \mu_{y1})^2}{N_1-1}$ respectively be the mean and variance of response group of size $N_1$, and $\mu_{y2} = \frac{\sum_{i=1}^{N_2} y_i}{N_2}$ and $\sigma_{y2}^2 = \frac{\sum_{i=1}^{N_2} (y_i - \mu_{y2})^2}{N_2-1}$ respectively be the mean and variance of non-response group of size $N_2$. Then the population mean is given by

$$\mu_y = W_1 \mu_{y1} + W_2 \mu_{y2}. \quad (I.26)$$

where $W_1 = \frac{N_1}{N}$ and $W_2 = \frac{N_2}{N}$. Not knowing $N_1$ poses a challenge of its own.

Let $\bar{y}_1 = \frac{\sum_{i=1}^{n_1} y_i}{n_1}$ be the sample mean for the response group, and $\bar{y}_2 = \frac{\sum_{i=1}^{n_2} y_i}{n_2}$ be the sample mean for the non-response group. One can note here that $\bar{y}_1$ and $\bar{y}_2$ are unbiased estimators for $\mu_{y1}$ and $\mu_{y2}$, respectively. But $\bar{y}_1$ has a bias $W_2(\mu_{y1} - \mu_{y2})$ in estimating the population mean of $\mu_y$.

Hansen and Hurwitz (1946) suggested an unbiased population mean estimator given by

$$\hat{\mu}_y = w_1 \bar{y}_1 + w_2 \bar{y}_2. \quad (I.27)$$

where $w_1 = \frac{n_1}{n}$ and $w_2 = \frac{n_2}{n}$. The variance of $\bar{y}$ is given by

$$Var(\hat{\mu}_y) = \left(\frac{N-n}{Nn}\right)\sigma_y^2 + \frac{W_2(f-1)}{n}\sigma_{y2}^2 \quad (I.28)$$

Their results showed that the mean estimation is more efficient and accurate since we obtain more information from the population.

So far we have talked about RRT models and Hansen and Hurwitz (1946) two-phase technique which are used to reduce the social desirability bias and the participation bias, respectively. There is another error called measurement error that we mentioned earlier. We will briefly introduce it in Section I.3 below.
I.3. Measurement Errors

Measurement error is also called observational error which is the difference between observed value and the true value of a variable. It usually can be divided into two components - random error and systematic error. Random errors occur because of random and inherently unpredictable events in the measurement process. Systematic errors are errors that are not determined by chance but are a consequence of a problem in the measurement system that affects all measurements in the same way. For example, scientists study global warming and need to measure the temperatures. If the temperatures were measured with a simple thermometers and the data were recorded by hand, it may cause some random errors because people sometimes make errors in reading the thermometers or recording the temperatures. However, if the scientists’ measurements are mostly temperatures near an urban area, it may cause some systematic error because all the temperatures are probably higher than in the urban area as compared to rural areas. Urban areas tend to be warmer than rural areas because of heat released by human activities. Measurement errors are very common in sampling surveys. We could reduce measurement errors by double checking all the measurements for accuracy, taking average of multiple measurements, and making sure the instrument has the highest precision etc.

Most of the time we assume measurement errors are very small and neglect them. But if measurement errors are not small enough, then we get unreliable estimates. Therefore, we will incorporate measurement errors in our mean estimation.

As mentioned at beginning, a proper sampling method can determine whether or not a sample is truly representative sample or not. In section I.4, two sampling methods will be presented to reflect on this problem.
I.4. Simple Random Sampling and Stratified Random Sampling

Simple random sampling is a sampling method where every sample of the same size has an equal change of being selected. For example, to choose a simple random sample of 10 universities from all the universities in a state, you could assign a number to each university and select a sample by letting a computer randomly generate 10 numbers. This is the most commonly used method because it is likely to provide a representative sample as long as the sample size is large enough. However, when sub-groups within a population vary, simple random sampling may not be a good choice.

Stratified random sampling is a sampling method that is used when researchers are either trying to draw conclusions from different sub-groups or strata that share some common characteristics, or when the population is not very homogeneous. While using stratified random sampling, the population is divided into different strata based on their common characteristics, such as gender, education level, geographic location, nationality and age etc. Then researchers can randomly select a simple random sample from each stratum and the estimates are aggregated over the strata.

Suppose we want to estimate the average income of individuals in a town. Assume that the town has 2000 residents with Master’s degree or above; 3000 residents with college degree; and 5000 residents with high school degree or below. We may choose a simple random sample of size 100 from this town. However, because the incomes are extremely different for people with different education levels, we may get a better estimation if we collect income data from residents with each education level. We could take a proportional random sample of sizes 20, 30, and 50 from education level high to low groups, respectively. In this way, we can create a more representative sample.
Simple random sampling and stratified random sampling are two of the most important and commonly used sampling methods. Therefore, we will study mean estimation under both methods in this dissertation.

I.5. Outline of the Dissertation

Chapter I provided the background of this study and an introduction to the techniques that will be used in the study, including several RRT models, Hansen and Hurwitz (1946) two-phase sampling, and the basic idea about measurement errors.

Chapter II presents the literature review. It includes mean estimations under RRT models, measurement errors, non-response, and stratified random sampling.

Chapter III presents mean estimators under measurement errors using the ORRT model. In this Chapter, the efficiency and the privacy of a general linear combination RRT model and a simple additive RRT model will be compared. A better criterion factoring in both efficiency and privacy of a RRT model is used for the entire study. A simulation study is also conducted to show the performance of various mean estimators.

Chapter IV presents mean estimators under simultaneous presence of measurement errors and non-response using ORRT model. We will introduce a modified version of Hansen and Hurwize two-phase sampling, then study some mean estimators using this new technique. A simulation study is also conducted to show the performance of the mean estimators.

Chapter V presents mean estimators under measurement errors and non-response under stratified random sampling using ORRT model. A simulation study is also conducted to show the performance of the mean estimators under stratified sampling.
Chapter VI presents a general discussion on the research carried out in the dissertation and some future directions.
CHAPTER II
LITERATURE REVIEW

In Chapter I, we discussed the background and the techniques that will be used in the study. Some literature review will be presented in Chapter II. All the literature review revolves around the topic of mean estimation, but under different conditions. We divided this chapter into four parts - mean estimation under RRT models; RRT models and measurement errors; non-response; and stratified random sampling.

II.1. Mean Estimation under RRT Models

Researchers have been working on the mean estimation for sensitive variables for years. They have discussed different estimators under the same RRT model or the same estimator under different RRT models. Many researchers have studied the mean estimation when the primary variable is sensitive and there is no auxiliary variables. These include Gupta and Shabbir (2004)[18], Gupta et al. (2002, 2010)[17][19], Wu et al. (2008)[86], Saha (2008)[56], and Perri(2008)[52] etc. Also, many others have used auxiliary information to improve the efficiency of the estimators, such as Kadilar and Cingi (2005, 2006)[28][29], Kadilar et al. (2007)[30], Shabbir & Gupta (2007, 2010)[59][60], Turgut and Cingi(2008)[83], Nangsue(2009)[51], Koyuncu and Kadilar (2009)[44], Sousa et al. (2010)[77], Subramani and Kumarapandiyam (2012)[81], Gupta et al. (2012, 2015, 2016) [20][22][23], Tarray & Singh (2015) [82], Kalucha et al. (2015)[31], and Zhang et al. (2018)[89] etc.

In this dissertation, we focus on estimating the mean of a sensitive variable using non-sensitive auxiliary variable that is highly correlated with the primary variable.
In this section, we will discuss in detail some existing mean estimators using RRT or ORRT models.

II.1.1 Mean Estimation under RRT Models (Pollock & Bek 1976 and Sousa et al. 2010)

Ratio and product estimators provide more accurate estimates than the ordinary mean estimator when an auxiliary variable exists that is highly correlated with the study variable. In sample surveys, there are some situations when the variable of interest (Y) is sensitive but there is a nonsensitive auxiliary variable (X) which is highly correlated with it. For example, Y may be the number of sexual partners a woman might have had in her life and X may be her age. In such cases, one can estimate mean of Y using one of the RRT models and improve the estimator by using auxiliary information.

Sousa et al. (2010) [77] proposed a ratio estimator where the mean of Y is estimated using the Pollock & Bek (1976) RRT model and it is further improved by an auxiliary variable X. Again, let Y be the sensitive variable of interest. Let X be a non-sensitive auxiliary variable which is observed directly and also is positively correlated with Y. We assume that the mean $\mu_x$ and variance $\sigma_x^2$ for X are known. Let S be a scrambling variable independent of Y and X. The respondents are asked to provide scrambled responses for Y given by $Z=Y+S$ but report the true responses for X. We assume the population mean of X $\mu_x$ is known. And population mean of S, $\mu_s = E(S)=0$. Thus, $E(Z)=E(Y)$. 


If the auxiliary variable X is ignored, then an unbiased ordinary estimator of $\mu_y$ is given by

$$\hat{\mu}_o = \bar{z} \quad (\text{II.1})$$

and the MSE of $\hat{\mu}_0$ is given by

$$MSE(\hat{\mu}_0) = \lambda(\sigma^2_y + \sigma^2_s), \quad (\text{II.2})$$

where $\lambda = (N - n)/Nn$.

The proposed ratio estimator for the population mean of Y using the auxiliary variable X is given by

$$\hat{\mu}_R = \bar{z} \frac{\mu_x}{\bar{x}} \quad (\text{II.3})$$

The MSE of the estimator $\hat{\mu}_R$, correct up to first order of approximation, is given by

$$MSE^{(1)}(\hat{\mu}_R) \approx \lambda \mu^2_y(C^2_z + C^2_x - 2\rho_{xz}C_zC_x), \quad (\text{II.4})$$

and correct up to second order of approximation, is given by

$$MSE^{(2)}(\hat{\mu}_R) \approx MSE^{(1)}(\hat{\mu}_R) + 3\mu^2_y\lambda^2C^2_x[(1 + 2\rho^2_{xz})C^2_z + 3C^2_x - 6\rho_{xz}C_zC_x], \quad (\text{II.5})$$

where $\lambda = \frac{N-n}{Nn}$.

Comparing the first order of approximation in (II.4) and (II.2), Sousa et al. (2010) [77] showed that the ratio estimator $\hat{\mu}_R$ is more efficient than the RRT mean estimator $\hat{\mu}_0$ when Y and X have strong positive correlation.
II.1.2 Mean Estimation under ORRT Models (Kalucha et al. 2015 and Zhang et al. 2018)

Sousa et al. (2010) [77] was the first to use ratio estimators under RRT models. They estimated $\mu_y$ using a non-optional RRT model with the utilization of a non-sensitive auxiliary variable. However, Gupta et al. (2002) [17] introduced ORRT models and showed that they perform better than non-optional RRT. Based on this result, Gupta et al. (2014) [21] improved Sousa et al. (2010) [77] by using optional scrambling. Additionally, in the Gupta et al. (2010) [19], they use a split-sample approach using different scrambling variables in the two sub-samples. Kalucha et al. (2015) [31] and Zhang et al (2018) [89] also improved the Sousa et al. (2010) estimator further by using a split sample ORRT model.

If a proportion $W$ of the respondents feel the survey question is sensitive, then according to Gupta et al.(2010), the reported response $Z_i$ in the $i^{th}$ sub-sample is given by

$$Z_i = \begin{cases} Y & \text{with probability } 1-W, \\ Y + S_i & \text{with probability } W \end{cases}, \quad i = 1, 2. \quad (II.6)$$

Kalucha et al. (2015) [31] proposed two ratio estimators of finite population mean using ORRT model and called them the additive ratio estimator and the multiplicative ratio estimator, respectively. Let $\bar{x}_i$ and $\bar{z}_i$ ($i=1, 2$) respectively be the means of the auxiliary variable and the reported response in the $i^{th}$ sub-sample. These estimators with associated MSEs, correct up to the first order of approximation, are given by:

$$\hat{\mu}_{AR} = \left( \frac{\mu_{s_2} \bar{z}_1 - \mu_{s_1} \bar{z}_2}{\mu_{s_2} - \mu_{s_1}} \right) \left( \frac{\mu_x}{\bar{x}_1} + \frac{\mu_x}{\bar{x}_2} \right) \left( \frac{1}{2} \right), \quad (II.7)$$
\[
\hat{\mu}_{MR} = \left( \frac{\mu_{s_2} \bar{z}_1 - \mu_{s_1} \bar{z}_2}{\mu_{s_2} - \mu_{s_1}} \right) \left( \frac{\mu_x}{\bar{x}_1} \right) \left( \frac{\mu_x}{\bar{x}_2} \right),
\]

(II.8)

\[
MSE^{(1)}(\hat{\mu}_{AR}) = E(\mu_{AR} - \mu_y)^2 \\
\approx \lambda_1 \left[ \left( \frac{\mu_{s_2}}{\mu_{s_2} - \mu_{s_1}} \right)^2 \sigma_{z_1}^2 + \frac{1}{4} \mu_y C_x^2 - \mu_y \rho_{yx} \sigma_y \left( \frac{\mu_{s_2}}{\mu_{s_2} - \mu_{s_1}} \right) C_x \right] + \lambda_2 \left[ \left( \frac{\mu_{s_1}}{\mu_{s_2} - \mu_{s_1}} \right)^2 \sigma_{z_2}^2 + \frac{1}{4} \mu_y C_x^2 + \mu_y \rho_{yx} \sigma_y \left( \frac{\mu_{s_1}}{\mu_{s_2} - \mu_{s_1}} \right) C_x \right],
\]

(II.9)

and

\[
MSE^{(1)}(\hat{\mu}_{MR}) = E(\mu_{MR} - \mu_y)^2 \\
\approx \lambda_1 \left[ \left( \frac{\mu_{s_2}}{\mu_{s_2} - \mu_{s_1}} \right)^2 \sigma_{z_1}^2 + \frac{1}{4} \mu_y C_x^2 - 2 \mu_y \rho_{yx} \sigma_y \left( \frac{\mu_{s_2}}{\mu_{s_2} - \mu_{s_1}} \right) C_x \right] + \lambda_2 \left[ \left( \frac{\mu_{s_1}}{\mu_{s_2} - \mu_{s_1}} \right)^2 \sigma_{z_2}^2 + \frac{1}{4} \mu_y C_x^2 + 2 \mu_y \rho_{yx} \sigma_y \left( \frac{\mu_{s_1}}{\mu_{s_2} - \mu_{s_1}} \right) C_x \right].
\]

(II.10)

Kalucha et al. (2015) showed that the additive ratio estimator \( \hat{\mu}_{AR} \) is more efficient than the ordinary RRT estimator \( \hat{\mu}_y \) when the correlation between the study variable and the auxiliary variable is greater than \( \frac{1}{2} \). However, the multiplicative ratio estimator was not found to be as efficient as the ordinary RRT estimator (\( \hat{\mu}_y \)) or the additive ratio estimator (\( \hat{\mu}_{AR} \)). But Zhang et al (2018) [89] modified the multiplicative ratio estimator in (II.8) and proposed a new geometric mean ratio estimator. It is given by

\[
\hat{\mu}_{GMR} = \left( \frac{\mu_{s_2} \bar{z}_1 - \mu_{s_1} \bar{z}_2}{\mu_{s_2} - \mu_{s_1}} \right) \sqrt{\left( \frac{\mu_x}{\bar{x}_1} \right) \left( \frac{\mu_x}{\bar{x}_2} \right)},
\]

(II.11)

The MSE of \( \hat{\mu}_{GMR} \), correct up to the first order of approximation, is given by

\[
MSE^{(1)}(\hat{\mu}_{GMR}) \approx \lambda_1 \left[ \left( \frac{\mu_{s_2}}{\mu_{s_2} - \mu_{s_1}} \right)^2 \sigma_{z_1}^2 + \frac{1}{4} \mu_y C_x^2 - \mu_y \rho_{yx} \sigma_y \left( \frac{\mu_{s_2}}{\mu_{s_2} - \mu_{s_1}} \right) C_x \right] + \lambda_2 \left[ \left( \frac{\mu_{s_1}}{\mu_{s_2} - \mu_{s_1}} \right)^2 \sigma_{z_2}^2 + \frac{1}{4} \mu_y C_x^2 + \mu_y \rho_{yx} \sigma_y \left( \frac{\mu_{s_1}}{\mu_{s_2} - \mu_{s_1}} \right) C_x \right].
\]

(II.12)
Comparing the MSE of the geometric mean ratio estimator with Kalucha et al. ratio estimators and the ordinary mean estimator in both the equal and the unequal sample split, they concluded that up to the first order approximation:

- The geometric mean ratio estimator is always more efficient than the multiplicative ratio estimator.
- The geometric mean ratio estimator is more efficient than the ordinary RRT estimator when the correlation coefficient between X and Y is greater than $\frac{1}{2}$.
- The geometric mean ratio estimator is as efficient as the additive ratio estimator up to the first order of approximation.

Since the MSE of the geometric mean ratio estimator, up to the first order of approximation, is same as that of the additive ratio estimator, the biases of these two estimators are compared. One can verify that

$$\text{Bias}^{(1)}(\hat{\mu}_{GMR}) - \text{Bias}^{(1)}(\hat{\mu}_{AR}) = -\frac{1}{8}\mu_y C_x^2 (\lambda_1 + \lambda_2),$$

which means the geometric mean ratio estimator has less bias if the $\mu_y$ is positive.

II.2. Mean Estimation under RRT Models and Measurement Errors

As in Section II.1, there are lots of population mean estimators that have proposed under RRT models. In addition to social desirability response bias, there are some other non-sampling errors such as measurement errors that may also affect the population mean estimation. Many researches have studied measurement errors while utilizing auxiliary information, including Shalabh (1997)[62], Manisha and Singh (2001) [48], Srivastava and Shalabh (2001) [80], Allen et al. (2003) [02], Singh and

Although Blattman et al (2014) [04] developed a survey validation technique for qualitative variables to check for measurement errors when dealing with sensitive attributes, not much effort has been devoted to estimating the finite population mean of a sensitive variable in the presence of measurement errors. We know that use of RRT models reduces non-sampling errors when the variable of interest is sensitive, one may also want to check the impact of measurement errors. Khalil et al. (2018) [34] studied mean estimation for sensitive variables in the presence of measurement errors under a non-optional RRT model. The details are provided below.

II.2.1 Mean Estimation under RRT Models in the Presence of Measurement Errors (Khalil et al. 2018)

According to Pollock & Bek (1976) RRT model, the respondent is asked to provide a scrambled value for Y given by $Z = Y + S$, and report a true response for the non-sensitive auxiliary variable $X$. Let the measurement errors for the scrambled response variable ($Z$) and the auxiliary variable ($X$) on $i_{th}$ unit respectively be $U_i$ and $V_i$. $U_i$ and $V_i$ are assumed to be random and independent with mean zero and variance $\sigma_u^2$ and $\sigma_v^2$ respectively.

There are some commonly used existing mean estimators and their MSEs in the presence of measurement errors:

- The ordinary RRT mean estimator is given by

\[
\hat{\mu}_0 = \frac{\sum_{i=1}^{n} z_i}{n} = \bar{z}.
\] (II.14)
The MSE of \( \hat{\mu}_0 \) is given by

\[
MSE^*(\hat{\mu}_0) = \lambda(\sigma_z^2 + \sigma_u^2),
\]

where \( \lambda = (N - n)/Nn. \)

- A ratio estimator proposed by Sousa et al. (2010) is given by

\[
\hat{\mu}_R = \bar{z} \frac{\mu_x}{\bar{x}}.
\]

The MSE of \( \hat{\mu}_R \) is given by

\[
MSE^*(\hat{\mu}_R) = \lambda(\sigma_z^2 + R^2\sigma_x^2 - 2R\rho_{zx}\sigma_x\sigma_z) + \lambda(\sigma_u^2 + R^2\sigma_v^2),
\]

where \( R = \mu_z/\mu_x \).

Khalil et al. (2018) generalized the estimator in (II.16) and proposed a generalized randomized response estimator for the mean of a sensitive study variable \( Y \) in the presence of a highly correlated (positively) auxiliary variable and measurement errors. The proposed estimator is given by

\[
\hat{\mu}_N = (\bar{z} + k(\mu_x - \bar{x}))(\frac{\bar{W}}{\bar{w}})^g,
\]

where \( \bar{w} = \phi(\alpha\bar{x} + \beta) + (1 - \phi)(\alpha\mu_x + \beta) \), \( \bar{W} = \alpha\mu_x + \beta \), \( k \) and \( g \) are suitable constants, and \( \phi \) is assumed to be an unknown constant whose value is to be determined from optimality considerations. \( \alpha \) (\( \alpha \neq 0 \)) and \( \beta \) are assumed to be some known parameters of the auxiliary variable \( X \), such as coefficient of variation \( (C_x) \), kurtosis, and correlation coefficient \( (\rho_{zx}) \) etc.
The minimum MSE of $\hat{\mu}_N$, correct up to the first order of approximation, is given by

$$MSE_{\min}(\hat{\mu}_N) \approx \lambda(\sigma^2_z + \sigma^2_u - \frac{\rho_{xz}^2 \sigma^2_x \sigma^2_z}{\sigma^2_x + \sigma^2_z}). \tag{II.19}$$

Their result showed that MSE increases when measurement errors exist. And the generalized estimator ($\hat{\mu}_N$) is more efficient than the ordinary RRT mean estimator ($\hat{\mu}_o$) and ratio estimator ($\hat{\mu}_R$) both with and without measurement errors, particularly if $Y$ and $X$ are highly correlated.

Khalil et al. in their study used what are known as full RRT (or non-optional RRT models) where all respondents provide a scrambled response. As mentioned earlier, Gupta et al. (2002) [17] ORRT model is generally more efficient than the corresponding non-optional RRT model. Also, in a recent publication by Gupta et al. (2018) [24], it is shown that there is no extra loss of privacy in using ORRT models as compared to the corresponding RRT models. So why not use ORRT models instead of full RRT model? Using ORRT model in this situation has not been done by anyone so far. This is our main motivation for this dissertation. The work will be introduced in Chapter III.

II.3. Mean Estimation under Non-response

Non-response is another common non-sampling error we have seen in sampling. The problem of non-response has been discussed in many papers. These include Hansen and Hurwitz (1946) [25], Foradori (1961) [12], Srinath (1971) [79], Khare and Srivastava (1993, 1995, 1997, and 2010) [37][38][39][40], Singh and Kumar (2008, 2009, and 2011) [67][69][73], Singh et al. (2010) [71], Kumar and Bhoulag (2011) [47], Shabbir and Khan (2013) [61], Singh and Sharma (2015)[76], and Azzem and Hanif
Most of these researchers suggested different types of estimators for population parameters based on Hansen and Hurwitz (1946) double sampling plan. Some of them also used different conditions such as mean estimation in the presence of both non-response and measurement errors, or mean estimation under non-response and RRT models.

**II.3.1 Mean Estimation under Non-response and Measurement Errors**

(Singh and Sharma 2015)

Singh and Sharma (2015) have studied the problem of estimating the finite population mean in the presence of non-response and measurement errors. In their study, they assume non-response and measurement errors happened on both the study and auxiliary variables and utilized Hansen and Hurwitz (1946) two-phase sampling. Assume that $n_1$ units provided response on the first call and $n_2 = n - n_1$ units did not respond. Then a sub-sample of size $n_s = \frac{n_2}{f}$ ($f > 1$) is taken from the $n_2$ non-responding units. Let $N_1$ and $N_2$ respectively be the sizes of the respondent group and the non-respondent group in the population. The proportions of the response group and the non-response group in the population are $W_1 = \frac{N_1}{N}$ and $W_2 = \frac{N_2}{N}$, respectively. Let the first phase measurement errors on the study variable $Y$ and auxiliary variable $X$ on the $i^{th}$ unit be $U_i$ and $V_i$; and let the second phase measurement errors of $Y$ and $X$ on the $i^{th}$ unit be $U_{2i}$ and $V_{2i}$. Assume these measurement errors are random and independent with variances of $\sigma_u^2$, $\sigma_v^2$, $\sigma_{u2}^2$ and $\sigma_{v2}^2$, respectively. In order to compare efficiency of mean estimators, some adapted estimators under both Hansen and Hurwitz (1946) two-phase sampling and measurement errors are here.
• The ordinary mean estimator is given by

\[ \hat{\mu}_0 = w_1 \bar{y}_1 + w_2 \bar{y}_2, \]  

(II.20)

where \( w_1 = \frac{n_1}{n} \) and \( w_2 = \frac{n_2}{n} \). The expected value of \( \hat{\mu}_0 \) is given by

\[ E(\hat{\mu}_0) = W_1 \mu_{y_1} + W_2 \mu_{y_2} = \mu_y. \]  

(II.21)

The variance of \( \hat{\mu}_0 \) under measurement errors is given by

\[ \text{Var}(\hat{\mu}_0) = \theta (\sigma_y^2 + \sigma_u^2) + \lambda (\sigma_{y(2)}^2 + \sigma_{u(2)}^2), \]  

(II.22)

where \( \theta = \frac{N-n}{Nn} \) and \( \lambda = \frac{W_2(f-1)}{n} \).

• A ratio estimator is given by

\[ \hat{\mu}_R = \bar{y}^* \bar{x}^* \mu_x. \]  

(II.23)

The MSE of \( \hat{\mu}_R \) under measurement errors is given by

\[ \text{MSE}(\hat{\mu}_R) = \theta \mu_y^2 (C_y^2 + C_x^2 + \frac{\sigma_y^2}{\mu_y^2} + \frac{\sigma_u^2}{\mu_x^2} - 2 \rho_{yx} C_y C_x) + \lambda \mu_y^2 (\sigma_{y(2)}^2 + \sigma_{x(2)}^2 + \frac{\sigma_{y(2)}^2}{\mu_y^2} + \frac{\sigma_{x(2)}^2}{\mu_x^2}). \]  

(II.24)

Singh and Sharma (2015) proposed a class of estimators given by

\[ \hat{\mu}_P = m_1 \bar{y}^* + m_2 \frac{\bar{y}^*}{\bar{x}^*} \mu_x. \]  

(II.25)

Their proposed class of estimators is a specific version of the ordinary estimator and ratio estimator if we let \((m_1, m_2) = (1, 0)\) and \((m_1, m_2) = (0, 1)\), respectively. By taking the optimum values of \((m_1, m_2)\) with \( m_2^* = \frac{1}{R} \left( \frac{O}{M} \right) \) and \( m_1^* = 1 - m_2^* \), where \( R \)
\[ \mu_y, \text{ the minimum MSE of the mean estimator } \hat{\mu}_P \text{ is given by} \]

\[ MSE(\hat{\mu}_P) = M[1 - \frac{O^2}{MN}] \quad (\text{II.26}) \]

where \( M = \frac{1}{n}(\sigma^2_y + \sigma^2_u) + \frac{(k-1)W_2}{n}(\sigma^2_{y(2)} + \sigma^2_{u(2)}) \), \( N = \frac{1}{n}(\sigma^2_x + \sigma^2_v) + \frac{(k-1)W_2}{n}(\sigma^2_{x(2)} + \sigma^2_{v(2)}) \)

and \( O = \frac{1}{n}\rho_{yx}\sigma_x\sigma_y + \frac{(k-1)W_2}{n}\rho_{xy2}\sigma_x(2)\sigma_y(2) \). Their results show that \( \hat{\mu}_P \) is more efficient than \( \hat{\mu}_0 \) if

\[ M - M(1 - \frac{O^2}{MN}) = \frac{O^2}{MN} > 0; \quad (\text{II.27}) \]

and \( \hat{\mu}_P \) is more efficient than \( \hat{\mu}_R \) if

\[ (M + N - 2O) - M(1 - \frac{O^2}{MN}) = N - 2O + \frac{O^2}{MN} > 0. \quad (\text{II.28}) \]

### II.3.2 Mean Estimation under Non-response and RRT Models

As we mentioned in the previous sections, the respondents are unlikely to provide true response in face-to-face interview if the survey question is sensitive. To reduce the bias caused by sensitive questions, one could use randomized response technique (RRT) models when we target the non-response group. Respondents may refuse to respond on the first call but may provide scrambled response on the second call with personal interview. Diana et al. (2014)[09] proposed an unbiased population mean estimator under Hansen and Hurwitz (1946) two-phase sampling. Their estimator reduces non-response but increases the estimator variance due to the use of RRT model in the non-respondent group. Later, Ahmed et al. (2017) [01] proposed generalized ratio and regression estimators utilizing known coefficient of variation of the study variable in case of second sample by using RRT approach. This estimator improved the efficiency when the auxiliary variable and the study variable are highly correlated. Here we only discuss Diana et al. (2014) in detail.
Diana et al. (2014) used a RRT model in the second phase non-respondents group where the scrambled response is given by $Z = T Y + S$. T and S are two scrambling variables that are independent of Y. Then a modified version of Hansen and Hurwitz (1946) estimator is given by

$$\hat{\mu}_{0HH} = w_1 \bar{y}_1 + w_2 \hat{y}_2,$$  \hspace{1cm} (II.29)

where $\hat{y}_2 = \sum_{i=1}^{n_x} (\frac{z_i}{n_x})$ and $z_i$ is the scrambled response from the second face-to-face interview step. The variance of the unbiased estimator $\hat{y}^*$ is given by

$$\text{Var}(\hat{\mu}_{0HH}) = \theta \sigma_y^2 + \lambda \sigma_{y(2)}^2 + G,$$  \hspace{1cm} (II.30)

where $G = \frac{W_{th}}{n} \left[ \frac{\sigma_y^2 (\sigma_y^2 + \mu_{y(2)}^2) + \sigma_s^2 + 2 \sigma_{st} \mu_x^2}{\mu_y^2} \right]$ is the penalty for using RRT models.

They suggested a regression estimator where auxiliary information is used. Assume the population mean $\mu_x$ of the auxiliary variable is known and non-response only happened on Y. The estimator is given by

$$\hat{\mu}_{\text{regHH}} = \hat{\mu}_{0HH} + \hat{\beta}^*_{yx} (\mu_x - \bar{x}),$$  \hspace{1cm} (II.31)

where $\hat{\mu}_{0HH}$ is the modified version of the Hansen and Hurwitz (1946) estimator and $\hat{\beta}^*_{yx} = \frac{\hat{\delta}_{yx}}{\hat{s}_{yx}^2}$. The MSE of $\hat{\mu}_{\text{regHH}}$ is given by

$$\text{MSE}(\hat{\mu}_{\text{regHH}}) = \theta \sigma_y^2 (1 - \rho_{yx}^2) + \lambda \sigma_{y(2)}^2 + G.$$  \hspace{1cm} (II.32)

Diana et al. (2014) also considered a situation when non-response is present in the auxiliary variable X, and also suggested a regression estimator

$$\hat{\mu}_{\text{regHH1}} = \hat{\mu}_{0HH} + \hat{\beta}^{**}_{yx} (\mu_x - \bar{x}^*),$$  \hspace{1cm} (II.33)
where $\hat{\beta}_{yx}^{**} = \frac{\hat{\sigma}_{yx}^2}{\sigma_x^2}$. The MSE of $\hat{\mu}_{regHH}$ is given by

$$MSE(\hat{\mu}_{regHH}) = \theta \sigma_y^2 (1 - \rho_{yx}^2) + \lambda (\sigma_{y(2)}^2 + \beta_{yx}^2 \sigma_{x(2)} - 2 \beta_{yx} \sigma_{yx(x)}) + \lambda \sigma_{y(2)}^2 + G.$$  

(II.34)

Researchers have studied mean estimation under non-response alone, both non-response and measurement errors, and both non-response and RRT models. But not many researchers have explored the performance of mean estimators for a sensitive variable under both non-response and measurement errors using ORRT model. Adding non-response to mean estimation for a sensitive question in the presence of measurement errors will be our second major motivation for this study. The work will be introduced in Chapter IV.

II.4. Mean Estimation under Stratified Random Sampling

In addition to simple random sampling, stratified random sampling is another commonly used method when subpopulations within an overall population vary. Much work has been done when study variables are directly observed in stratified random sampling, including Kadilar and Cingi (2003, 2005)[26][27], Shabbir and Gupta (2005,2006) [57][58], Koyuncu and Kadilar (2008, 2009, 2010)[41][42][43], Singh and Karpe (2010)[72], Zahid and Shabbir (2018)[88], and Khalil et al. (2017) [33] etc.. Some of this work also involves non-response, measurement errors, and sensitive questions.
II.4.1 Mean Estimation under Stratified Random Sampling in the Presence of Measurement Errors and Non-response (Zahid and Shabbir 2018)

Zahid and Shabbir (2018)[88] proposed a class of estimators in the presence of measurement errors and non-response under stratified random sampling. Assume measurement errors are found in both the study variable $Y$ and the auxiliary variable $X$ and non-response happened in each stratum. Their study used Hansen and Hurwitz (1946) two-phase sampling to reduce the impact of non-sampling errors caused by non-response in each stratum. Let a finite population $U = (U_1, U_2, U_3, ..., U_N)$ be divided into $L$ homogeneous strata, and $N_h$ represent the number of units in stratum $h$ such that $\sum_{h=1}^{L} N_h = N$. Let $X$ and $Y$ have population means $\mu_{xh} = \frac{1}{N_h} \sum_{i=1}^{N_h} x_{hi}$ and $\mu_{yh} = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi}$ respectively in stratum $h$. Let the respective measurement errors on the study variable $Y$ and the auxiliary variable $X$ in the $h^{th}$ stratum be given by $U_{hi}$ and $V_{hi}$. These measurement errors are assumed to be uncorrelated and having normal distribution with zero mean and variance $\sigma_{uh}^2$ and $\sigma_{vh}^2$, respectively. It is also assumed that the measurement errors are independent of $Y$ and $X$. Under Hansen and Hurwize two-phases sampling, $n_{1h}$ units provided response on the first call and remaining $n_{2h} = n_h - n_{1h}$ units do not respond. Then a sub-sample of size $n_{sh} = \frac{n_{2h}}{f_h}$ ($f_h > 1$) is taken from the $n_{2h}$ non-response units in the $h^{th}$ stratum.

The study provided some existing estimators under stratified random sampling and measurement errors as discussed below.

- The ordinary Hansen and Hurwitz (1946) mean estimator is given by

$$\hat{\mu}_o^* = \sum_{h=1}^{L} W_h \bar{y}_h^*, \quad (\text{II.35})$$

where $\bar{y}_h^* = \left(\frac{n_{1h}}{n_h}\right)\bar{y}_{1h} + \left(\frac{n_{2h}}{n_h}\right)\bar{y}_{2h}$ and $W_h = N_h/N$. The expected value of $\hat{\mu}_o^*$ is
given by

\[ E(\hat{\mu}_o^*) = \sum_{h=1}^{L} W_h \mu_{yh} = \mu_y. \]  

(II.36)

The variance of \( \hat{\mu}_o^* \) is given by

\[ \text{Var}(\hat{\mu}_o^*) = \sum_{h=1}^{L} W_h^2 \left[ \theta_h (\sigma_{yh}^2 + \sigma_{uh}^2) + \lambda_h (\sigma_{(2)h}^2 + \sigma_{u(2)h}^2) \right], \]  

(II.37)

where \( \theta_h = \frac{N_h - n_h}{N_h n_h} \), \( \lambda_h = \frac{N_h (f_h - 1)}{N_h n_h} \).

- The ratio estimator is given by

\[ \hat{\mu}_r^* = \sum_{h=1}^{L} W_h \frac{\bar{y}_h^*}{\bar{x}_h} \mu_{xh}. \]  

(II.38)

The MSE of \( \hat{\mu}_r^* \) is given by

\[ \text{MSE}(\hat{\mu}_r^*) = \sum_{h=1}^{L} W_h^2 \left[ \theta_h (\sigma_{yh}^2 + \sigma_{xh}^2) + \lambda_h (\sigma_{x(2)h}^2 + \sigma_{v(2)h}^2) + \right. \]

\[ R_h^2 (\theta_h (\sigma_{xh}^2 + \sigma_{v(2)h}^2) + \lambda_h (\sigma_{x(2)h}^2 + \sigma_{v(2)h}^2)) - \]

\[ 2R_h (\theta_h \rho_{yhx} \sigma_{yhx} xh + \lambda_h \rho_{xhx} xh), \]  

(II.39)

where \( R_h = \frac{\mu_{yh}}{\mu_{xh}} \).

Zahid and Shabbir (2018) proposed a class of estimators given by

\[ \hat{\mu}_{ZS}^* = \sum_{h=1}^{L} W_h \left[ m_{1h} \bar{y}_h^* + m_{2h} (\mu_{xh} - \bar{x}_h^*) \left( \frac{\mu_{xh}}{\bar{x}_h} \right)^{\alpha_h} \exp(1 - \alpha_h) \left( \frac{\mu_{xh} - \bar{x}_h^*}{\mu_{xh} + \bar{x}_h^*} \right) \right], \]  

(II.40)

where \( m_{1h} \) and \( m_{2h} \) are constants whose values are to be determined and \( \alpha_h \) is a scalar.
By substituting optimal values of \( m_{1h} \) and \( m_{2h} \), the minimum MSE of \( \hat{\mu}_{ZS}^* \) is given by

\[
MSE(\hat{\mu}_{ZS}^*) = \sum_{h=1}^{L} W_h^2 \mu_{y_h}^2 - \frac{A_{h1}E_{h1}^2 + B_{h1}D_{h1}^2 - 2C_{h1}D_{h1}E_{h1}}{A_{h1}B_{h1} - C_{h1}^2}, \tag{II.41}
\]

where

\[
A_{h1} = \mu_{y_h}^2 + A_h + e_{h1}^2 t_h^2 R_h C_h + 4 e_{h1} t_h R_h^2 C_h + 2 f_{h1} t_h^2 R_h^2 B_h,
\]

\[
B_{h1} = t_h^2 B_h, \quad C_{h1} = t_h C_h + 2 e_{h1} t_h R_h B_h,
\]

\[
D_{h1} = \mu_{y_h}^2 + e_{h1} t_h R_h^2 C_h + f_{h1} t_h^2 R_h^2 B_h,
\]

\[
E_{h1} = e_{h1} t_h^2 R_h B_h,
\]

\[
A_h = \theta_h (\sigma_{y_h}^2 + \sigma_{u_h}^2) + \lambda_h (\sigma_{y(2)h}^2 + \sigma_{u(2)h}^2),
\]

\[
B_h = \theta_h (\sigma_{x_h}^2 + \sigma_{v_h}^2) + \lambda_h (\sigma_{x(2)h}^2 + \sigma_{v(2)h}^2),
\]

and \( C_h = \theta_h \rho_{yxh} \sigma_{y_h} \sigma_{x_h} + \lambda_h \rho_{yx(2)} \).

The proposed estimator \( \hat{\mu}_{ZS}^* \) is more efficient than the ordinary and ratio estimators when the following conditions hold:

- \( MSE(\hat{\mu}_{ZS}^*) < Var(\hat{\mu}_{o}^*) \) if
  \[
  \sum_{h=1}^{L} W_h^2 \mu_{y_h}^2 - \sum_{h=1}^{L} W_h^2 \frac{A_{h1}E_{h1}^2 + B_{h1}D_{h1}^2 - 2C_{h1}D_{h1}E_{h1}}{A_{h1}B_{h1} - C_{h1}^2} - \sum_{h=1}^{L} W_h^2 A_h < 0
  \]

- \( MSE(\hat{\mu}_{ZS}^*) < MSE(\hat{\mu}_r^*) \) if
  \[
  \sum_{h=1}^{L} W_h^2 \mu_{y_h}^2 - \sum_{h=1}^{L} W_h^2 \frac{A_{h1}E_{h1}^2 + B_{h1}D_{h1}^2 - 2C_{h1}D_{h1}E_{h1}}{A_{h1}B_{h1} - C_{h1}^2} - \sum_{h=1}^{L} W_h^2 (A_h R_h^2 B_h - 2 R_h C_h) < 0
  \]

**II.4.2 Mean Estimation under Stratified Random Sampling using RRT in the Presence of Measurement Errors (Khalil et al. 2018)**

In Chapter II.2, we introduced Khalil et al. (2018) [34] mean estimation of a sensitive variable in the presence of measurement errors. The study used simple random sampling. It has been further extended to stratified random sampling in
Kahlil et al (2018)[35]. In the new study, they have modified some of the existing mean estimators in the context of measurement errors under stratified random sampling:

- The ordinary mean estimator:
  \[
  \hat{\mu}^{st}_o = \bar{z}^{st} = \sum_{h=1}^{L} W_h \bar{z}_h.
  \]  
  The MSE of \( \hat{\mu}^{st}_o \) is given by
  \[
  \text{MSE}(\hat{\mu}^{st}_o) = \sum_{h=1}^{L} W_h^2 \theta_h \left( \frac{\sigma^2_{zh}}{\gamma_{zh}} \right),
  \]
  where \( \theta_h = \frac{N_h - n_h}{N_h n_h} \) and \( \gamma_{zh} = \frac{\sigma^2_{zh}}{\sigma^2_{zh} + \sigma^2_{uh}} \).

- The ratio estimator:
  \[
  \hat{\mu}^{st}_r = \frac{\bar{z}^{st}_{st}}{\bar{x}^{st}} \mu_x = \sum_{h=1}^{L} W_h \frac{\bar{z}_h}{\bar{x}_h} \mu_x.
  \]  
  The MSE of \( \hat{\mu}^{st}_r \) is given by
  \[
  \text{MSE}(\hat{\mu}^{st}_r) = \sum_{h=1}^{L} W_h^2 \theta_h \left[ \frac{\sigma^2_{zh}}{\gamma_{zh}} + R \frac{\sigma^2_{xh}}{\gamma_{xh}} (R - 2 \beta_{xh} \gamma_{xh}) \right],
  \]
  where \( \gamma_{zh} = \frac{\sigma^2_{xh}}{\sigma^2_{zh} \gamma_{xh}} \), \( \gamma_{xh} = \frac{\sigma^2_{zh}}{\sigma^2_{zh} + \sigma^2_{uh}} \), and \( R = \frac{\mu_y}{\mu_x} \).

Kahlil et al (2018)[35] proposed a generalized mean estimator for the mean of a sensitive study variable \( Y \) in the presence of measurement error, which is given by

\[
\hat{\mu}^{st}_{GE} = [\bar{z}^{st} + k(\mu_x - \bar{x}^{st})] \left[ \frac{\alpha_{st} \mu_x + \beta_{st}}{w(\alpha_{st} \bar{x}^{st} + \beta_{st}) + (1 - w)(\alpha_{st} \mu_x + \beta_{st})} \right]^{g},
\]

where \( k \) and \( g \) are suitable constants, and \( w \) is an unknown constant whose value is to be determined from optimality consideration. \( \alpha_{st} (\alpha_{st} \neq 0) \) and \( \beta_{st} \) are assumed to be some known parameters of the auxiliary variable \( X \), such as coefficient of variation.
(C_x), kurtosis, and correlation coefficient (\rho_{zx}) etc. The optimum value of gw\phi which gives the minimum MSE is given by

\[(gw\phi)_{opt} = \frac{\mu_x}{\mu_y} \left( \frac{\sum_{h=1}^{L} W_h^2 \gamma_{xh} \sigma_{xh}}{\sum_{h=1}^{L} W_h^2 \gamma_{xh} \sigma_{xh}^2 / \theta_{xh}} - k \right) \]  \hspace{1cm} \text{(II.47)}

By substituting optimal value of \((gw\phi)_{opt}\), the minimum value of \(MSE^*(\hat{\mu}_{st GE})\) is given by

\[MSE^*(\hat{\mu}_{st GE}) \approx \sum_{h=1}^{L} \frac{W_h^2 \gamma_{xh} \sigma_{xh}^2}{\theta_{xh}} (1 - \rho_c^2), \]  \hspace{1cm} \text{(II.48)}

where

\[\rho_c = \frac{\sum_{h=1}^{L} W_h^2 \gamma_{xh} \sigma_{xh}}{\sqrt{\sum_{h=1}^{L} W_h^2 \gamma_{xh} \sigma_{xh}^2 / \theta_{xh}} \sqrt{\sum_{h=1}^{L} W_h^2 \gamma_{xh} \sigma_{xh}^2 / \theta_{xh}}}, \]  \hspace{1cm} \text{(II.49)}

Kahlil et al.(2018) showed that the proposed mean estimator \(\hat{\mu}_{st GE}\) is more efficient than the ordinary mean estimator and the ratio estimator when measurement errors are both present and absent, particularly when the study variable and the auxiliary variable are highly correlated.

Researchers have studied mean estimation under stratified random sampling and non-response; and under stratified random sampling, measurement errors, and RRT models. But not many researchers have explored the performance of mean estimators for a sensitive variable under both non-response and measurement errors using stratified random sampling. Using stratified random sampling for estimating the population mean of a sensitive variable in the presence of measurement errors and non-response simultaneously will be our third major objective for this dissertation. The work will be introduced in Chapter V.
CHAPTER III
MEAN ESTIMATION UNDER ORRT MODELS IN THE PRESENCE OF MEASUREMENT ERRORS

As mentioned in Chapter II.2, we will re-examine Khalil et al. (2018) [34] mean estimation of a sensitive variable in the presence of measurement errors but using ORRT models in this Chapter[36]. A RRT model could have different scrambling methods such as simple additive scrambling or multiplicative scrambling. In this Chapter, we also consider a broader class of scrambling methods for RRT models. In Sections III.1 and III.2, a general scrambling RRT model as well as its ORRT version will be discussed; in Section III.3, some existing mean estimators under ORRT model in the presence of measurement errors will be presented; a generalized mean estimator will be introduced in Section III.4; Section III.5 will present the simulation results; and Section III.6 will provide concluding remarks of this Chapter.

III.1. A General Scrambling Model

Let us introduce the notations again. Let Y be the sensitive study variable with unknown mean $\mu_{y}$ and unknown variance $\sigma_{y}^{2}$, and X be a non-sensitive auxiliary variable with known mean $\mu_{x}$ and known variance $\sigma_{x}^{2}$. Suppose X has a strong positive correlation with Y. Let T and S be two scrambling variables with known variances $\sigma_{T}^{2}$ and $\sigma_{S}^{2}$, respectively. Usually we choose T with a mean ($\mu_{T}$) of 1 and S with a mean ($\mu_{S}$) of 0. T, S, X and Y are mutually independent. Let W be the probability that the respondent finds the question sensitive. The respondent is asked to report a scrambled response for study variable (Y) if he/she considers the question sensitive, and a correct response otherwise. One could add noises to the study variable Y differently.
The most commonly used RRT model for quantitative response is the additive model given by Pollock and Bek (1976) [53] where the reported response is

\[ Z = Y + S. \]  (III.1)

Eichhorn and Hayre (1983)[10] proposed a multiplicative model given by

\[ Z = YS. \]  (III.2)

Diana and Perri (2011) [08] introduced a more general linear combination model given by

\[ Z = TY + S. \]  (III.3)

As mentioned previously, it is common to assume that \( E(T) = \mu_T = 1 \) and \( E(S) = \mu_s = 0 \) in model (III.3). One can easily note that models (III.1) and (III.2) are special cases of (III.3) if we assume \( \sigma_s^2 = 0 \) and \( \sigma_T^2 = 0 \), respectively.

The multiplicative scrambling compromises respondent anonymity and it is not very efficient. Hence, in this study only the other two models will be considered. It is easy to verify that in Pollock and Bek (1976) [53] additive model (III.1), \( E(Z) = E(Y) = \mu_y \) and

\[ Var(Z) = \sigma_z^2 = \sigma_y^2 + \sigma_s^2; \]  (III.4)

and for Diana and Perri (2011) [08] general model (III.3), \( E(Z) = E(Y) = \mu_y \) and

\[ Var(Z) = \sigma_z^2 = \sigma_T^2(\mu_y^2 + \sigma_y^2) + \sigma_y^2 + \sigma_s^2. \]  (III.5)

The comparison of variances in (III.1) and (III.3) indicates that the additive model is more efficient than the general model. However, efficiency is not the only
criterion that we use to evaluate RRT models. The primary objective of a RRT model is to protect respondents’ privacy. Privacy level could be another consideration to evaluate RRT models.

Using the privacy protection measure $\Delta = E(Z - Y)^2$ proposed by Yan et al. (2008)[87], we can easily calculate the privacy level of the Pollock & Bek (1976) [53] model in (III.1) and the Diana & Perri (2011) [08] model in (III.3). These are given respectively by

$$\Delta_{PB} = \sigma_s^2$$  \hspace{1cm} (III.6)

and

$$\Delta_{DP} = \sigma_T^2(\mu_y^2 + \sigma_y^2) + \sigma_s^2.$$

(III.7)

One can easily notice by comparing (III.6) with (III.7) that the Diana and Perri (2011) model offers greater privacy.

Efficiency and privacy are two important considerations we use to compare RRT models. If efficiency is same, a model with higher privacy is preferred; and if privacy is same, we choose a model with better efficiency. However, neither efficiency nor privacy is need to be kept fixed here. Instead of holding one of the measures constant, Gupta et al.(2018)[24] proposed a unified measure of model quality given by

$$\delta = \frac{Var(\hat{\mu})}{PL},$$

(III.8)

where $\hat{\mu}$ is the mean estimator and $PL$ is the privacy level for the model as given by Yan et al. (2009). In (III.8), $Var(\hat{\mu})$ can be replaced by $MSE(\hat{\mu})$ in case of biased estimators. The goal of this measure is to achieve a right trade-off between efficiency and privacy protection.
One may note that the model with smaller $\delta$ value is preferred in terms of either a larger privacy level or smaller value of $Var(\hat{\mu})$. It may be observed that

$$\delta_{PB} = 1 + \frac{\sigma_T^2}{\sigma_S^2} > 1 + \frac{\sigma_T^2}{\sigma_S^2} + \frac{\sigma_T^2}{\sigma_S^2} + \frac{\sigma_T^2}{\sigma_S^2} = \delta_{DP}. \quad (\text{III.9})$$

Hence, while working with the general RRT model will put a burden on the model efficiency, it is better in terms of the unified measure of both efficiency and privacy. Therefore, Diana and Perri (2011) model will be used in the current study, but with a reasonably small value of $\sigma_T^2$.

### III.2. ORRT Version of the General Scrambling Model

Since ORRT model is more efficient, we add optionally to the Diana and Perri (2011) model. In the ORRT version, the respondent may answer in the following two ways depending on whether the respondent considers the question sensitive or not:

$$Z = \begin{cases} 
Y & \text{with probability } 1-W \\
TY + S & \text{with probability } W,
\end{cases} \quad (\text{III.10})$$

where it is assumed that $\mu_T = E(T) = 1$ and $\mu_s = E(S) = 0$.

The mean and variance of scrambled response ($Z$) are respectively given by:


$$= E(Y) - E(Y)W + E(T)E(Y)W + E(S)W \quad (\text{III.11})$$

$$= E(Y) - E(Y)W + (1)E(Y)W + (0)W$$

$$= E(Y)$$
and

\[ \text{Var}(Z) = E(Z^2) - E^2(Z) \]

\[ = E(Y^2)(1 - W) + E[(TY + S)^2]W - E^2(Z) \]

\[ = E(Y^2) - E(Y^2)W + E(T^2Y^2)W + 2E(TYS)W + E(S^2)W - E^2(Z) \]

\[ = \text{Var}(Y) + \text{Var}(S)W - E(Y^2)W + E(T^2)E(Y^2)W \]

\[ = \text{Var}(Y) + \text{Var}(S)W - E(Y^2)W + [\text{Var}(T) + 1]E(Y^2)W \]

\[ = \text{Var}(Y) + \text{Var}(S)W + \text{Var}(T)E(Y^2)W \]

\[ = \text{Var}(Y) + \text{Var}(S)W + \text{Var}(T)[\text{Var}(Y) + E^2(Y)]W \]

\[ = \sigma_y^2 + \sigma_s^2W + \sigma_T^2(\sigma_y^2 + \mu_y^2)W. \]

(III.12)

Note that \( \text{Var}(Z) \) increases with \( W \), and hence there is gain in efficiency compared to the non-optional model where \( W=1 \).

Note that under the ORRT model, the correlation coefficient between \( Z \) and \( X \) is given by

\[ \rho_{zx} = \frac{\sigma_{zx}}{\sqrt{\sigma_x^2 \sqrt{\sigma_z^2}}} \]

\[ = \frac{\sigma_{zx}}{\sqrt{\sigma_x^2 \sqrt{\sigma_y^2} + \sigma_s^2W + \sigma_T^2(\sigma_y^2 + \mu_y^2)W}} \]

(III.13)

\[ = \frac{\rho_{yx}}{\sqrt{1 + \frac{\sigma_s^2W}{\sigma_y^2} + \frac{\sigma_T^2(\sigma_y^2 + \mu_y^2)W}{\sigma_y^2}}}. \]

This will be utilized later.
III.3. Some Existing Mean Estimators under Measurement Errors

Let a simple random sample of size $n$ be drawn without replacement from a finite population $U = (U_1, U_2, ..., U_N)$. Let $(x_i, y_i, z_i)$ be the observed values (factoring in measurement errors) and $(X_i, Y_i, Z_i)$ be true values for the auxiliary variable $X$, the study variable $Y$ and the scrambled response variable $Z$ respectively associated with the $i^{th}$ $(i=1, 2, ..., n)$ sample unit. The respective measurement errors associated with the scrambled response variable $(Z)$ and the auxiliary variable $(X)$ are given by

$$P_i = z_i - Z_i \quad (III.14)$$

and

$$V_i = x_i - X_i. \quad (III.15)$$

These measurement errors are assumed to be random and uncorrelated with mean zero and variance $\sigma_p^2$ and $\sigma_v^2$, respectively. Some other necessary notations are given below. Let

$$\Omega_z = \sum_{i=1}^{n} (z_i - \mu_y), \quad (III.16)$$

$$\Omega_x = \sum_{i=1}^{n} (x_i - \mu_x), \quad (III.17)$$

$$\Omega_p = \sum_{i=1}^{n} P_i, \quad (III.18)$$

and

$$\Omega_v = \sum_{i=1}^{n} V_i \quad (III.19)$$
Let \( e_0^* = \frac{1}{n\mu_y} (\Omega_z + \Omega_p) \) and \( e_1^* = \frac{1}{n\mu_x} (\Omega_x + \Omega_v) \). In other words, \( \bar{z}^* = (1 + e_0^*)\mu_y \) and \( \bar{x}^* = (1 + e_1^*)\mu_x \). Under the assumption of bivariate normality (Sukhatme et al.1970)[78]:

\[
E(e_0^*) = 0, \quad (\text{III.20})
\]

\[
E(e_1^*) = 0, \quad (\text{III.21})
\]

\[
E(e_0^{*2}) = \frac{1}{\mu_y^2} \theta (\sigma_z^2 + \sigma_p^2), \quad (\text{III.22})
\]

\[
E(e_1^{*2}) = \frac{1}{\mu_x^2} \theta (\sigma_x^2 + \sigma_v^2) \quad (\text{III.23})
\]

and

\[
E(e_0^*e_1^*) = \theta \rho_{zx} \frac{\sigma_x \sigma_z}{\mu_y \mu_x}, \quad (\text{III.24})
\]

where \( \theta = (N - n)/Nn \).

Some existing mean estimators in the presence of measurement errors using ORRT model are given below.

- The ordinary mean estimator is given by

\[
\hat{\mu}_{yw} = \frac{\sum_{i=1}^n z_i}{n} = \bar{z}^*. \quad (\text{III.25})
\]

It can be written as

\[
\hat{\mu}_{yw} = (1 + e_0^*)\mu_y. \quad (\text{III.26})
\]
The difference between the ordinary mean estimator and the true mean can be written as

\[ \hat{\mu}_{yw} - \mu_y = e_0^* \mu_y. \]  

\[ \text{III.27} \]

Taking square and then expected value on both side of (III.27), the MSE of \( \hat{\mu}_{yw} \) is given by

\[ \text{MSE}^*(\hat{\mu}_{yw}) = \theta (\sigma_z^2 + \sigma_p^2) \]

\[ = \theta (\sigma_y^2 + \sigma_s^2 W + \sigma_T^2 (\mu_y^2 + \mu_y^2) W + \sigma_p^2). \]

\[ \text{III.28} \]

• A ratio estimator corresponding to the one in Gupta et al. (2014) is given by

\[ \hat{\mu}_{rw} = \frac{\bar{z}^*}{\bar{x}^*} \mu_x = \hat{R}_{rw}^* \mu_x. \]

\[ \text{III.29} \]

It can be written as

\[ \hat{\mu}_{rw} = \frac{(1 + e_0^*) \mu_y}{(1 + e_1^*) \mu_x} \mu_x \]

\[ = \mu_y (1 + e_0^*) (1 + e_1^*)^{-1} \]

\[ = \mu_y (1 - e_1^* + e_1^{*2} - e_1^{*3} + ...) \]

\[ = \mu_y (1 - e_1^* + e_1^{*2} + e_0^* - e_0^* e_1^* + ...). \]

\[ \text{III.30} \]

Using second order approximation, the difference between the ratio estimator and the true mean can be written as

\[ \hat{\mu}_{yw} - \mu_y = \mu_y (-e_1^* + e_1^{*2} + e_0^* - e_0^* e_1^*). \]

\[ \text{III.31} \]
Taking square and then expected value on both side of (III.31), the MSE of $\hat{\mu}_{rw}$ is given by

$$MSE^*(\hat{\mu}_{rw}) = \theta(\sigma^2_z + R^2_w \sigma^2_z - 2R_w \rho_{zx} \sigma_x \sigma_z) + \theta(\sigma^2_p + R^2_w \sigma^2_v),$$

(III.32)

where $\sigma^2_z = \sigma^2_y + \sigma^2_s W + \sigma^2_T (\sigma^2_y + \mu^2_y) W$ and $R_w = \mu_y / \mu_z$. 

• A regression estimator proposed by Gupta et al. (2014) is given by

$$\hat{\mu}_{reg,w} = \bar{z}^* + \hat{\beta}_{zx}(\mu_x - \bar{x}^*),$$

(III.33)

where $\hat{\beta}_{zx} = \sigma_{zx} \sigma^2_x = \rho_{zx} \sigma_x \sigma_z$.

It can be written as

$$\hat{\mu}_{reg,w} = (1 + e^*_0)\mu_y + \hat{\beta}_{zx}(\mu_x - (1 + e^*_1)\mu_x)$$

$$= \mu_y(1 + e^*_0) - \hat{\beta}_{zx}(e^*_1 \mu_x).$$

(III.34)

The difference between the regression estimator and the true mean can be written as

$$\hat{\mu}_{reg,w} - \mu_y = e^*_0 \mu_y - \hat{\beta}_{zx}(e^*_1 \mu_x).$$

(III.35)

Taking square and then expected value on both side of (III.35), the MSE of $\hat{\mu}_{reg,w}$, up to second order approximation, is given by

$$MSE^*(\hat{\mu}_{reg,w}) = \theta \sigma^2_z (1 - \rho^2_{zx}) + \theta(\sigma^2_p + \hat{\beta}^2_{zx} \sigma^2_v).$$

(III.36)

The MSEs of the above mean estimators without measurement errors may be obtained by letting $\sigma^2_p = \sigma^2_v = 0$ in (III.28), (III.32) and (III.36).

With this background, we use the generalized mean estimator presented in Khalil et al. (2018) [34]. This estimator includes a wide variety of mean estimators as special cases. It is given below:

\[ \hat{\mu}_{pw} = (\bar{z}^* + k(\mu_x - \bar{x}^*))\left(\frac{D}{d}\right)^{\nu}, \]  

(III.37)

where \( \bar{d} = \phi(\alpha x + \beta) + (1 - \phi)(\alpha \mu_x + \beta) \), \( \bar{D} = \alpha \mu_x + \beta \), \( k \) and \( \nu \) are suitable constants. \( \phi \) is assumed to be an unknown constant whose value is to be determined from optimally considerations. \( \alpha \) (\( \alpha \neq 0 \)) and \( \beta \) are assumed to be some known parameters of the auxiliary variable \( X \), such as coefficient of variation (\( C_x \)), kurtosis, and correlation coefficient (\( \rho_{zx} \)) etc. Please note here that with different values of \( \alpha \) and \( \beta \), we can obtain various estimators. Also, with \( \nu = 1 \) we get various ratio estimators and with \( \nu = -1 \) we get various product estimators.

III.4.1 Bias and MSE of the Generalized Mean Estimator

The generalized mean estimator will be studied under both ORRT model and measurement errors. According to the notations in Chapter III.3, this estimator can be written as

\[ \hat{\mu}_{pw} = ((1 + e_{1}^*)\mu_y + k(\mu_x - (1 + e_{1}^*)\mu_x))(\frac{\alpha \mu_x + \beta}{\phi(\alpha(1 + e_{1}^*)\mu_x + \beta) + (1 - \phi)(\alpha \mu_x + \beta)})^{\nu}. \]  

(III.38)
Using Taylor’s approximation and retaining terms of order up to 2, the difference between the generalized mean estimator and the true mean can be written as

\[
\hat{\mu}_{pw} - \mu_y = ((1 + e_0^*)\mu_y + k(\mu_x - (1 + e_1^*)\mu_x))(\alpha\mu_x + \beta) - \mu_y
\]

\[
= (\mu_y + e_0^*\mu_y - ke_1^*\mu_x)(\alpha\mu_x + \beta) - \mu_y
\]

\[
\approx \mu_y + (e_0^* - 0)\mu_y + (e_1^* - 0)[(-k\bar{x}) + \mu_yv\left(\frac{-\alpha\phi\mu_x}{\alpha\mu_x + \beta}\right) + \frac{1}{2!}[(e_0^* - 0)^2(0) + 2(e_0^* - 0)

\]

\[
(v - 1)\mu_yv\left(\frac{-\alpha\phi\mu_x}{\alpha\mu_x + \beta}\right) + \frac{1}{2!}2\left(\frac{-\alpha\phi\mu_x}{\alpha\mu_x + \beta}\right)
\]

\[
\approx \mu_y + (e_0^* - 0)\mu_y + (e_1^* - 0)[(-k\bar{x}) + \mu_yv\left(\frac{-\alpha\phi\mu_x}{\alpha\mu_x + \beta}\right) + \frac{1}{2!}[(e_0^* - 0)^2(0) + 2(e_0^* - 0)

\]

\[
2\left(\frac{-\alpha\phi\mu_x}{\alpha\mu_x + \beta}\right)
\]

\[
\Rightarrow (III.39)
\]

Taking expectation on both side of (III.39), the bias of the generalized mean estimator \(\hat{\mu}_{pw}\), correct to the second order or approximation, is given by

\[
\text{Bias}^*(\hat{\mu}_{pw}) \approx \frac{\theta}{\mu_y} \left(\frac{v(v + 1)}{2}\Phi^2R_{pw}^2\sigma_x^2 - \mu_y\sigma_{\mu_x}^2 + \nu\sigma_{\mu_y}^2 + \nu\mu_{\sigma_x}^2 + \mu_{\sigma_y}^2\right)
\]

\[
+ \frac{\theta}{\mu_y} \left(\frac{v(v + 1)}{2}\Phi^2R_{pw}^2\sigma_v^2 + \nu\sigma_{\mu_y}^2\right)
\]

\[
(III.40)
\]

where \(R_{pw} = \frac{\alpha\mu_y}{\alpha\mu_x + \beta}\). The bias of \(\hat{\mu}_{pw}\) without measurement errors may be obtained by setting \(\sigma_v^2 = 0\) in above equation.

To determine the expression for MSE of the generalized mean estimator in (III.37), we take square of (III.39) on both sides and retaining terms of order up to 2
to get

\[(\hat{\mu}_{pw} - \mu_y)^2 = \varepsilon_0^2 \mu^2_y + k^2 \mu^2_x \varepsilon_1^2 + \left( \frac{\alpha \phi v}{\alpha \mu_x + \beta \mu_x \mu_y} \right)^2 - 2 \rho_{yx} \varepsilon_1^2 k \mu_x \mu_y - 2 \rho_{yx} \varepsilon_1^2 \frac{\alpha \phi v}{\alpha \mu_x + \beta \mu_x \mu_y}.\]  

(III.41)

Taking the expected value on both side of (III.41), the expression for MSE of \(\hat{\mu}_{pw}\), correct to the first order approximation is given by

\[\text{MSE}^*(\hat{\mu}_{pw}) \approx \theta \left( \sigma^2_z + v^2 \phi^2 R_{pw}^2 \sigma^2_x + k^2 \sigma^2_x - 2 v \phi R_{pw} \rho_{zx} \sigma_x \sigma_y - 2 k \rho_{zx} \sigma_x \sigma_y + 2 v \phi k R_{pw} \sigma^2_y \right) + \theta \left( \sigma^2_p + v^2 \phi^2 R_{pw}^2 \sigma^2_v + k^2 \sigma^2_v + 2 v \phi k R_{pw} \sigma^2_v \right),\]  

(III.42)

where \(\theta = \frac{N-n}{N n}\) and \(R_{pw} = \frac{\alpha \mu_y}{\alpha \mu_x + \beta}\).

The optimum value of \(\phi\) by taking the first derivative which gives the minimum MSE, is given by

\[\phi_{\text{opt}} = \frac{\rho_{zx} \sigma_x \sigma_y - k (\sigma^2_x + \sigma^2_v)}{v R_{pw} (\sigma^2_x + \sigma^2_v)}.\]  

(III.43)

By substituting (III.43) in (III.42), the minimum value of \(\text{MSE}^*(\hat{\mu}_{pw})\) is given by

\[\text{MSE}_{\text{min}}^*(\hat{\mu}_{pw}) \approx \theta (\sigma^2_z + \sigma^2_p) \frac{\rho_{zx} \sigma_z \sigma_x}{\sigma^2_x + \sigma^2_v}.\]  

(III.44)

The expression for the minimized MSE of proposed estimator without measurement errors may be obtained by putting \(\sigma^2_u = \sigma^2_v = 0\) in (III.44), which gives

\[\text{MSE}_{\text{min}}(\hat{\mu}_{pw}) = \theta \sigma^2_z (1 - \rho^2_{zx}).\]  

(III.45)

III.4.2 Efficiency Comparisons

The \(\text{MSE}_{\text{min}}(\hat{\mu}_{pw})\) in (III.44) is same as that of the approximate MSE of the usual linear regression estimator \(\text{MSE}(\hat{\mu}_{\text{reg,w}})\) (III.36). Comparing the minimum
MSE(\(\hat{\mu}_{pw}\)) (III.44) with measurement errors to MSEs of existing estimators MSE(\(\hat{\mu}_{yw}\)) (III.28) and MSE(\(\hat{\mu}_{rw}\)) (III.32), it is easy to verify that

\[
MSE_{\min}^*(\hat{\mu}_{pw}) < MSE^*(\hat{\mu}_{yw}) \text{ if } \rho_{zx}^2 \frac{\sigma_x^2 \sigma_z^2}{\sigma_x^2 + \sigma_z^2} > 0
\]  

(III.46)

and

\[
MSE_{\min}^*(\hat{\mu}_{pw}) < MSE^*(\hat{\mu}_{rw}) \text{ if } (R_{pw} \sqrt{\frac{\sigma_x^2}{\sigma_x^2 + \sigma_v^2}} - \frac{\rho_{zx} \sigma_z \sigma_x}{\sqrt{\sigma_x^2 + \sigma_v^2}})^2 > 0.
\]  

(III.47)

These two conditions always hold true.

Comparing MSE(\(\hat{\mu}_{yw}\)) (III.28) and MSE(\(\hat{\mu}_{rw}\)) (III.32), it is easy to verify that

\[
MSE^*(\hat{\mu}_{rw}) < MSE^*(\hat{\mu}_{yw}) \text{ if } \theta[R_w^2(\sigma_x^2 + \sigma_v^2) - 2R_w \rho_{zx} \sigma_x \sigma_z] < 0.
\]  

(III.48)

This condition holds true only when \(R_w^2(\sigma_x^2 + \sigma_v^2)\) is small or/and \(2R_w \rho_{zx} \sigma_x \sigma_z\) is large. The population parameters are fixed for a finite population, and the correlation between X and Z (\(\rho_{zx}\)) is positively associated with the correlation between X and Y (\(\rho_{yx}\)). In other words, the ratio estimator \(\hat{\mu}_{rw}\) is more efficient than the ordinary mean estimator \(\hat{\mu}_{yw}\) only when the measurement errors of X (\(\sigma_v^2\)) is small and the correlation between X and Y (\(\rho_{yx}\)) is high. This conditional superiority of the ratio estimator is reasonable because measurement errors on both X and Y will bring more burden on efficiency than the measurement errors on Y alone. However, the generalized mean estimator has no such restrictions. It is always more efficient than the ordinary mean estimator no matter how large the measurement errors are since the generalized estimator uses variety of other information also.
III.5. Simulation Study

In this section, we examine the performance of the generalized mean estimator with the ordinary mean estimator and the ratio estimator, by way of a simulation study. In the generalized mean estimator, we choose \( v \) and \( k \) to be 1, and \( \phi \) to be its optimum value. As for \( \alpha \) and \( \beta \), we have used various parameters associated with the auxiliary variable such as the coefficient of variation \( (C_x) \) and kurtosis, but these choices do not impact the results in any meaningful way. As we can see in (III.44) above, minimized MSE is independent of \( \alpha \) and \( \beta \), and empirical MSEs also are almost the same for all choices of \( \alpha \) and \( \beta \). We will show different cases where \( \alpha \) and \( \beta \) take different values.

We consider a finite population of size 5000 generated from bivariate normal distribution with means and covariances of \((Y, X)\) as given below.

\[
\text{Population} \quad \mu = \begin{bmatrix} 10 \\ 6 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 16 & 9.0510 \\ 9.0510 & 8 \end{bmatrix}, \quad \rho_{yx} = 0.8
\]

To explain the simulation process further, we started with a sample of size 5000 from a normal population with parameters:

\[
\mu_x = 6, \sigma_x^2 = 8, \mu_y = 10, \sigma_y^2 = 16, \rho_{yx} = 0.8 \quad (A)
\]

However, the real parameters of the set of 5000 data points we generated using R are very close to the parameter values in (A) but not exactly same. For the simulation study, we used parameter values in (B) and not those in (A).

\[
\mu_x = 6.0228, \sigma_x^2 = 8.1830, \mu_y = 9.9864, \sigma_y^2 = 16.1215, \rho_{yx} = 0.8024 \quad (B)
\]
The scrambling variable $S$ is taken to be a normal variate with mean equal to zero and vary variances $(0.2\sigma_x^2, 0.5\sigma_x^2$ and $1\sigma_x^2)$. And $T$ is also taken to be a normal variate but with mean equal to one and varying variances $(0, 0.5, 1)$.

The observed values of $Z$ and $X$ are given by: $z = Z + p$ and $x = X + v$, where $p$ and $v$ are represent measurement errors. The measurement errors are taken to be a normal distributions with mean equal to zero and varying variances $(0, 5, 10)$. The observed response $z$ is given by $z=TY+S+P$ with probability $W$, and by $z = Y+P$ with probability $1-W$.

We consider samples of size $n = 500$ using SRSWOR (simple random sampling without replacement). Coding for the simulations was done in R and results are averaged over 5,000 iterations. The empirical MSE of the estimator $\hat{\mu}_w$ is computed by

$$MSE^*(\hat{\mu}_w) = \frac{1}{5000} \sum_{i=1}^{5000} (\hat{\mu}_w - \mu_y)^2,$$

where $\hat{\mu}_w = \hat{\mu}_{yw}, \hat{\mu}_{rw}, \hat{\mu}_{pw}$. Here, $\mu_y$ is the population mean of the sensitive study variable. The percent relative efficiencies (PREs) of the estimators ($\hat{\mu}_w$) with respect to mean estimator ($\hat{\mu}_{yw}$) is defined as

$$PRE = \frac{MSE^*(\hat{\mu}_{yw})}{MSE^*(\hat{\mu}_w)} * 100.$$  

We will also use the unified measure $\delta$ of the efficiency and the privacy as defined in Gupta et al. (2018)[24]. It is given by

$$\delta = \frac{MSE^*(\hat{\mu}_w)}{\Delta_{DP}}.$$  

In (III.51), MSE is used in place of Var(.) to account for biased estimators.
Table III.1. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators when $\sigma_v^2 = \sigma_p^2 = 1$ and $\sigma_s^2 = 0.2*\sigma_x^2$.

<table>
<thead>
<tr>
<th>Est. W</th>
<th>$\sigma_T^2$</th>
<th>MSE</th>
<th>PRE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Without ME</td>
<td>With ME</td>
<td>Without ME</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.0305</td>
<td>0.0323</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0311</td>
<td>0.0335</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.0820</td>
<td>0.0838</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.0822</td>
<td>0.0884</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.1335</td>
<td>0.1352</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1330</td>
<td>0.1340</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.0313</td>
<td>0.0331</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.0309</td>
<td>0.0334</td>
<td>100.0000</td>
</tr>
<tr>
<td>$\hat{\mu}_{yw}$</td>
<td>0.5</td>
<td>0.1137</td>
<td>0.1155</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1120</td>
<td>0.1130</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.1961</td>
<td>0.1979</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1930</td>
<td>0.1933</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.0319</td>
<td>0.0337</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.0312</td>
<td>0.0339</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1349</td>
<td>0.1367</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1334</td>
<td>0.1351</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.2379</td>
<td>0.2397</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.2350</td>
<td>0.2368</td>
<td>100.0000</td>
</tr>
<tr>
<td>$\hat{\mu}_{rw}$</td>
<td>0.5</td>
<td>0.0160</td>
<td>0.0176</td>
<td>190.6250</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.0165</td>
<td>0.0231</td>
<td>188.4848</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0675</td>
<td>0.0742</td>
<td>121.4815</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.0678</td>
<td>0.0794</td>
<td>121.2389</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1189</td>
<td>0.1257</td>
<td>112.2792</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1286</td>
<td>0.1249</td>
<td>103.4215</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.0168</td>
<td>0.0236</td>
<td>186.3095</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.0174</td>
<td>0.0241</td>
<td>177.5862</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0992</td>
<td>0.1060</td>
<td>114.6169</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1006</td>
<td>0.1055</td>
<td>111.3320</td>
</tr>
<tr>
<td>µ</td>
<td>pw</td>
<td>0.1816</td>
<td>0.1883</td>
<td>107.9846</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>0</td>
<td>0.0174</td>
<td>0.0242</td>
<td>183.3333</td>
<td>139.2562</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1204</td>
<td>0.1271</td>
<td>112.0432</td>
<td>107.5531</td>
</tr>
<tr>
<td>1</td>
<td>0.2234</td>
<td>0.2301</td>
<td>106.4906</td>
<td>104.1721</td>
</tr>
<tr>
<td>0</td>
<td>0.0118</td>
<td>0.0156</td>
<td>258.4746</td>
<td>207.0513</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0633</td>
<td>0.0671</td>
<td>129.5419</td>
<td>124.8882</td>
</tr>
<tr>
<td>1</td>
<td>0.1148</td>
<td>0.1186</td>
<td>116.2892</td>
<td>113.9966</td>
</tr>
<tr>
<td>0</td>
<td>0.0127</td>
<td>0.0165</td>
<td>246.4567</td>
<td>200.6061</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0951</td>
<td>0.0989</td>
<td>119.5584</td>
<td>116.7846</td>
</tr>
<tr>
<td>1</td>
<td>0.1774</td>
<td>0.1813</td>
<td>110.5411</td>
<td>109.1561</td>
</tr>
<tr>
<td>0</td>
<td>0.0132</td>
<td>0.0171</td>
<td>241.6667</td>
<td>197.0760</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1163</td>
<td>0.1201</td>
<td>115.9931</td>
<td>113.8218</td>
</tr>
<tr>
<td>1</td>
<td>0.2192</td>
<td>0.2230</td>
<td>108.5310</td>
<td>107.4888</td>
</tr>
</tbody>
</table>

\[ \hat{\mu}_{pw} \]
Table III.2. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators when $\sigma^2_v = \sigma^2_p = 1$ and $\sigma^2_s = 0.5^{*}\sigma^2_x$.

<table>
<thead>
<tr>
<th>Est. $W$</th>
<th>$\sigma^2_T$</th>
<th>MSE Without ME</th>
<th>MSE With ME</th>
<th>PRE Without ME</th>
<th>PRE With ME</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0327</td>
<td>0.0335</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0084</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.0842</td>
<td>0.0895</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0014</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1357</td>
<td>0.1354</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0012</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.0348</td>
<td>0.0344</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0090</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.1173</td>
<td>0.1157</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0019</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1997</td>
<td>0.1967</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0017</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.0363</td>
<td>0.0355</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0093</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.1393</td>
<td>0.1387</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0023</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.2423</td>
<td>0.2405</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0021</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.0181</td>
<td>0.0189</td>
<td>180.6630</td>
<td>138.1526</td>
<td>0.0061</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.0697</td>
<td>0.0702</td>
<td>120.8034</td>
<td>112.5654</td>
<td>0.0012</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1212</td>
<td>0.1211</td>
<td>111.9637</td>
<td>107.5059</td>
<td>0.0011</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.0203</td>
<td>0.0212</td>
<td>171.4286</td>
<td>135.0554</td>
<td>0.0066</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.1028</td>
<td>0.1044</td>
<td>114.1051</td>
<td>108.7671</td>
<td>0.0018</td>
<td></td>
</tr>
<tr>
<td>$\hat{\mu}_{pw}$</td>
<td>1</td>
<td>0.1851</td>
<td>0.1919</td>
<td>107.8876</td>
<td>104.9505</td>
<td>0.0016</td>
</tr>
<tr>
<td>----------------</td>
<td>---</td>
<td>---------</td>
<td>---------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.1864</td>
<td>0.1900</td>
<td>105.5258</td>
<td>103.7368</td>
<td>0.0016</td>
</tr>
<tr>
<td></td>
<td>0.1248</td>
<td>0.1316</td>
<td>111.6186</td>
<td>107.2188</td>
<td>0.0021</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2278</td>
<td>0.2345</td>
<td>106.3652</td>
<td>104.0938</td>
<td>0.0020</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1170</td>
<td>0.1208</td>
<td>115.9829</td>
<td>113.8245</td>
<td>0.0010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1076</td>
<td>0.1104</td>
<td>110.3315</td>
<td>108.9827</td>
<td>0.0016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0986</td>
<td>0.1014</td>
<td>116.6331</td>
<td>115.0888</td>
<td>0.0017</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.1810</td>
<td>0.1848</td>
<td>110.8264</td>
<td>108.9827</td>
<td>0.0016</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.1808</td>
<td>0.1821</td>
<td>108.7942</td>
<td>108.2372</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.0176</td>
<td>0.0214</td>
<td>206.2500</td>
<td>178.0374</td>
<td>0.0052</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1207</td>
<td>0.1245</td>
<td>115.4101</td>
<td>113.3333</td>
<td>0.0020</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2236</td>
<td>0.2275</td>
<td>108.3631</td>
<td>107.2967</td>
<td>0.0019</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.2247</td>
<td>0.2276</td>
<td>107.0316</td>
<td>106.3708</td>
<td>0.0019</td>
<td></td>
</tr>
</tbody>
</table>

54
Table III.3. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators when $\sigma^2_v = \sigma^2_p = 1$ and $\sigma^2_s = 1*\sigma^2_x$.

<table>
<thead>
<tr>
<th>Est. $\hat{\mu}_{yw}$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_T$</td>
<td>MSE</td>
</tr>
<tr>
<td></td>
<td>Without ME</td>
</tr>
<tr>
<td>0</td>
<td>0.0363</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0879</td>
</tr>
<tr>
<td>1</td>
<td>0.1393</td>
</tr>
<tr>
<td>0</td>
<td>0.0407</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1232</td>
</tr>
<tr>
<td>1</td>
<td>0.2055</td>
</tr>
<tr>
<td>0</td>
<td>0.0436</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1467</td>
</tr>
<tr>
<td>1</td>
<td>0.2497</td>
</tr>
<tr>
<td>0</td>
<td>0.0218</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0733</td>
</tr>
<tr>
<td>1</td>
<td>0.1248</td>
</tr>
<tr>
<td>0</td>
<td>0.0261</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1087</td>
</tr>
</tbody>
</table>

55
<table>
<thead>
<tr>
<th>$\bar{\mu}_{p\bar{w}}$</th>
<th>0</th>
<th>0.1910</th>
<th>0.1977</th>
<th>107.5916</th>
<th>104.8558</th>
<th>0.0016</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0.1927</td>
<td>0.1961</td>
<td>105.2413</td>
<td>103.6206</td>
<td>0.0016</td>
</tr>
<tr>
<td></td>
<td>0.0291</td>
<td>0.0358</td>
<td>149.8282</td>
<td>126.8156</td>
<td>0.0044</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0306</td>
<td>0.0371</td>
<td>139.8693</td>
<td>123.4501</td>
<td>0.0046</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1322</td>
<td>0.1389</td>
<td>110.9682</td>
<td>106.9114</td>
<td>0.0021</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1348</td>
<td>0.1406</td>
<td>109.0504</td>
<td>105.6188</td>
<td>0.0022</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2382</td>
<td>0.2419</td>
<td>104.8279</td>
<td>103.9686</td>
<td>0.0020</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2352</td>
<td>0.2435</td>
<td>105.9099</td>
<td>102.9158</td>
<td>0.0020</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0176</td>
<td>0.0214</td>
<td>206.2500</td>
<td>178.0374</td>
<td>0.0026</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0183</td>
<td>0.0224</td>
<td>205.4645</td>
<td>178.5714</td>
<td>0.0028</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0692</td>
<td>0.0730</td>
<td>127.0231</td>
<td>122.7397</td>
<td>0.0011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0686</td>
<td>0.0731</td>
<td>127.5510</td>
<td>124.0766</td>
<td>0.0011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1207</td>
<td>0.1245</td>
<td>115.4101</td>
<td>113.3333</td>
<td>0.0010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1204</td>
<td>0.1238</td>
<td>115.6977</td>
<td>113.4087</td>
<td>0.0010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0220</td>
<td>0.0258</td>
<td>185.0000</td>
<td>164.3411</td>
<td>0.0032</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0228</td>
<td>0.0270</td>
<td>177.1930</td>
<td>160.0000</td>
<td>0.0033</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1045</td>
<td>0.1083</td>
<td>117.8947</td>
<td>115.4201</td>
<td>0.0017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1054</td>
<td>0.1075</td>
<td>115.6546</td>
<td>114.3256</td>
<td>0.0016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1869</td>
<td>0.1907</td>
<td>109.9518</td>
<td>108.7048</td>
<td>0.0016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1870</td>
<td>0.1882</td>
<td>108.4492</td>
<td>107.9702</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0249</td>
<td>0.0287</td>
<td>175.1004</td>
<td>158.1882</td>
<td>0.0036</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0258</td>
<td>0.0295</td>
<td>165.8915</td>
<td>155.2542</td>
<td>0.0037</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1280</td>
<td>0.1318</td>
<td>114.6094</td>
<td>112.6707</td>
<td>0.0020</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1300</td>
<td>0.1330</td>
<td>113.0769</td>
<td>111.6541</td>
<td>0.0020</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2310</td>
<td>0.2348</td>
<td>108.0952</td>
<td>107.1124</td>
<td>0.0019</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2330</td>
<td>0.2358</td>
<td>106.9099</td>
<td>106.2765</td>
<td>0.0019</td>
<td></td>
</tr>
</tbody>
</table>

56
Tables III.1, III.2, and III.3 present the theoretical and empirical MSEs and PREs of the ORRT mean estimators when both the variances of measurement errors on \(X\) and \(Z\) are set equal to 1 and the variance of \(S\) is set equal to \(0.2*\sigma_x^2\), \(0.5*\sigma_x^2\), and \(1*\sigma_x^2\), respectively. Comparing these three tables, the mean estimation is less efficient as the variance of \(S\) increases. For instance, when \(\sigma_T^2=1\), the sensitively level \(W\) is is equal to 0.8, and the measurement errors are present, the MSEs of the generalized mean estimator are 0.1813, 0.1848, 0.1907 respectively corresponding to the Var(\(S\)) is equal to \(0.2*\sigma_x^2\), \(0.5*\sigma_x^2\), and \(1*\sigma_x^2\). These results consistent with the theoretical results. Larger variance of \(S\) introduces more penalty for using RRT models.

From all three tables, one can observe that the MSE of the mean estimators increases as \(W\) increases, both when measurement errors are present, and absent. For example in the Table III.1, the MSE of the generalized mean estimator increased from 0.0671 to 0.1201 as \(W\) increased from 0.5 to 1 when \(\sigma_T^2 = 0.5\) and the measurement errors are present. It indicates the ORRT model gains some efficiency when some the respondents feel the survey question is not sensitive. Also, as the variance of \(T\) increases, the MSE increases while \(\delta\) decreases with a reasonably small value of \(\sigma_T^2\). Again, we can select some values from Table III.1 as an example. When the sensitively level \(W\) is 0.5 and the measurement errors are present, the MSE of the ordinary mean estimator increased from 0.0323 to 0.1352 as the variance of \(T\) increased from 0 to 1, while the \(\delta\) value decreased from 0.0198 to 0.0012. In other words, mean estimators under the simple additive model \((Z=Y+S)\) are more efficient as compared to the general linear combination model \((Z=TY+S)\). However, the general linear combination model is better if both efficiency and privacy are considered simultaneously.
Table III.4. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under the Conditions of $\sigma_v^2 = \sigma_p^2 = 1, 5, 10$ when $W = 0.8$, $\sigma_T^2 = 0.5$ and $\sigma_s^2 = 0.5*\sigma_x^2$.

<table>
<thead>
<tr>
<th>Est.</th>
<th>MSE</th>
<th>PRE</th>
<th>MSE</th>
<th>PRE</th>
<th>MSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>$\hat{\mu}_{yw}$</td>
<td>0.1191</td>
<td>0.1262</td>
<td>0.1351</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>0.1167</td>
<td>0.1241</td>
<td>0.1335</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
</tr>
<tr>
<td>$\hat{\mu}_{rw}$</td>
<td>0.1095</td>
<td>0.1364</td>
<td>0.1699</td>
<td>108.7671</td>
<td>92.5220</td>
<td>79.5174</td>
</tr>
<tr>
<td></td>
<td>0.1092</td>
<td>0.1359</td>
<td>0.1703</td>
<td>106.8681</td>
<td>91.3171</td>
<td>78.3911</td>
</tr>
<tr>
<td>$\hat{\mu}_{pw}$</td>
<td>0.1024</td>
<td>0.1146</td>
<td>0.1267</td>
<td>116.3086</td>
<td>110.1222</td>
<td>106.6298</td>
</tr>
<tr>
<td></td>
<td>0.1014</td>
<td>0.1133</td>
<td>0.1260</td>
<td>115.0888</td>
<td>109.5322</td>
<td>105.9524</td>
</tr>
</tbody>
</table>

Table III.4 presents the theoretical and empirical MSEs and PREs of the ORRT mean estimators under different variances of measurement errors on X and Z when the sensitivity level W is equal to 0.8, variance of T is equal to 0.5 and variance of S is equal to $0.5*\sigma_x^2$. As the variance of measurement errors increase, the MSE of each mean estimator increases, which means larger measurement errors have larger negative impact on mean estimation.

Also, from Tables III.1, III.2, III.3 and III.4, it is more clear that the generalized mean estimator $\hat{\mu}_{pw}$ is more efficient than the other two mean estimators no matter how large the measurement errors are. However, as the measurement errors increase, the ratio estimator $\hat{\mu}_{rw}$ become less efficient than the ordinary mean estimator $\hat{\mu}_{yw}$ because the ordinary mean estimator is not impacted by the measurement error in X. This was not so for the generalized mean estimator because the use of the regression term was able to overcome the measurement error burden due to X. Therefore, the generalized mean estimator may be preferred in mean estimation since it is more efficient without restrictions.
Table III.5 presents the theoretical and empirical MSEs and PREs of the ORRT Generalized Mean Estimator under different $\alpha$ and $\beta$ values when $\sigma_v^2 = \sigma_p^2 = 1$, $W = 0.8$ and $\sigma_s^2 = 0.5\sigma_x^2$.

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\sigma_T^2$</th>
<th>MSE</th>
<th>PRE</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.0162</td>
<td>214.8148</td>
<td>118.9189</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.0986</td>
<td>118.9655</td>
<td>106.6298</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.1810</td>
<td>110.3315</td>
<td>104.0172</td>
</tr>
<tr>
<td>1</td>
<td>Cx</td>
<td>0.5</td>
<td>0.0162</td>
<td>214.8148</td>
<td>118.9189</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.0986</td>
<td>118.9655</td>
<td>106.6298</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.1810</td>
<td>110.3315</td>
<td>104.0172</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0.5</td>
<td>0.0162</td>
<td>214.8148</td>
<td>118.9189</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.0986</td>
<td>118.9655</td>
<td>106.6298</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.1810</td>
<td>110.3315</td>
<td>104.0172</td>
</tr>
<tr>
<td>0.5</td>
<td>Cx</td>
<td>0.5</td>
<td>0.0162</td>
<td>214.8148</td>
<td>118.9189</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5</td>
<td>0.0986</td>
<td>118.9655</td>
<td>106.6298</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>0.1810</td>
<td>110.3315</td>
<td>104.0172</td>
</tr>
</tbody>
</table>

Table III.5 presents the theoretical and empirical MSEs and PREs of the ORRT estimators under different $\alpha$ and $\beta$ values when the sensitivity level $W$ is equal to 0.8.
and variance of T is equal to 0.5. It is clear that different values of \( \alpha \) and \( \beta \) have no impact on the efficiency.

### III.6. Concluding Chapter Remarks

The main contribution in this chapter is the mean estimation of a sensitive variable, in the presence of measurement errors, using ORRT models. While such mean estimation has been attempted before by Khalil et al. (2018) using non-optional RRT models. It has not been done using the more efficient ORRT models. The resultant gain in efficiency using ORRT models is obvious from both the theoretical and the empirical results. The simple additive RRT model is more efficient in terms of PRE. But the general RRT model is better, if we examine the performance of various estimators with respect to the unified measure of efficiency and privacy. It is also clear from the theoretical conditions (III.46) and (III.47) and the simulation results that the generalized mean estimator is more efficient than the ordinary mean estimator and the ratio estimator.

Non-response is another common non-sampling error we have seen in sampling. Will the generalized estimator in Chapter III still be more efficient than the other existing estimators in the presence of non-response? The mean estimation of a sensitive variable under both measurement errors and non-response will be discussed in the next chapter.
CHAPTER IV
MEAN ESTIMATION IN THE SIMULTANEOUS PRESENCE OF
MEASUREMENT ERRORS AND NON-RESPONSE USING ORRT MODELS

We have briefly introduced utilizing non-optional RRT in Hansen and Hurwitz (1946) two-phase sampling in Section II.3. But we aim to work with the more efficient ORRT models. In Section IV.1, a modified version of Hansen and Hurwitz (1946) two-phase sampling using ORRT models will be introduced; some existing mean estimators under the modified two-phase sampling in the presence of measurement errors will be discussed in Section IV.2; Section IV.3 will talk about the generalized mean estimator; Section IV.4 will present the simulation results; Section IV.5 will provide concluding remarks for this Chapter.

IV.1. Modified Hansen and Hurwitz (HH) Two-phase Sampling Technique

As mentioned in Section I.2, Hansen and Hurwitz two-phase sampling uses mail or phone survey at the first attempt and then uses face-to-face interview at the second phase to obtain more information. However, it may cause non-response bias if the variable of interest is sensitive. The respondent may provide untruthful response in the face-to-face interview. In order to encourage the respondents to answer a sensitive survey question truthfully, we give the respondents the opportunity to scramble the response using ORRT in the second phase of HH procedure when there is a face-to-face interview. In this case, we are modifying the HH procedure assuming that in the first phase, respondent group gives direct answer to both X and Y; and then in the second phase, ORRT model is used to get response from the group of non-respondents.
Using the standard terminology as used before, let \( \mu_y = \frac{\sum_{i=1}^{N} y_i}{N} \) and \( \sigma^2_y = \frac{\sum_{i=1}^{N} (y_i - \mu_y)^2}{N-1} \) be the population mean and variance of the study variable \( Y \). Let \( \mu_y(1) = \frac{\sum_{i=1}^{N_1} y_i}{N_1} \) and \( \sigma^2_y(1) = \frac{\sum_{i=1}^{N_1} (y_i - \mu_y(1))^2}{N_1-1} \) be the population mean and variance of respondent group of size \( N_1 \), \( \mu_y(2) = \frac{\sum_{i=1}^{N_2} y_i}{N_2} \) and \( \sigma^2_y(2) = \frac{\sum_{i=1}^{N_2} (y_i - \mu_y(2))^2}{N_2-1} \) be the population mean and variance of non-respondent group of size \( N_2 \). Let \( \mu_x = \frac{\sum_{i=1}^{N} x_i}{N} \) and \( \sigma^2_x = \frac{\sum_{i=1}^{N} (x_i - \mu_x)^2}{N-1} \) be the population mean and variance of the auxiliary variable \( X \). Let \( \mu_x(1) = \frac{\sum_{i=1}^{N_1} x_i}{N_1} \) and \( \sigma^2_x(1) = \frac{\sum_{i=1}^{N_1} (x_i - \mu_x(1))^2}{N_1-1} \) be the population mean and variance of respondent group of size \( N_1 \), \( \mu_x(2) = \frac{\sum_{i=1}^{N_2} x_i}{N_2} \) and \( \sigma^2_x(2) = \frac{\sum_{i=1}^{N_2} (x_i - \mu_x(2))^2}{N_2-1} \) be the population mean and variance of non-respondent group of size \( N_2 \). We assume that only \( n_1 \) units provide response on the first call and remaining \( n_2 = n - n_1 \) units do not respond. Then a subsample of size \( n_s = \frac{n_2}{f} \) \( (f>0) \) is taken. Let \( \rho_{yx} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \) be the correlation coefficient between \( X \) and \( Y \). Similarly let \( \rho_{yx}(1) = \frac{\sigma_{xy}(1)}{\sigma_x \sigma_y} \) be the correlation coefficient for the respondent group, and \( \rho_{yx}(2) = \frac{\sigma_{xy}(2)}{\sigma_x \sigma_y} \) be the correlation coefficient for the non-respondents group.

In Section III.2, we have proved that the general linear combination RRT model is better if both efficiency and privacy are considered together. Therefore, when we apply ORRT in the second phase, the scrambled response is given by

\[
Z = \begin{cases} 
Y & \text{with probability 1-W} \\
TY + S & \text{with probability W,}
\end{cases}
\]  

(IV.1)

where it is assumed that \( \mu_T = \text{E}(T) = 1 \) and \( \mu_s = \text{E}(S) = 0 \).

We can write randomized linear model as \( Z = (YT + S)J + Y(1-J) \), where \( J \sim \text{Bernoulli}(W) \). Therefore, \( \text{E}(J) = W, \text{Var}(J) = W(1-W) \) and \( \text{E}(J^2) = \text{Var}(J) + E^2(J) = W \).
The expectation and variance under randomization mechanism are given by

\[ E_R(Z) = E_R(TYJ + SJ + Y - YJ) \]
\[ = YE_R(TJ) + E_R(SJ) + Y - YE_R(J) \]  
\[ = Y\mu_TW + \mu_sW + Y -YW \]
\[ = (\mu_TW + 1 - W)Y + \mu_sW \]  
\[ \text{(IV.2)} \]

and

\[ V_R(Z) = V_R(TYJ + SJ + Y - YJ) \]
\[ = V_R(TYJ) + V_R(SJ) + V_R(YJ) + 2Cov(TYJ, SJ) - 2Cov(TYJ, YJ) \]
\[ - 2Cov(SJ, YJ) \]
\[ = Y^2[(\sigma^2_T + \mu^2_T)W - \mu^2_TW^2] + [(\sigma^2_s + \mu^2_s)W - \mu^2_sW^2] + Y^2[W(1 - W) + \]
\[ 2Y\mu_TW(1 - W) - 2Y^2[\mu_TW(1 - W)] - 2Y[\mu_sW(1 - W)] \]
\[ = (Y^2\sigma^2_T + \sigma^2_s)W. \]  
\[ \text{(IV.3)} \]

Let \( \hat{y}_i \) be a transformation of the randomized response on the \( i \)th unit whose expectation under the randomization mechanism is the true response \( y_i \). It is given by

\[ \hat{y}_i = \frac{z_i - \mu_sW}{\mu_TW + 1 - W} \]  
\[ \text{(IV.4)} \]

with

\[ E_R(\hat{y}_i) = y_i \]  
\[ \text{(IV.5)} \]
(from IV.2) and
\[
V_R(\hat{y}_i) = \frac{V_R(z_i)}{(\mu_W + 1 - W)^2} = \frac{[\hat{y}_i^2 \sigma_2^2 + \sigma_2^2]W}{(\mu_W + 1 - W)^2} = \tau_i
\]  
(IV.6)

(from IV.3).

With ORRT model added, a modified version of HH estimator is given by
\[
\hat{y} = w_1 \bar{y}_1 + w_2 \hat{y}_2,
\]  
(IV.7)

where \( \hat{y}_2 = \sum_{i=1}^{n_s} (\hat{y}_i) \).

Let \( E_i \) and \( V_i \) be the expectation and variance in the \( i^{th} \) phase (\( i=1,2 \)) under the two-phase sampling. It is easy to verify that
\[
E(\hat{y}) = E_1 E_2 [w_1 \bar{y}_1 + w_2 \hat{y}_2]
= E_1 [w_1 \bar{y}_1 + w_2 E_R(\hat{y}_2)]
= E_1 [w_1 \bar{y}_1 + w_2 \bar{y}_2]
= W_1 \mu_y(1) + W_2 \mu_y(2)
= \mu_y
\]  
(IV.8)

since \( E_R(\hat{y}_2) = \frac{1}{n_s} \sum_{i=1}^{n_s} E_R(\hat{y}_i) = \bar{y}_2 \).
The variance of \( \hat{y} \) can be written as

\[
\text{Var}(\hat{y}) = E[V_2(\hat{y})] + V_1[E_2(\hat{y})]
\]

\[
= E[V_2(w_1\hat{y}_1 + w_2\hat{y}_2)] + V_1[w_1\hat{y}_1 + w_2\hat{y}_2]
\]

\[
= E[0 + V_2(w_2\hat{y}_2)] + V_1[w_1\hat{y}_1 + w_2\hat{y}_2]
\]

\[
= E[V_2(w_2\hat{y}_2)] + V_1(\bar{y})
\]

\[
= E\left[\frac{w_2}{n_s} \sum_{i=1}^{N_2} \frac{(\sigma_y^2 + \sigma_{y(2)}^2) W}{\mu_T W + 1 - W^2}\right] + V(\bar{y})
\]

\[
= \text{Var}(\bar{y}) + \frac{W_2 f}{n} \sum_{i=1}^{N_2} \tau_i.
\]  

Note \( E(y_i^2) = \sigma_y^2 + \mu_y^2 \), and

\[
E\left(\frac{w_2^2}{n_s}\right) = E\left(\frac{n_s^2}{n^2} f\right) = E\left(\frac{n_s^2}{n^2} n_2\right) = \frac{f}{n} E(n_2) = \frac{f}{n} (nW_2) = \frac{W_2 f}{n},
\]

if we assume \( \frac{n}{N} \approx \frac{n_2}{N_2} \).

Since \( \bar{y} \) is the original HH mean estimator, the variance of \( \hat{y} \) is given by

\[
\text{Var}(\hat{y}) = \theta \sigma_y^2 + \lambda \sigma_{y(2)}^2 + \frac{W_2 f}{n} \left[\left(\frac{\sigma_y^2 + \mu_y^2}{n_2} + \sigma_{y(2)}^2 + \sigma_y^2\right) W\right],
\]

where \( \theta = \frac{(N-n)}{N n} \) and \( \lambda = \frac{(f-1)W_2}{n} \). It is easy to notice that \( \frac{W_2 f}{n} \left[\left(\frac{\sigma_y^2 + \mu_y^2}{n_2} + \sigma_{y(2)}^2 + \sigma_y^2\right) W\right] \) is the penalty for using ORRT model.

### IV.2. Some Existing Mean Estimators under Measurement Errors and Non-response

Let the measurement error of the auxiliary variable \( X \) in the population be given by \( V_i = x_i - X_i \). Let the respective measurement errors associated with the study variable \( Y \) in the population and the scrambled variable \( Z \) in the face-to-face phase be given by \( U_i = y_i - Y_i \) and \( P_i = z_i - Z_i \). These measurement errors are
assumed to be random and uncorrelated with mean zero and variances $\sigma_v^2$, $\sigma_u^2$, and $\sigma_p^2$ respectively.

Assume population mean $\mu_x$ of the auxiliary variable X is known, and non-response happens on both X and Y. Some notations are given below

\[ \Omega_y = \sum_{i=1}^{n} (y_i - \mu_y), \quad \text{(IV.12)} \]
\[ \Omega_x = \sum_{i=1}^{n} (x_i - \mu_x), \quad \text{(IV.13)} \]
\[ \Omega_u = \sum_{i=1}^{n_1} U_i + \sum_{i=1}^{n_2} P_i, \quad \text{(IV.14)} \]

and

\[ \Omega_v = \sum_{i=1}^{n} V_i, \quad \text{(IV.15)} \]

where $U_i$, $P_i$, $V_i$ are measurement errors on Y, Z and X, respectively. Let $e_0^* = \frac{1}{n\mu_y} (\Omega_y + \Omega_u)$ and $e_1^* = \frac{1}{n\mu_x} (\Omega_x + \Omega_u)$. In other words, $\hat{y}^* = (1 + e_0^*)\mu_y$ and $\bar{x}^* = (1 + e_1^*)\mu_x$, where $\hat{y}^* = w_1\bar{y}_1^* + w_2\bar{y}_2^*$ and $\bar{x}^* = w_1\bar{x}_1^* + w_2\bar{x}_2^*$ in the presence of measurement errors.

Under the assumption of bivariate normality (Sukhatme et al.1970)[78]:

\[ E(e_0^*) = E(e_1^*) = 0; \quad \text{(IV.16)} \]

\[ E(e_0^{*2}) = \frac{1}{\mu_y} \left( \theta (\sigma_y^2 + \sigma_u^2) + \lambda (\sigma_y^2 + \sigma_p^2) + \frac{W_2}{n} f \left[ \frac{([\sigma_y^2 + \mu_y (2) \sigma_T^2] + \sigma_s^2)W}{(\mu_T W + 1 - W)^2} \right] \right); \quad \text{(IV.17)} \]
\[
E(e_1^2) = \frac{1}{\mu_x^2} \left[ \theta(\sigma_x^2 + \sigma_u^2) + \lambda(\sigma_x^2(2) + \sigma_v^2) \right]; \tag{IV.18}
\]
and
\[
E(e_0e_1^* ) = \theta \rho_{yx} \frac{\sigma_y \sigma_x}{\mu_y} + \lambda \rho_{zx(2)} \frac{\sigma_z \sigma_x(2)}{\mu_y}, \tag{IV.19}
\]
where \( \theta = \frac{(N - n)}{Nn} \), \( \lambda = \frac{(f - 1)W}{n} \), \( \sigma_z^2 = \sigma_y^2 + \sigma_s^2W + \sigma_T^2(\sigma_y^2 + \mu_y^2)W \), and
\[
\rho_{zx(2)} = \frac{\rho_{yx(2)}}{\sqrt{1 + \frac{\sigma_y^2W}{\sigma_y^2(2)} + \frac{\sigma_T^2(\sigma_y(2) + \mu_y^2)W}{\sigma_y^2(2)}}}. \tag{IV.20}
\]

Some existing mean estimators in the presence of measurement errors and non-response using the modified HH two-phase sampling are listed below:

- The ordinary mean estimator is given by
  \[
  \hat{\mu}_{yw}^{HH} = \hat{y}^* = w_1 \bar{y}_1^* + w_2 \bar{y}_2^*. \tag{IV.21}
  \]
  It can be written as
  \[
  \hat{\mu}_{yw}^{HH} = (1 + e_0^*) \mu_y. \tag{IV.22}
  \]
  The difference between the ordinary mean estimator and the true mean can be written as
  \[
  \hat{\mu}_{yw}^{HH} - \mu_y = e_0^* \mu_y. \tag{IV.23}
  \]
  Taking square and then expected value on both side of (IV.23), the MSE of \( \hat{\mu}_{yw} \) is given by
  \[
  MSE^*(\hat{\mu}_{yw}^{HH}) = \theta(\sigma_y^2 + \sigma_u^2) + \lambda(\sigma_y^2(2) + \sigma_p^2) + G \tag{IV.24}
  \]
where $G = w_2 f \left[ \frac{[\sigma_y^2 + \mu_y^2 \sigma_z^2]}{(\mu_y W + 1 - W)^2} \right].$

- A ratio estimator corresponding to the one in Gupta et al. (2014) is given by
  \[ \hat{\mu}_{rw}^{HH} = \frac{\bar{y}^*}{\bar{x}^*} \hat{\mu}_x = \hat{R}_{w}^{HH} \hat{\mu}_x, \]
  (IV.25)
  where $\bar{y}^* = w_1 \bar{y}_1^* + w_2 \bar{y}_2^*.$

It can be written as
  \[ \hat{\mu}_{rw}^{HH} = \frac{(1 + e_0^*) \mu_y}{(1 + e_1^*) \mu_x} \mu_x \]
  \[ = \mu_y (1 + e_0^*) (1 + e_1^*)^{-1} \]
  \[ = \mu_y (1 + e_0^*) (1 - e_1^* + e_1^{*2} - e_1^{*3} + ...) \]
  \[ = \mu_y (1 - e_1^* + e_1^{*2} + e_0^* - e_0^* e_1^* + ...). \]
  (IV.26)

Using second order approximation, the difference between the ratio estimator and the true mean can be written as
  \[ \hat{\mu}_{rw}^{HH} - \mu_y = \mu_y (-e_1^* + e_1^{*2} + e_0^* - e_0^* e_1^*). \]
  (IV.27)

Taking square and then expected value on both side of (IV.27), the MSE of $\hat{\mu}_{rw}^{HH}$ is given by
  \[ MSE^{*}(\hat{\mu}_{rw}^{HH}) = \theta (\sigma_y^2 + R^2 \sigma_x^2 - 2 R \rho_{yx} \sigma_y \sigma_x) + \lambda (\sigma_{y(2)}^2 + R^2 \sigma_{x(2)}^2) - 2 R \rho_{zx(2)} \sigma_z \sigma_{x(2)} + \theta (\sigma_u^2 + R^2 \sigma_v^2) + \lambda (\sigma_p^2 + R^2 \sigma_v^2) + G, \]
  (IV.28)

where $R = \frac{\mu_y}{\mu_x}.$

The MSEs of the above mean estimators without measurement error may be obtained by letting $\sigma_u^2 = \sigma_p^2 = \sigma_v^2 = 0$ in (IV.24) and (IV.28).
IV.3. Generalized Estimator under ORRT Models and HH Two-phase Sampling Technique in the Presence of Measurement Errors

With this background, we use the generalized mean estimator used in Khalil et al. (2018) [34] and Chapter III. This estimator includes a wide variety of mean estimators as special cases. The non-response version is given by:

\[
\hat{\mu}_{HH}^{pw} = (\hat{y}^* + k(\mu_x - \bar{x}^*))\left(\frac{\bar{D}}{d}\right)^v, \tag{IV.29}
\]

where

\[
\hat{y}^* = w_1\tilde{y}_1^* + w_2\tilde{y}_2^*, \quad \bar{x}^* = w_1\bar{x}_1^* + w_2\bar{x}_2^*, \quad \bar{d} = \phi(\alpha\bar{x}^* + \beta) + (1 - \phi)(\alpha\mu_x + \beta),
\]

\[
\bar{D} = \alpha\mu_x + \beta, \quad k \text{ and } v \text{ are suitable constants. } \phi \text{ is assumed to be an unknown constant whose value is to be determined from optimally considerations. } \alpha (\alpha \neq 0) \text{ and } \beta \text{ are assumed to be some known parameters of the auxiliary variable } X, \text{ such as coefficient of variation } (C_x), \text{ kurtosis, and correlation coefficient } (\rho_{yx}) \text{ etc. Please note here with different values of } \alpha \text{ and } \beta, \text{ we can obtain various estimators. Also, with } v=1 \text{ we get various ratio estimators and with } v=-1 \text{ we get various product estimators.}
\]

IV.3.1 Bias and MSE of the Generalized Mean Estimator

The generalized mean estimator will be studied under both modified HH two-phase sampling and measurement errors. According to the notations in Section IV.2, it can be written as

\[
\hat{\mu}_{HH}^{pw} = ((1 + e_0^*)\mu_y + k(\mu_x - (1 + e_1^*)\mu_x))\left(\frac{\alpha\mu_x + \beta}{\phi(\alpha(1 + e_1^*)\mu_x + \beta) + (1 - \phi)(\alpha\mu_x + \beta)}\right)^v. \tag{IV.30}
\]
Using Taylor’s approximation and retaining terms of order up to 2, the difference between the generalized mean estimator and the true mean can be written as

\[
\hat{\mu}_{RH}^{pw} - \mu_y = ((1 + e_0^*)\mu_y + k(\mu_x - (1 + e_1^*)\mu_x)) \left( \frac{\alpha \mu_x + \beta}{\phi(\alpha(1 + e_1^*)\mu_x + \beta) + (1 - \phi)(\alpha \mu_x + \beta)} \right)^v - \mu_y
\]

\[
= e_0^*\mu_y + e_1^*[-k \mu_x + \mu_y v(\frac{-\alpha \phi \mu_x}{\alpha \mu_x + \beta})] + \frac{1}{2!} [2e_0^*e_1^*\mu_y v(\frac{-\alpha \phi \mu_x}{\alpha \mu_x + \beta}) + e_1^2 (2k v \mu_x (\frac{\alpha \phi \mu_x}{\alpha \mu_x + \beta}) + v(v + 1) \mu_y (\frac{-\alpha \phi \mu_x}{\alpha \mu_x + \beta})^2)]
\]

\[\text{(IV.31)}\]

Taking expectation on both side of (IV.31), the bias of the generalized mean estimator \(\hat{\mu}_{pw}^{HH}\), correct to second order or approximation, is given by

\[
\text{Bias}^*(\hat{\mu}_{pw}^{HH}) \approx \theta [(kH + \frac{v + 1}{v} \mu_y H^2)(\sigma_x^2 + \sigma_v^2) - H \rho_{yx} \sigma_y \sigma_x] + \lambda [(kH + \frac{v + 1}{v} \mu_y H^2)(\sigma_{x(2)}^2 + \sigma_v^2) - H \rho_{zx(2)} \sigma_z \sigma_{x(2)}],
\]

\[\text{(IV.32)}\]

where \(H = \frac{\alpha \phi v}{\alpha \mu_x + \beta}\). The bias of \(\hat{\mu}_{pw}^{HH}\) without measurement error may be obtained by setting \(\sigma_v^2 = 0\) in above equation.

To determine the expression for MSE of the generalized mean estimator (IV.29), we take square of (IV.31) on both sides and retaining terms of order up to 2 which is given by

\[
(\hat{\mu}_{pw}^{HH} - \mu_y)^2 = e_0^{*2} \mu_y^2 + k^2 \mu_x^2 e_1^{*2} + (H \mu_x \mu_y e_1^*)^2 - 2e_0^*e_1^* k \mu_x \mu_y - 2e_0^*e_1^* H \mu_x \mu_y + 2e_1^{*2} k H \mu_x^2 \mu_y.
\]

\[\text{(IV.33)}\]
Taking the expected value on both side of (IV.33), the expression for MSE of
\( \mu_{pw}^H \), correct to the first order approximation is given by

\[
MSE^*(\hat{\mu}_{pw}^H) = E(\hat{\mu}_{pw}^H - \mu_y)^2
\]

\[
\approx \theta(\sigma_y^2 + k^2\sigma_x^2 + \phi^2v^2R_{pw}^2\sigma_x^2) + 2k\phi vR_{pw}\sigma_x^2 - 2k\rho_{yx}\sigma_x\sigma_y - 2\phi vR_{pw}\rho_{yx}
\]

\[
\sigma_x\sigma_y) + \lambda(\sigma_y^2 + k^2\sigma_x^2) + \phi^2v^2R_{pw}^2\sigma_x^2 + 2k\phi vR_{pw}\sigma_x^2 - 2k\rho_{zx}(2)
\]

\[
\sigma_z\sigma_x(2) - 2\phi vR_{pw}\rho_{zx}(2)\sigma_x\sigma_z(2)) + \theta(\sigma_y^2 + k^2\sigma_u^2 + \phi^2v^2R_{pw}^2\sigma_u^2 + 2k\phi vR_{pw}\sigma_u^2 + G
\]

\[
= \theta[\sigma_y^2 + (k + \phi vR_{pw})^2\sigma_x^2 - 2(k + \phi vR_{pw})\rho_{yx}\sigma_x\sigma_y] +
\]

\[
\lambda[\sigma_y^2 + (k + \phi vR_{pw})^2\sigma_x^2(2) - 2(k + \phi vR_{pw})\rho_{zx}(2)\sigma_x\sigma_z(2)] +
\]

\[
\theta[\sigma_u^2 + (k + \phi vR_{pw})^2\sigma_v^2] + \lambda[\sigma_y^2 + (k + \phi vR_{pw})^2\sigma_u^2] + G,
\]

(IV.34)

where \( R_{pw} = \frac{\alpha_{\mu_y}}{\alpha_{\mu_x} + \beta} \). Minimization of the above expression (IV.34) with respect to \( \phi \)
yields its optimum value as:

\[
\phi_{opt} \approx \frac{\theta(\rho_{yx}\sigma_x\sigma_y - k(\sigma_x^2 + \sigma_y^2)) + \lambda(\rho_{zx}(2)\sigma_z\sigma_x(2) - k(\sigma_x^2(2) + \sigma_y^2(2)))}{vR_{pw}[\theta(\sigma_x^2 + \sigma_y^2) + \lambda(\sigma_x^2(2) + \sigma_y^2(2))]}.
\]

(IV.35)

Substitution of \( \phi_{opt} \) in MSE(\( \hat{\mu}_{pw}^H \)) yields the minimum value as:

\[
MSE_{\text{min}}^*(\hat{\mu}_{pw}^H) \approx \theta(\sigma_y^2 + P^2\sigma_x^2 - 2P\rho_{yx}\sigma_x\sigma_y) + \lambda(\sigma_y^2(2) + P^2\sigma_x^2(2) - 2P\rho_{zx}(2)\sigma_z\sigma_x(2)) +
\]

\[
\theta(\sigma_u^2 + P^2\sigma_v^2) + \lambda(\sigma_p^2 + P^2\sigma_v^2) + G,
\]

(IV.36)

where \( P = \frac{\theta\rho_{yx}\sigma_x\sigma_y + \lambda\rho_{zx}(2)\sigma_z\sigma_x(2)}{\theta(\sigma_x^2 + \sigma_y^2) + \lambda(\sigma_x^2(2) + \sigma_y^2(2))}. \)
The expression for the minimized MSE of the generalized estimator without ME may be obtained by putting $\sigma^2_u = \sigma^2_v = \sigma^2_p = 0$ in the (IV.36), which gives

$$MSE_{\text{min}}(\hat{\mu}_{pw}^{HH}) \geq \theta \sigma^2_y + P^2 \sigma^2_x - 2P \rho_{yx} \sigma_x \sigma_y + \lambda \sigma^2_{y(2)} + P^2 \sigma^2_{x(2)} - 2P \rho_{zx(2)} \sigma_z \sigma_{x(2)} + G,$$

where $G = \frac{W^2f}{n} \left[ \frac{\sigma^2_x \sigma^2_y + \sigma^2_{y(2)} + \mu^2_{x(2)}}{(\mu_T W + 1 - W)^2} \right]$.

**IV.3.2 Efficiency Comparisons**

Comparing the MSE expressions of $\hat{\mu}_{yw}^{HH}$ (IV.24), $\hat{\mu}_{rw}^{HH}$ (IV.28), and $\hat{\mu}_{pw}^{HH}$ (IV.36) with measurement errors, it can be verified easily that

- $MSE_{\text{min}}(\hat{\mu}_{pw}^{HH}) < MSE^*(\hat{\mu}_{yw}^{HH})$ if

  $$MSE_{\text{min}}(\hat{\mu}_{pw}^{HH}) - MSE^*(\hat{\mu}_{yw}^{HH}) = P^2 \left( \theta \sigma^2_x + \sigma^2_v \right) + \lambda \left( \sigma^2_{y(2)} + \sigma^2_{x(2)} \right) = - \frac{(\theta \rho_{yx} \sigma_x \sigma_y + \lambda \rho_{zx(2)} \sigma_z \sigma_{x(2)})^2}{\theta \sigma^2_x + \sigma^2_v + \lambda \sigma^2_{x(2)} + \sigma^2_{x(2)}} < 0; \quad (IV.38)$$

- $MSE_{\text{min}}(\hat{\mu}_{pw}^{HH}) < MSE^*(\hat{\mu}_{rw}^{HH})$ if

  $$MSE_{\text{min}}(\hat{\mu}_{pw}^{HH}) - MSE^*(\hat{\mu}_{rw}^{HH}) = (P^2 - R^2) \left( \theta \sigma^2_x + \sigma^2_v \right) + \lambda \left( \sigma^2_{x(2)} + \sigma^2_{x(2)} \right) - 2(P + R) \left( \theta \rho_{yx} \sigma_x \sigma_y + \lambda \rho_{zx(2)} \sigma_z \sigma_{x(2)} \right) < 0; \quad (IV.39)$$
In other words, if
\[
\frac{(P^2 - R^2)(\theta(\sigma_x^2 + \sigma_v^2) + \lambda(\sigma_{x(2)}^2 + \sigma_v^2))}{2(P + R)(\theta \rho_{yx} \sigma_x \sigma_y + \lambda \rho_{zx(2)} \sigma_z \sigma_{x(2)})} = \frac{(P - R)(\theta(\sigma_x^2 + \sigma_v^2) + \lambda(\sigma_{x(2)}^2 + \sigma_v^2))}{2(\theta \rho_{yx} \sigma_x \sigma_y + \lambda \rho_{zx(2)} \sigma_z \sigma_{x(2)})}
\]
\[
= \frac{P - R}{2P}
\]
\[
= \frac{1}{2} - \frac{R}{2P}
\]
\[
= \frac{1}{2} - \frac{\mu_y}{2\mu_x} \frac{\theta(\sigma_x^2 + \sigma_v^2) + \lambda(\sigma_{x(2)}^2 + \sigma_v^2)}{\theta \rho_{yx} \sigma_x \sigma_y + \lambda \rho_{zx(2)} \sigma_z \sigma_{x(2)}} < 1;
\]

and

\[\text{MSE}^*(\hat{\mu}_{yw}^{HH}) < \text{MSE}^*(\hat{\mu}_{rw}^{HH})\] if

\[
\text{MSE}^*(\hat{\mu}_{rw}^{HH}) - \text{MSE}^*(\hat{\mu}_{yw}^{HH}) = R^2(\theta(\sigma_x^2 + \sigma_v^2) + \lambda(\sigma_{x(2)}^2 + \sigma_v^2)) - 2R(\theta \rho_{yx} \sigma_x \sigma_y + \lambda \rho_{zx(2)} \sigma_z \sigma_{x(2)}) < 0; \]

\[\text{MSE}^*(\hat{\mu}_{yw}^{HH}) < \text{MSE}^*(\hat{\mu}_{rw}^{HH})\] if

\[
\frac{R^2 \theta(\sigma_x^2 + \sigma_v^2) + \lambda(\sigma_{x(2)}^2 + \sigma_v^2)}{2R \theta \rho_{yx} \sigma_x \sigma_y + \lambda \rho_{zx(2)} \sigma_z \sigma_{x(2)}} = \frac{\mu_y}{2\mu_x} \frac{\theta(\sigma_x^2 + \sigma_v^2) + \lambda(\sigma_{x(2)}^2 + \sigma_v^2)}{\theta \rho_{yx} \sigma_x \sigma_y + \lambda \rho_{zx(2)} \sigma_z \sigma_{x(2)}} < 1. \]

The conditions (IV.38) and (IV.40) always hold true. From (IV.42), the ratio estimator is more efficient than the ordinary mean estimator only if the measurement error on auxiliary variable X (\(\sigma_v^2\)) is small, and X and Y are highly correlated.

**IV.4. Simulation Study**

In this section, we will evaluate the performance of the generalized mean estimator under non-response and measurement errors with the other two estimators by a simulation study. In the generalized mean estimator, we choose \(v\) and \(k\) to be 1,
and ϕ to be its optimum value. As demonstrated in Chapter III simulations, we could use various parameters associated with the auxiliary variable such as the coefficient of variation \(C_x\) or kurtosis for \(α\) and \(β\), but these choices do not impact the results in any meaningful way. We will only show the results where \(α = 1\) and \(β = 0\). The scrambling variable \(S\) is taken to be a normal variate with mean equal to zero and vary variance \((0.2*σ_x^2, 0.5*σ_x^2, \text{ and } 1*σ_x^2)\). And \(T\) is also taken to be a normal variate but with mean equal to one and varying variances \((0, 0.5, 1)\). The measurement errors of \(X\) have a normal distribution with mean zero in both phases; the measurement errors of \(Y\) in the first phase and \(Z\) in the second phase have a normal distribution with mean zero. We demonstrate different variances \((0, 5, 10)\) for measurement errors. We consider a finite population of size 5000 generated from bivariate normal distribution with means and covariance of \((Y, X)\) as given below.

**Population**  
\[
\mu = \begin{bmatrix} 10 \\ 6 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 16 & 9.051 \\ 9.051 & 8 \end{bmatrix}, \quad ρ_{yx} = 0.8
\]

The real parameters of the set of 5000 data points we generated using R are very close to the parameter values in (A) but not exactly same. For the simulation study, we used parameter values in (B) and not those in (A).

\[
\begin{align*}
μ_x &= 6, \quad σ_x^2 = 8, \quad μ_y = 10, \quad σ_y^2 = 16, \quad ρ_{yx} = 0.8 \\
μ_x &= 6.0228, \quad σ_x^2 = 8.1830, \quad μ_y = 9.9864, \quad σ_y^2 = 16.1215, \quad ρ_{yx} = 0.8024
\end{align*}
\]

We consider samples of size \(n = 500\) using SRSWOR (simple random sampling without replacement) and assume response rate is 40% in the first phase. This means in the first phase only 200 \((n_1)\) subjects provide a response to the survey question and 300 \((n_2)\) of them do not. In the second phase, we take another sample \((n_s = \frac{n_2}{T})\) from
non-respondent group by using $f = 2, 3, 4$, respectively. Different response rates of 20%, 40% and 60% are also compared in the simulation study. Coding for the simulations was done in R and results are averaged over 5,000 iterations. The empirical MSE of the estimator $\hat{\mu}_y$ is computed by

$$MSE^*(\hat{\mu}_w) = \frac{1}{5000} \sum_{i=1}^{5000} (\hat{\mu}_w^{HH} - \mu_y)^2,$$

where $\hat{\mu}_w^{HH} = \hat{\mu}_{yw}^{HH}, \hat{\mu}_{rw}^{HH},$ and $\hat{\mu}_{pw}^{HH}$. Here, $\mu_y$ is the population mean of the sensitive study variable.

The percent relative efficiencies (PREs) of the estimators ($\hat{\mu}_w^{HH}$) with respect to the ordinary mean estimator ($\hat{\mu}_{yw}^{HH}$) is defined as

$$PRE = \frac{MSE^*(\hat{\mu}_{yw}^{HH})}{MSE^*(\hat{\mu}_w^{HH})} \times 100.$$

We will also use the unified measure $\delta$ of efficiency and privacy as defined in Gupta et al. (2018)[24]. It is given by

$$\delta = \frac{MSE^*(\hat{\mu}_w^{HH})}{\Delta_{DP}}.$$

In (IV.45), MSE is used in place of Var(.) to account for biased estimators.
Table IV.1. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators when Response Rate = 40%, $\sigma^2_v = \sigma^2_u = \sigma^2_p = 1$, $f = 2$ and $\sigma^2_s = 0.2\sigma^2_x$.

<table>
<thead>
<tr>
<th>Est. $\hat{\mu}_{HH}$</th>
<th>W</th>
<th>$\sigma^2_T$</th>
<th>MSE Without ME</th>
<th>MSE With ME</th>
<th>PRE Without ME</th>
<th>PRE With ME</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\mu}_{HH}$</td>
<td>0</td>
<td>0.5</td>
<td>0.0508</td>
<td>0.0537</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0330</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0514</td>
<td>0.0551</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0339</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.1190</td>
<td>0.1220</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0021</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1140</td>
<td>0.1164</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td>0.1865</td>
<td>0.1895</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0016</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1787</td>
<td>0.1810</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0016</td>
</tr>
<tr>
<td>$\mu_{yq}$</td>
<td>0.8</td>
<td>0.5</td>
<td>0.1519</td>
<td>0.1554</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0337</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0517</td>
<td>0.0554</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0340</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.1603</td>
<td>0.1633</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1603</td>
<td>0.1631</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td>0.2670</td>
<td>0.2700</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.2698</td>
<td>0.2723</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0023</td>
</tr>
<tr>
<td>$\hat{\mu}_{HH}$</td>
<td>0.8</td>
<td>0.5</td>
<td>0.1527</td>
<td>0.1557</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0342</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0529</td>
<td>0.0567</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0348</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.1875</td>
<td>0.1905</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1790</td>
<td>0.1851</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td>0.3197</td>
<td>0.3227</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.3169</td>
<td>0.3158</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0027</td>
</tr>
<tr>
<td>$\mu_{wq}$</td>
<td>0</td>
<td>0.5</td>
<td>0.0261</td>
<td>0.0373</td>
<td>194.6360</td>
<td>143.9678</td>
<td>0.0229</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0269</td>
<td>0.0388</td>
<td>191.0781</td>
<td>142.0103</td>
<td>0.0238</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.0943</td>
<td>0.1055</td>
<td>126.1930</td>
<td>115.6398</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0910</td>
<td>0.1055</td>
<td>125.2747</td>
<td>110.3318</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td>0.1618</td>
<td>0.1730</td>
<td>115.2658</td>
<td>109.5376</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1566</td>
<td>0.1650</td>
<td>114.1124</td>
<td>109.6970</td>
<td>0.0014</td>
</tr>
<tr>
<td>$\mu_{HH}$</td>
<td>0.8</td>
<td>0.5</td>
<td>0.0273</td>
<td>0.0385</td>
<td>190.1099</td>
<td>142.5974</td>
<td>0.0237</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.0283</td>
<td>0.0385</td>
<td>182.6855</td>
<td>143.8961</td>
<td>0.0237</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.1356</td>
<td>0.1468</td>
<td>118.2153</td>
<td>111.2398</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.1381</td>
<td>0.1453</td>
<td>116.0753</td>
<td>112.2505</td>
<td>0.0025</td>
</tr>
</tbody>
</table>
\[
\begin{array}{cccccc}
\mu_{\mu} & \hat{\mu}_{\mu}^{HH} & \hat{\mu}_{\mu}^{\nu} & \hat{\mu}_{e}^{\nu} & \hat{\mu}_{e}^{HH} & \hat{\mu}_{e}^{\nu} \\
0.2423 & 0.2534 & 110.1940 & 106.5509 & 0.0022 & \\
0.0280 & 0.0393 & 188.2143 & 141.7303 & 0.0242 & \\
0.1628 & 0.1740 & 115.1720 & 109.4828 & 0.0029 & \\
0.2950 & 0.3062 & 108.3729 & 105.3886 & 0.0026 & \\
0.0191 & 0.0255 & 265.9686 & 210.5882 & 0.0157 & \\
0.0873 & 0.0937 & 136.3116 & 130.2026 & 0.0016 & \\
0.1549 & 0.1612 & 120.4003 & 117.5558 & 0.0014 & \\
0.0203 & 0.0267 & 255.6650 & 205.6180 & 0.0164 & \\
0.1286 & 0.1350 & 124.6501 & 120.9630 & 0.0023 & \\
0.2353 & 0.2417 & 113.4722 & 111.7087 & 0.0021 & \\
0.0211 & 0.0274 & 249.7630 & 203.2847 & 0.0168 & \\
0.1559 & 0.1622 & 120.2694 & 117.4476 & 0.0027 & \\
0.2880 & 0.2944 & 111.0069 & 109.6128 & 0.0025 & \\
\end{array}
\]
Table IV.2. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators when Response Rate = 40% , $\sigma^2_v = \sigma^2_u = \sigma^2_p = 1$, $f = 2$ and $\sigma^2_s = 0.5*\sigma^2_x$.

<table>
<thead>
<tr>
<th>Est. $\hat{\mu}_{HH}$</th>
<th>W</th>
<th>$\sigma^2_T$</th>
<th>MSE Without ME</th>
<th>MSE With ME</th>
<th>PRE Without ME</th>
<th>PRE With ME</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.537</td>
<td>0.0567</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0139</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.0545</td>
<td>0.0584</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0143</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td>0.1219</td>
<td>0.1249</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td>0.1172</td>
<td>0.1196</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.1895</td>
<td>0.1924</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0016</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.1820</td>
<td>0.1843</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0016</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.8</td>
<td>0.0566</td>
<td>0.0596</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0146</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.8</td>
<td>0.0559</td>
<td>0.0598</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0146</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.8</td>
<td>0.1650</td>
<td>0.1680</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0027</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.8</td>
<td>0.1651</td>
<td>0.1679</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0027</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.8</td>
<td>0.2716</td>
<td>0.2746</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.8</td>
<td>0.2746</td>
<td>0.2771</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0.0586</td>
<td>0.0616</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0151</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0.0587</td>
<td>0.0625</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0153</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0.1933</td>
<td>0.1963</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0032</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0.1847</td>
<td>0.1916</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0.3255</td>
<td>0.3285</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0.3225</td>
<td>0.3226</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0027</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0</td>
<td>0.0290</td>
<td>0.0403</td>
<td>185.1724</td>
<td>140.6948</td>
<td>0.0099</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0</td>
<td>0.0301</td>
<td>0.0420</td>
<td>181.0631</td>
<td>139.0476</td>
<td>0.0103</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0972</td>
<td>0.1084</td>
<td>125.4115</td>
<td>115.2214</td>
<td>0.0018</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0940</td>
<td>0.1028</td>
<td>124.6809</td>
<td>116.3424</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.1648</td>
<td>0.1759</td>
<td>114.9879</td>
<td>109.3803</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.1597</td>
<td>0.1679</td>
<td>113.9637</td>
<td>107.7677</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0319</td>
<td>0.0432</td>
<td>177.4295</td>
<td>137.9630</td>
<td>0.0106</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0330</td>
<td>0.0454</td>
<td>169.3939</td>
<td>131.7181</td>
<td>0.0111</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0</td>
<td>0.1403</td>
<td>0.1514</td>
<td>117.6051</td>
<td>110.9643</td>
<td>0.0025</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0</td>
<td>0.1425</td>
<td>0.1498</td>
<td>115.8596</td>
<td>112.0828</td>
<td>0.0024</td>
</tr>
<tr>
<td>$\hat{\mu}_{HH}^{pw}$</td>
<td>$\hat{\mu}_{HH}^{pw}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2469</td>
<td>0.2581</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2531</td>
<td>0.2588</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0339</td>
<td>0.0451</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0345</td>
<td>0.0451</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1686</td>
<td>0.1798</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1640</td>
<td>0.1741</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3008</td>
<td>0.3119</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2940</td>
<td>0.3058</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0220</td>
<td>0.0284</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0228</td>
<td>0.0296</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0903</td>
<td>0.0966</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0864</td>
<td>0.0912</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1578</td>
<td>0.1641</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1518</td>
<td>0.1562</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1333</td>
<td>0.1397</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1348</td>
<td>0.1392</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2399</td>
<td>0.2463</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2451</td>
<td>0.2485</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2069</td>
<td>0.3033</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2076</td>
<td>0.3039</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1617</td>
<td>0.1680</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1553</td>
<td>0.1629</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2938</td>
<td>0.3002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2847</td>
<td>0.2944</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\hat{\mu}_{HH}$</th>
<th>0.2581</th>
<th>110.0041</th>
<th>106.3929</th>
<th>0.0022</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2531</td>
<td>108.4947</td>
<td>107.0711</td>
<td>0.0022</td>
<td></td>
</tr>
<tr>
<td>0.0345</td>
<td>170.1449</td>
<td>138.5809</td>
<td>0.0110</td>
<td></td>
</tr>
<tr>
<td>0.1640</td>
<td>112.6220</td>
<td>110.0517</td>
<td>0.0028</td>
<td></td>
</tr>
<tr>
<td>0.2940</td>
<td>109.6939</td>
<td>105.4938</td>
<td>0.0026</td>
<td></td>
</tr>
<tr>
<td>0.0228</td>
<td>239.0351</td>
<td>197.2973</td>
<td>0.0072</td>
<td></td>
</tr>
<tr>
<td>0.0864</td>
<td>135.6481</td>
<td>131.1404</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td>0.1518</td>
<td>119.8946</td>
<td>117.9898</td>
<td>0.0013</td>
<td></td>
</tr>
<tr>
<td>0.0252</td>
<td>221.8254</td>
<td>185.7143</td>
<td>0.0079</td>
<td></td>
</tr>
<tr>
<td>0.1348</td>
<td>122.4777</td>
<td>120.6178</td>
<td>0.0023</td>
<td></td>
</tr>
<tr>
<td>0.2485</td>
<td>112.0359</td>
<td>111.5091</td>
<td>0.0021</td>
<td></td>
</tr>
<tr>
<td>0.0276</td>
<td>212.6812</td>
<td>184.3658</td>
<td>0.0083</td>
<td></td>
</tr>
<tr>
<td>0.1629</td>
<td>118.9311</td>
<td>117.6182</td>
<td>0.0027</td>
<td></td>
</tr>
<tr>
<td>0.2944</td>
<td>113.2771</td>
<td>109.5788</td>
<td>0.0025</td>
<td></td>
</tr>
</tbody>
</table>
Table IV.3. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators when Response Rate = 40%, $\sigma_v^2 = \sigma_u^2 = \sigma_p^2 = 1$, $f = 2$ and $\sigma_s^2 = 1\sigma_z^2$.

<table>
<thead>
<tr>
<th>Est.</th>
<th>W</th>
<th>$\sigma_T^2$</th>
<th>MSE Without ME</th>
<th>MSE With ME</th>
<th>PRE Without ME</th>
<th>PRE With ME</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\hat{\mu}_{HH}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0586</td>
<td>0.0616</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0076</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.1268</td>
<td>0.1298</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1943</td>
<td>0.1973</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.0644</td>
<td>0.0674</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0083</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.1728</td>
<td>0.1757</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0027</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.2793</td>
<td>0.2823</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.0683</td>
<td>0.0713</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0088</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.2030</td>
<td>0.2060</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0031</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.3351</td>
<td>0.3380</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0027</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.3320</td>
<td>0.3334</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0027</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.0339</td>
<td>0.0451</td>
<td>172.8614</td>
<td>136.5854</td>
<td>0.0055</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.1021</td>
<td>0.1133</td>
<td>124.1920</td>
<td>114.5631</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1696</td>
<td>0.1808</td>
<td>114.5637</td>
<td>109.1261</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.0398</td>
<td>0.0510</td>
<td>161.8090</td>
<td>132.1569</td>
<td>0.0063</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
<td>0.1481</td>
<td>0.1592</td>
<td>116.6779</td>
<td>110.3643</td>
<td>0.0024</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1499</td>
<td>0.1573</td>
<td>115.2769</td>
<td>111.6974</td>
<td>0.0024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\hat{\mu}_{HH}$</td>
<td>$\hat{\mu}_{pw}$</td>
<td>$\sigma^2_v$</td>
<td>$\sigma^2_u$</td>
<td>$\sigma^2_p$</td>
<td>$\text{MSE}$</td>
<td>$\text{PRE}$</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
<td>0.2546</td>
<td>0.2658</td>
<td>109.7015</td>
<td>106.2077</td>
<td>0.0022</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2606</td>
<td>0.2663</td>
<td>108.3653</td>
<td>106.9846</td>
<td>0.0022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.0437</td>
<td>0.0549</td>
<td>156.2929</td>
<td>129.8725</td>
<td>0.0067</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0445</td>
<td>0.0550</td>
<td>153.9326</td>
<td>131.2727</td>
<td>0.0068</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.1783</td>
<td>0.1895</td>
<td>113.8531</td>
<td>108.7071</td>
<td>0.0029</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1730</td>
<td>0.1842</td>
<td>118.0347</td>
<td>109.7177</td>
<td>0.0028</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.3104</td>
<td>0.3215</td>
<td>107.9575</td>
<td>105.1322</td>
<td>0.0026</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.3129</td>
<td>0.3160</td>
<td>106.1042</td>
<td>105.5063</td>
<td>0.0026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.0269</td>
<td>0.0333</td>
<td>217.8439</td>
<td>184.9850</td>
<td>0.0041</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0279</td>
<td>0.0348</td>
<td>213.9785</td>
<td>183.0460</td>
<td>0.0043</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.0952</td>
<td>0.1015</td>
<td>133.1933</td>
<td>127.8818</td>
<td>0.0015</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0916</td>
<td>0.0964</td>
<td>133.6245</td>
<td>129.3568</td>
<td>0.0015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1627</td>
<td>0.1690</td>
<td>119.4222</td>
<td>116.7456</td>
<td>0.0014</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1571</td>
<td>0.1616</td>
<td>119.2234</td>
<td>117.2649</td>
<td>0.0013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.0328</td>
<td>0.0392</td>
<td>196.3415</td>
<td>171.9388</td>
<td>0.0048</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0328</td>
<td>0.0399</td>
<td>192.6829</td>
<td>168.4211</td>
<td>0.0049</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.1411</td>
<td>0.1474</td>
<td>122.4663</td>
<td>119.1995</td>
<td>0.0022</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1423</td>
<td>0.1469</td>
<td>121.4336</td>
<td>119.6052</td>
<td>0.0022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.2477</td>
<td>0.2540</td>
<td>112.7574</td>
<td>111.1417</td>
<td>0.0021</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.2477</td>
<td>0.2562</td>
<td>114.0089</td>
<td>111.2022</td>
<td>0.0021</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.0367</td>
<td>0.0431</td>
<td>186.1035</td>
<td>165.4292</td>
<td>0.0053</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0375</td>
<td>0.0436</td>
<td>182.6667</td>
<td>165.5963</td>
<td>0.0054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.1713</td>
<td>0.1777</td>
<td>118.5055</td>
<td>115.9257</td>
<td>0.0027</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1744</td>
<td>0.1731</td>
<td>117.0872</td>
<td>116.7533</td>
<td>0.0026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.3034</td>
<td>0.3097</td>
<td>110.4483</td>
<td>109.1379</td>
<td>0.0025</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.3037</td>
<td>0.3049</td>
<td>109.3184</td>
<td>109.3473</td>
<td>0.0025</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tables IV.1, IV.2, IV.3 present the theoretical and empirical MSEs and PREs of the ORRT mean estimators when all the variances of measurement errors ($\sigma^2_v$, $\sigma^2_u$ and $\sigma^2_p$) are set equal to 1 and response rate in Phase I is set equal to 40% with different variances of S ($0.2*\sigma^2_x$, $0.5*\sigma^2_x$, $1*\sigma^2_x$), respectively. Comparing these three tables,
the mean estimation is less efficient as the variance of S increases in the presence of non-response. For example, under the situation when variance of T is equal to 0.5, the sensitivity level W is equal to 0.8, and in the presence of measurement errors, the MSEs of the generalized mean estimator are respectively equal to 0.1350, 0.1397 and 0.1774 for the variance of S equal to \(0.2\sigma^2_x\), \(0.5\sigma^2_x\), and \(\sigma^2_x\). These results are consistent with the theoretical results. Larger variance of S introduces larger penalty for using RRT models.

For all three tables, the MSE of the mean estimators increases as W increases under non-response and measurement errors. For example, in Table IV.1, the MSE of the generalized mean estimator increased from 0.0937 to 0.1622 as the sensitivity level increased from 0.5 to 1 when variance of T is equal to 0.5. It indicates that the ORRT model gains some efficiency when some of the respondents feel the survey question is not sensitive. Furthermore, as the variance of T increases, the MSE increases while \(\delta\) decreases with a reasonably small value of \(\sigma^2_T\). For instance, in Table IV.1, when the sensitivity level W is equal to 0.5, the MSE of the generalized mean estimator increases from 0.0333 to 0.1690 as the variance of T increases from 0 to 1, while the \(\delta\) value decreases from 0.0041 to 0.0014. Similar to Chapter III, mean estimators under the general linear combination model \((Z=TY+S)\) is better than under the simple additive model \((Z=Y+S)\) when non-response is present if both efficiency and privacy are considered through the unified measures.
Table IV.4. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under the Conditions of $\sigma^2_\nu = \sigma^2_u = \sigma^2_p = 1, 5, 10$ when Response Rate $= 40\%$, and $\sigma^2_s = 0.5*\sigma^2_x$.

<table>
<thead>
<tr>
<th>Est.</th>
<th>$f$</th>
<th>MSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>$\hat{\mu}_{yw}$</td>
<td>2</td>
<td>0.1680</td>
<td>0.1799</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1679</td>
<td>0.1799</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.2470</td>
<td>0.2637</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2397</td>
<td>0.2653</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.3261</td>
<td>0.3476</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3173</td>
<td>0.3484</td>
</tr>
<tr>
<td>$\hat{\mu}_{rw}$</td>
<td>2</td>
<td>0.1514</td>
<td>0.1960</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1498</td>
<td>0.1894</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.2236</td>
<td>0.2859</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2223</td>
<td>0.2882</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.2957</td>
<td>0.3758</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2854</td>
<td>0.3761</td>
</tr>
<tr>
<td>$\hat{\mu}_{pw}$</td>
<td>2</td>
<td>0.1397</td>
<td>0.1601</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1392</td>
<td>0.1585</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.2071</td>
<td>0.2357</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1972</td>
<td>0.2353</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.2745</td>
<td>0.3113</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2653</td>
<td>0.3114</td>
</tr>
</tbody>
</table>

Table IV.4 presents the theoretical and empirical MSEs and PREs of the ORRT mean estimators under different variances of measurement errors (1, 5, 10) when the sensitivity level $W$ is 0.8, variance of $T$ is 0.5 and response rate in Phase I is 40%. As the measurement errors increase, the MSE of each mean estimator increases. For instance, the MSE of the generalized mean estimator increased from 0.1297 to 0.1894 as the variance of measurement errors increased from 1 to 10 when the value of $f$ is
2. It is obvious that larger measurement errors have larger negative impact on mean estimation under non-response.

Also, from Tables IV.1, IV.2, IV.3 and Table IV.4, it is clear that the generalized mean estimator $\hat{\mu}_{pw}$ is more efficient than the other two mean estimators even when very large measurement errors are present. However, the ratio estimator $\hat{\mu}_{rw}$ becomes less efficient than the ordinary mean estimator $\hat{\mu}_{yw}$ as the measurement errors increase. For example in Table IV.4, the MSE of the generalized mean estimator 0.1804 is less than the MSE of the ordinary mean estimator 0.1948 when the variance of measurement errors is 10. However, the MSE of the ratio estimator 0.2518 is larger than other two estimators because the measurement errors are large. The reason is the measurement errors take place on both X and Y for the ratio estimator and only on Y for the ordinary mean estimator. This result shows the superiority of the generalized mean estimator in the presence of measurement errors and non-response because it is not affected as badly as the ratio estimator by measurement errors on X.

Table IV.5 presents the theoretical and empirical MSEs and PREs of the ORRT mean estimators under different response rates when the variance of measurement errors is equal to 1, sensitivity level W is equal to 0.8, and variance of T is equal to 0.5. The efficiency of each estimator gets better as the response rate increases. In other words, the larger the sample we collect from the first call, the higher is the efficiency of the mean estimation.

In addition, from both Tables IV.4 and IV.5, the efficiency of each estimator gets worse as the value of $f$ increases. For example, the MSE of the generalized mean estimator increased from 0.1601 to 0.3113 as the value of $f$ increased from 2 to 4 when the variance of measurement errors is 5. It is reasonable because larger $f$ value means we obtain smaller sample from the second call.
Table IV.5. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under the Conditions of Response Rate (RR) = 20%, 40%, 60% when $\sigma_v^2 = \sigma_u^2 = \sigma_p^2 = 1$, $W = 0.8$, and $\sigma_T^2 = 0.5\sigma_x^2$

<table>
<thead>
<tr>
<th>RR. f</th>
<th>$\hat{\mu}_{yw}^{HH}$</th>
<th>$\hat{\mu}_{rw}^{HH}$</th>
<th>$\hat{\mu}_{pw}^{HH}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>PRE</td>
<td>MSE</td>
</tr>
<tr>
<td>2</td>
<td>0.2134</td>
<td>0.1680</td>
<td>0.1231</td>
</tr>
<tr>
<td>3</td>
<td>0.3184</td>
<td>0.2470</td>
<td>0.1764</td>
</tr>
<tr>
<td>4</td>
<td>0.4234</td>
<td>0.3261</td>
<td>0.2298</td>
</tr>
<tr>
<td></td>
<td>0.1951</td>
<td>0.1514</td>
<td>0.1086</td>
</tr>
<tr>
<td>2</td>
<td>0.2914</td>
<td>0.2236</td>
<td>0.1570</td>
</tr>
<tr>
<td>3</td>
<td>0.3878</td>
<td>0.2957</td>
<td>0.2055</td>
</tr>
<tr>
<td>4</td>
<td>0.1818</td>
<td>0.1397</td>
<td>0.0983</td>
</tr>
<tr>
<td>3</td>
<td>0.2718</td>
<td>0.2071</td>
<td>0.1436</td>
</tr>
<tr>
<td>4</td>
<td>0.3618</td>
<td>0.2745</td>
<td>0.1888</td>
</tr>
</tbody>
</table>

IV.5. Concluding Chapter Remarks

The main contribution in this chapter is the mean estimation of a sensitive variable, in the presence of measurement errors and non-response, using modified HH two-phase sampling. ORRT model leads to better results than non-optional RRT model under the presence of non-response and measurement errors simultaneously. Measurement errors have a negative impact on mean estimation, especially when they
are large. A simulation study verifies the theoretical results. It is also clear from the theoretical conditions (IV.38), (IV.40), (IV.42) and the simulation results that the generalized mean estimator is always more efficient than the ordinary RRT mean estimator and the ratio estimator, while the ratio estimator is less efficient than the ordinary mean estimator if the measurement errors on X are large.

We have used SRS in the current study. We want to explore the performance of mean estimation under the same conditions as in this Chapter, but using stratified random sampling. The mean estimation of a sensitive variable under both measurement errors and non-response using stratified random sampling will be discussed in the next chapter.
CHAPTER V
MEAN ESTIMATION IN THE SIMULTANEOUS PRESENCE OF MEASUREMENT ERRORS AND NON-RESPONSE USING ORRT MODELS UNDER THE STRATIFIED RANDOM SAMPLING DESIGN

We have so far discussed mean estimation under measurement errors, and/or non-response using simple random sampling[92]. In this Chapter, we will continue the Chapter IV work but using stratified random sampling. In Section V.1, some existing mean estimators under measurement errors and non-response will be presented; in Section V.2, the generalized mean estimator will be discussed; Section V.3 will present a simulation study; and Section V.4 will provide concluding remarks for this Chapter.

V.1. Some Existing Mean Estimators under Measurement Errors and Non-response using Stratified Random Sampling

Let a finite population $U = (U_1, U_2, U_3, ..., U_N)$ is divided in $L$ homogeneous strata with $N_h$ representing the number of units in stratum $h$ such that $\sum_{h=1}^{L} N_h = N$. From $h^{th}$ stratum, a simple random sample of size $n_h$ is drawn without replacement such that $\sum_{h=1}^{L} n_h = n$. Under the situation where non-response is present, we assume that $n_{1h}$ units provided response on the first call and remaining $n_{2h} = n_h - n_{1h}$ units do not respond. Then a sub-sample of size $n_{sh} = \frac{n_{2h}}{f_h}$ ($f_h > 1$) is taken from the $n_{2h}$ non-response units in the $h^{th}$ stratum. We use standard terminology, as used in Chapter IV, but with term '$h$' ($h = 1, 2, ..., h$) in the subscript to represent terms in the $h^{th}$ stratum. In the $h^{th}$ stratum, let the measurement errors of the auxiliary variable ($X$) be given by $V_{ih} = x_{ih} - X_{ih}$. Let the respective measurement errors
associated with the study variable \((Y)\) in the population and the scrambled variable \((Z)\) in the face-to-face phase be given by 

\[ U_{ih} = y_{ih} - Y_{ih} \quad \text{and} \quad P_{ih} = z_{ih} - Z_{ih}. \]

These measurement errors are assumed to be random and uncorrelated with mean zero and variances \(\sigma^2_{vih}, \sigma^2_{uhi}, \text{and } \sigma^2_{ph}, \) respectively.

Assume population mean of the auxiliary variable \(\mu_x\) is known, and non-response happens on both \(X\) and \(Y\). Some notations are given below

\[ \Omega_y = \sum_{h=1}^{L} \sum_{i=1}^{n_h} (y_{ih} - \mu_{yih}), \quad (V.1) \]

\[ \Omega_x = \sum_{h=1}^{L} \sum_{i=1}^{n_h} (x_{ih} - \mu_{xih}), \quad (V.2) \]

\[ \Omega_u = \sum_{h=1}^{L} (\sum_{i=1}^{n_{1h}} U_{ih} + \sum_{i=1}^{n_{2h}} P_{ih}), \quad (V.3) \]

and

\[ \Omega_v = \sum_{h=1}^{L} \sum_{i=1}^{n_h} V_{ih}. \quad (V.4) \]

Let \(e_0^* = \frac{1}{n_{\mu_y}} (\Omega_y + \Omega_u)\) and \(e_1^* = \frac{1}{n_{\mu_x}} (\Omega_x + \Omega_v)\). In other words, \(\hat{y}^{*st} = (1 + e_0^*) \mu_y\) and \(\bar{x}^{*st} = (1 + e_1^*) \mu_x\), where \(\hat{y}^{*st} = \sum_{h=1}^{L} \pi_h (w_{1h} \hat{y}_{1h} + w_{2h} \hat{y}_{2h})\) and \(\bar{x}^{*st} = \sum_{h=1}^{L} \pi_h (w_{1h} \bar{x}_{1h} + w_{2h} \bar{x}_{2h})\) in the presence of measurement errors, and \(\pi_h = N_h / N\).

Under the assumption of bivariate normality (Sukhatme et al.1970)[78]:

\[ E(e_h^*) = E(e_1^*) = 0; \]

\[ E(e_h^*) = E(e_1^*) = 0; \quad (V.5) \]
\[ E(e_0^2) = \frac{1}{\mu_y^2} \sum_{h=1}^{L} \pi_h^2 \left[ \sigma_{yh}^2 + \sigma_{ah}^2 \right] + \lambda_h \left( \sigma_{y(2)h}^2 + \mu_y^2 \right) + \frac{W_{2h} f_h}{n_h} \left( \frac{\sigma_{yh}^2 W_h + \sigma_{2h}^2 W_h(\sigma_{y(2)h}^2 + \mu_y^2 + \mu_{y(2)h}^2)}{\left( \mu_{y(h)} W_h + 1 + W_h \right)^2} \right], \] (V.6)

\[ E(e_1^2) = \frac{1}{\mu_x^2} \sum_{h=1}^{L} \pi_h^2 \left[ \theta_h \left( \sigma_{xh}^2 + \sigma_{vh}^2 \right) + \lambda_h \left( \sigma_{x(2)h}^2 + \sigma_{v(2)h}^2 \right) \right], \] (V.7)

and

\[ E(e_0^2 e_1^2) = \sum_{h=1}^{L} \pi_h^2 \left[ \theta_h \rho_{yxh} \frac{\sigma_{yh}^2 \sigma_{xh}^2}{\mu_y \mu_x} + \lambda_h \rho_{x(2)h} \frac{\sigma_{xh}^2 \sigma_{x(2)h}}{\mu_x \mu_x} \right], \] (V.8)

where \( \theta_h = \frac{N_y}{N_h n_h} \), \( \lambda_h = \frac{N_{2h}(f_h-1)}{N_h n_h} \), \( W_{2h} = \frac{N_{2h}}{N_h} \), and

\[ \rho_{xxh} = \frac{\sigma_{y(2)h}}{\sqrt{1 + \frac{\sigma_{yh}^2 W_h + \sigma_{2h}^2 W_h(\sigma_{y(2)h}^2 + \mu_y^2 + \mu_{y(2)h}^2)}{\sigma_{y(2)h}^2}}}. \] (V.9)

The two existing mean estimators in the presence of measurement errors using the modified HH two-phase method under stratified random sampling are given by:

- The ordinary mean estimator is given by

\[ \hat{\mu}_{yw} = \sum_{h=1}^{L} \pi_h \hat{y}_h = \sum_{h=1}^{L} \pi_h \left( w_{1h} \hat{y}_{1h} + w_{2h} \hat{y}_{2h} \right), \] (V.10)

where \( \pi_h = \frac{N_h}{N} \). It can be written as

\[ \hat{\mu}_{yw} = (1 + e_0^*) \mu_y. \] (V.11)

The difference between the ordinary mean estimator and the true mean can be written as

\[ \hat{\mu}_{yw} - \mu_y = e_0^* \mu_y. \] (V.12)
Taking square and then expected value on both side of (V.12), the MSE of $\hat{\mu}_{yw}^st$ is given by

$$\text{MSE}^*(\hat{\mu}_{yw}^st) = \sum_{h=1}^{L} \pi_h^2 [\theta_h (\sigma_{y_h}^2 + \sigma_{u_h}^2) + \lambda_h (\sigma_{y_{(2)h}}^2 + \sigma_{p_h}^2) + G_h],$$  \hspace{1cm} (V.13)

where $G_h = \frac{W_{2h} f_h}{n_h} \left[ \frac{\sigma_{y_h}^2 W_h + \frac{1}{2} \sigma_{y_{(2)h}}^2 W_h + \sigma_{p_h}^2 + \mu_{y_{(2)h}}^2}{(\mu_{y_h} W_h + 1 - W_h)^2} \right].$

- A ratio estimator proposed by Gupta et al. (2014) is given by

$$\hat{\mu}_{rw}^st = \frac{\hat{y}^st}{\bar{x}^st} \mu_x = \hat{R}_{rw}^st \mu_x.$$  \hspace{1cm} (V.14)

It can be written as

$$\hat{\mu}_{rw}^st = \frac{(1 + e_0^*) \mu_y}{(1 + e_1^*) \mu_x} \mu_x$$

$$= \mu_y (1 + e_0^*) (1 + e_1^*)^{-1}$$

$$= \mu_y (1 - e_1^* + e_1^{*2} + e_0^* - e_0^* e_1^* + ...).$$  \hspace{1cm} (V.15)

Using second order approximation, the difference between the ratio estimator and the true mean can be written as

$$\hat{\mu}_{yw}^st - \mu_y = \mu_y (-e_1^* + e_1^{*2} + e_0^* - e_0^* e_1^*).$$  \hspace{1cm} (V.16)

Taking square and then expected value on both side of (V.16), the MSE of $\hat{\mu}_{rw}^st$ is given by

$$\text{MSE}^*(\hat{\mu}_{rw}^st) = \sum_{h=1}^{L} \pi_h^2 \theta_h (\sigma_{y_h}^2 + \sigma_{u_h}^2 + R^2 (\sigma_{x_h}^2 + \sigma_{v_h}^2)) - 2 \rho_{yxh} \sigma_{y_h} \sigma_{x_h} +$$

$$\lambda_h (\sigma_{y_{(2)h}}^2 + \sigma_{p_h}^2 + R^2 (\sigma_{x_{(2)h}}^2 + \sigma_{v_{(2)h}}^2)) - 2 \rho_{zxh} (2) \sigma_{y_{(2)h}} \sigma_{x_{(2)h}} + G_h],$$  \hspace{1cm} (V.17)

where $R = \mu_y / \mu_x.$
The MSE of \( \hat{\mu}_{yw}^{st} \) and \( \hat{\mu}_{rw}^{st} \) without measurement error, may be obtained by putting \( \sigma_{vh}^2 = \sigma_{uh}^2 = \sigma_{ph}^2 = 0 \) in the equations (V.13)(V.17).

V.2. Generalized Estimator in the Presence of Measurement Errors and Non-response using ORRT Models under the Stratified Random Sampling Design

With this background, we use the generalized mean estimator used in Khalil et al. (2018) [34] and previous two Chapters. This estimator includes a wide variety of mean estimators as special cases. The non-response version under stratified random sampling is given by:

\[
\hat{\mu}_{pw}^{st} = \left( \hat{\bar{y}}^{*st} + k(\mu_x - \bar{x}^{*st}) \right) \left( \bar{D}^{st} \right)^v,
\]

where

\[
\hat{\bar{y}}^{*st} = \sum_{h=1}^{L} \pi_h (w_{1h} \hat{y}_{1h} + w_{2h} \hat{y}_{2h})
\]

is the ordinary mean estimator in (V.10),

\[
\hat{x}^{*st} = \sum_{h=1}^{L} \pi_h (w_{1h} \bar{x}_{1h} + w_{2h} \bar{x}_{2h}), \quad \bar{d} = \phi(\alpha^{st} \bar{x} + \beta^{st}) + (1 - \phi)(\alpha^{st} \mu_x + \beta^{st}), \quad \bar{D} = \alpha^{st} \mu_x + \beta^{st},
\]

\( k \) and \( v \) are suitable constants. \( \phi \) is assumed to be an unknown constant whose value is to be determined from optimally considerations. \( \alpha^{st} (\alpha^{st} \neq 0) \) and \( \beta^{st} \) are assumed to be some known parameters of the auxiliary variable \( X \), such as coefficient of variation \( (C_x) \), kurtosis, and correlation coefficient \( (\rho_{yx}) \) etc. Please note here with different values of \( \alpha^{st} \) and \( \beta^{st} \), we can obtain various estimators. Also, with \( v=1 \) we get various ratio estimators and with \( v=-1 \) we get various product estimators.

V.2.1 Bias and MSE of the Generalized Mean Estimator

The generalized mean estimator will be studied under both modified HH two-phase sampling and measurement errors using stratified random sampling. According
to the notations in Section V.1, it can be written as

\[ \hat{\mu}_{pw}^{st} = \left( (1 + e_0^*)\mu_y + k(\mu_x - (1 + e_1^*)\mu_x) \right) \frac{\alpha^{st}\mu_x + \beta^{st}}{\phi(\alpha^{st}(1 + e_1^*)\mu_x + \beta^{st}) + (1 - \phi)(\alpha^{st}\mu_x + \beta^{st})^2}. \]  

(V.19)

Using Taylor’s approximation and retaining terms of order up to 2, the difference between the generalized mean estimator and the true mean can be written as

\[ \hat{\mu}_{pw}^{st} - \mu_y = \left( (1 + e_0^*)\mu_y + k(\mu_x - (1 + e_1^*)\mu_x) \right) \frac{\alpha^{st}\mu_x + \beta^{st}}{\phi(\alpha^{st}(1 + e_1^*)\mu_x + \beta^{st}) + (1 - \phi)(\alpha^{st}\mu_x + \beta^{st})^2} \]

\[ - \mu_y = e_0^*\mu_y + e_1^*[-k\mu_x + \mu_y(\frac{-\alpha^{st}\phi\mu_x}{\alpha^{st}\mu_x + \beta^{st}})] + \frac{1}{2!}[2e_0^*e_1^*\mu_y(\frac{-\alpha^{st}\phi\mu_x}{\alpha^{st}\mu_x + \beta^{st}}) + e_1^*2(vk\mu_x(\frac{\alpha^{st}\phi\mu_x}{\alpha^{st}\mu_x + \beta^{st}}) + v(v + 1)\mu_y(\frac{-\alpha^{st}\phi\mu_x}{\alpha^{st}\mu_x + \beta^{st}})^2)]. \]

(V.20)

Taking expectation on both side of (V.20), the bias of the generalized mean estimator \( \hat{\mu}_{pw}^{st} \), correct to second order of approximation, is given by

\[ \text{Bias}(\hat{\mu}_{pw}^{st}) = \sum_{h=1}^{L} \pi_h^2 \{ \theta_h [(kH^{st} + \frac{v + 1}{v}\mu_yH^{st2})(\sigma_{xh}^2 + \sigma_{vh}^2) - H^{st}\rho_{xyh}\sigma_{yh}\sigma_{xh}] + \lambda_h [(kH^{st} + \frac{v + 1}{v}\mu_yH^{st2})(\sigma_{(x)h}^2 + \sigma_{v(h)}^2) - H^{st}\mu_y\rho_{zz(2)h}\sigma_{z(h)\sigma_{x(2)h}}] \}, \]

(V.21)

where \( H^{st} = \frac{\alpha^{st}\phi v}{\alpha^{st}\mu_x + \beta^{st}}. \)

The bias of \( \hat{\mu}_{pw} \) without measurement error may be obtained by setting \( \sigma_v^2 = 0 \) in equation (V.21).
To determine the expression for MSE of the generalized mean estimator (V.18), we take square of (V.20) on both sides and retain terms of order up to 2 which is given by

\[(\hat{\mu}_{st}^{pw} - \mu_y)^2 = e_0^2 \mu_y^2 + k^2 \mu_x e_1^2 + (H_{st}^{st} \mu_x \mu_y e_1^2)^2 - 2e_0^* e_1^* k \mu_x \mu_y - 2e_1^* H_{st}^{st} \mu_x \mu_y^2 + 2e_1^2 k H_{st}^{st} \mu_x \mu_y.\]  

(V.22)

Taking the expected value on both side of (V.22), the expression for MSE of \(\hat{\mu}_{st}^{pw}\), correct to the first order approximation is given by

\[\text{MSE}^* (\hat{\mu}_{st}^{pw}) = E(\hat{\mu}_{st}^{pw} - \mu_y)^2 \]

\[\simeq \sum_{h=1}^{L} \pi_h^2 \left[ \theta_h (\sigma_{y|h}^2 + k^2 \sigma_{x|h}^2 + \phi^2 v^2 R_{st|pw}^{st} \sigma_{x|h}^2 + 2k \phi v R_{st|pw}^{st} \sigma_{x|h}^2 - 2k \rho_{yx|h} \sigma_{x|h} \sigma_{y|h} + 2k \phi v R_{st|pw}^{st} \sigma_{x|h}^2 - 2k \rho_{yx|h} \sigma_{x|h} \sigma_{y|h}) + 2k \phi v R_{st|pw}^{st} \sigma_{x|h}^2 - 2k \rho_{yx|h} \sigma_{x|h} \sigma_{y|h} - 2\phi v R_{st|pw}^{st} \mu_y \rho_{yx|h} \sigma_{x|h} \sigma_{y|h}) + 2k \phi v R_{st|pw}^{st} \sigma_{x|h}^2 - 2k \rho_{yx|h} \sigma_{x|h} \sigma_{y|h} + 2k \phi v R_{st|pw}^{st} \sigma_{x|h}^2) + G_h \right] \]

\[= \sum_{h=1}^{L} \pi_h^2 \left[ \theta_h (\sigma_{y|h}^2 + \phi^2 v^2 R_{st|pw}^{st} \sigma_{x|h}^2 + 2k \phi v R_{st|pw}^{st} \sigma_{x|h}^2) + G_h \right] \]

\[= \sum_{h=1}^{L} \pi_h^2 \left[ \theta_h (\sigma_{y|h}^2 + (k + \phi v R_{st|pw}^{st})^2 \sigma_{x|h}^2 - 2(k + \phi v R_{st|pw}^{st}) \rho_{yx|h} \sigma_{x|h} \sigma_{y|h}) + \lambda_h (\sigma_{x|h}^2 + \phi^2 v^2 R_{st|pw}^{st} \sigma_{x|h}^2 + 2k \phi v R_{st|pw}^{st} \sigma_{x|h}^2) + G_h, \right] \]

(V.23)

where \(R_{st|pw}^{st} = \frac{\alpha^* \mu_x}{\alpha^* \mu_x + \beta^*}.\)
Minimization of the above expression (V.23) with respect to \( \phi \) yields its optimum value as:

\[
\phi_{opt} \approx \frac{\sum_{h=1}^{L} \pi_h^2 \left[ \theta_h (\rho_{yxh} \sigma_{xh} \sigma_{yh} - k(\sigma_{xh}^2 + \sigma_{vh}^2)) + \lambda_h (\mu_y \rho_{zxh} \sigma_{zh} \sigma_{z(2)h} - k(\sigma_{z(2)h}^2 + \sigma_{vh}^2)) \right]}{\nu R_{pw}^2 \sum_{h=1}^{L} \pi_h^2 \left[ \theta_h (\sigma_{xh}^2 + \sigma_{vh}^2) + \lambda_h (\sigma_{z(2)h}^2 + \sigma_{vh}^2) \right]}
\]  

(V.24)

Substitution of \( \phi_{opt} \) in MSE(\( \hat{\mu}_{pw} \)) yields the minimum value as:

\[
MSE_{min}^*(\hat{\mu}_{pw}) \approx \sum_{h=1}^{L} \pi_h^2 \left[ \theta_h (\sigma_{y(2)h}^2 + P_{stx(2)h}^2 \sigma_{xh}^2 - 2P_{stx(2)h} \rho_{zxh} \sigma_{zh} \sigma_{z(2)h}) + \lambda_h (\sigma_{y(2)h}^2 + P_{stx(2)h}^2 \sigma_{z(2)h}^2 + \sigma_{ph}^2 + P_{stx(2)h}^2 \sigma_{ph}^2) + G_h \right],
\]

where \( P_{st} = \sum_{h=1}^{L} \pi_h^2 \theta_h \rho_{yxh} \sigma_{xh} \sigma_{yh} + \lambda_h \rho_{zxh} \sigma_{zh} \sigma_{z(2)h} \).

The expression for the minimized MSE of generalized estimator without measurement errors may be obtained by putting \( \sigma_{uh}^2 = \sigma_{vh}^2 = \sigma_{ph}^2 = 0 \) in the above expression, which gives

\[
MSE_{min}^*(\hat{\mu}_{pw}) \approx \sum_{h=1}^{L} \pi_h^2 \left[ \theta_h (\sigma_{y(2)h}^2 + P_{stx(2)h}^2 \sigma_{xh}^2 - 2P_{stx(2)h} \rho_{zxh} \sigma_{zh} \sigma_{z(2)h}) + \lambda_h (\sigma_{y(2)h}^2 + P_{stx(2)h}^2 \sigma_{z(2)h}^2 + \sigma_{ph}^2 + P_{stx(2)h}^2 \sigma_{ph}^2) + G_h \right],
\]

(V.26)

where \( G_h = \frac{W_{2h} f_h}{\mu_{pw} h} \left[ \frac{\sigma_{y(2)h}^2 + \sigma_{z(2)h}^2 + \mu_{y(2)h}^2}{\mu_{pw} h + 1 - W_h} \right] \).

**V.2.2 Efficiency Comparisons**

Comparing MSE expressions of \( \hat{\mu}_{yw}^* \) (V.13), \( \hat{\mu}_{rw}^* \) (V.17), and \( \hat{\mu}_{pw}^* \) (V.25) with measurement errors, it can be verified that
• \(MSE^*_\text{min}(\hat{\mu}_{pw}^*) < MSE^*(\hat{\mu}_{yw}^*)\) if

\[
MSE^*_\text{min}(\hat{\mu}_{pw}^*) - MSE^*(\hat{\mu}_{yw}^*)
= -\sum_{h=1}^{L} \pi_h^2 \left( \frac{\theta_h \rho_{yxh} \sigma_{xh} \sigma_{yh} + \lambda_h \rho_{zz(x(2)h)} \sigma_{zh} \sigma_{x(2)h}}{\theta_h (\sigma_{xh}^2 + \sigma_{vh}^2) + \lambda_h (\sigma_{x(2)h}^2 + \sigma_{vh}^2)} \right)^2 < 0, \tag{V.27}
\]

• \(MSE^*_\text{min}(\hat{\mu}_{pw}^*) < MSE^*(\hat{\mu}_{rw}^*)\) if

\[
MSE^*_\text{min}(\hat{\mu}_{pw}^*) - MSE^*(\hat{\mu}_{rw}^*)
= \sum_{h=1}^{L} \pi_h^2 \left[ (P_{st}^2 - R^2) (\theta_h (\sigma_{xh}^2 + \sigma_{vh}^2) + \lambda_h (\sigma_{x(2)h}^2 + \sigma_{vh}^2)) \right]
- 2(P_{st} + R) (\theta_h \rho_{yxh} \sigma_{xh} \sigma_{yh} + \lambda \rho_{zz(x(2)h)} \sigma_{zh} \sigma_{x(2)h})^2 < 0; \tag{V.28}
\]

In other words, if

\[
\sum_{h=1}^{L} \pi_h^2 \left( P_{st}^2 - R^2 \right) (\theta_h (\sigma_{xh}^2 + \sigma_{vh}^2) + \lambda_h (\sigma_{x(2)h}^2 + \sigma_{vh}^2))
- 2(P_{st} + R) (\theta_h \rho_{yxh} \sigma_{xh} \sigma_{yh} + \lambda \rho_{zz(x(2)h)} \sigma_{zh} \sigma_{x(2)h})^2 < 0; \tag{V.29}
\]

and

\[
\frac{1}{2} - \frac{R}{2P_{st}}
= \frac{1}{2} - \frac{\mu_y}{2\mu_x} \sum_{h=1}^{L} \pi_h^2 \theta_h (\sigma_{xh}^2 + \sigma_{vh}^2) + \lambda_h (\sigma_{x(2)h}^2 + \sigma_{vh}^2) < 1,
\]

and
\[ \text{MSE}^*_{\text{min}}(\hat{\mu}_{rw}^{st}) < \text{MSE}^*(\hat{\mu}_{yw}^{st}) \] if
\[ \text{MSE}^*(\hat{\mu}_{rw}^{st}) - \text{MSE}^*(\hat{\mu}_{yw}^{st}) = \sum_{h=1}^{L} \pi_h^2 [R^2(\theta_h(\sigma_{xh}^2 + \sigma_{vh}^2) + \lambda_h(\sigma_{x(2)h}^2 + \sigma_{vh}^2)) - 2R(\theta_h \rho_{yxh} \sigma_{xh} \sigma_{yh} + \lambda_h \rho_{zxx(2)h} \sigma_{zh} \sigma_{x(2)h})] < 0; \]

\[ (V.30) \]

In other words, if
\[ \sum_{h=1}^{L} \pi_h^2 \frac{R^2\theta_h(\sigma_{xh}^2 + \sigma_{vh}^2) + \lambda_h(\sigma_{x(2)h}^2 + \sigma_{vh}^2)}{2R\theta_h \rho_{yxh} \sigma_{xh} \sigma_{yh} + \lambda_h \rho_{zxx(2)h} \sigma_{zh} \sigma_{x(2)h}} = \frac{\mu_y}{2\mu_x} \sum_{h=1}^{L} \pi_h^2 \theta(\sigma_{xh}^2 + \sigma_{vh}^2) + \lambda_h(\sigma_{x(2)h}^2 + \sigma_{vh}^2) < 1. \]

\[ (V.31) \]

The conditions (V.27) and (V.29) always hold true. From (V.31), the ratio estimator is more efficient than the ordinary mean estimator only if the measurement error on auxiliary variable X (\(\sigma_{xh}^2\)) is small and X and Y are highly correlated in each stratum.

**V.3. Simulation Study**

We will evaluate the performance of the generalized mean estimator under non-response and measurement errors using stratified random sampling with the other two estimators by a simulation study in this section. In the generalized mean estimator, we choose \(v\) and \(k\) to be 1, \(\alpha = 1\), \(\beta = 0\), and \(\phi\) to be its optimum value. The scrambling variable \(S\) is taken to be a normal variate with mean equal to zero and vary variance (0.2\(\sigma_x^2\), 0.5\(\sigma_x^2\), \(\sigma_x^2\)). And \(T\) is also taken to be a normal variate but with mean equal to one and varying variances (0, 0.5, 1). The measurement errors of X have a normal distribution with mean zero in both phases; the measurement
errors of \( Y \) in the first phase and \( Z \) in the second phase have a normal distribution with mean zero. We use different variances for measurement errors.

We consider three bivariate normal distributions with different covariance matrices to represent the distribution of \( Y \) and \( X \) in three strata. We assume each stratum with size of 2000 and take a sample of size 200 using SRSWOR from each stratum. We assume the first phase response rate in each stratum is 40%. This means in the first phase only 80 \( (n_1) \) subjects provide a response to the survey question and 120 \( (n_2) \) of them do not. In the second phase, we take another sample \( (n_s = \frac{n_2}{f}) \) from non-respondent group by using \( f = 2, 3, 4 \), respectively. Different response rates of 20\%, 40\% and 60\% also compared in the simulation study. The three strata have covariance matrices \( \Sigma \) as given below:

\[
\text{Stratum 1} \quad \mu = \begin{bmatrix} 10 \\ 6 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 16 & 7.8384 \\ 7.8384 & 6 \end{bmatrix}, \quad \rho_{yx} = 0.8
\]

\[
\text{Stratum 2} \quad \mu = \begin{bmatrix} 9 \\ 5 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 16 & 6.2610 \\ 6.2610 & 5 \end{bmatrix}, \quad \rho_{yx} = 0.7
\]

\[
\text{Stratum 3} \quad \mu = \begin{bmatrix} 7 \\ 4 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 16 & 6.8586 \\ 6.8586 & 6 \end{bmatrix}, \quad \rho_{yx} = 0.7
\]

The real parameters of the set of 5000 data points we generated using R are very close to the parameter values in (A) but not exactly same. For the simulation study, we used parameter values in (B) and not those in (A).

\[\text{Stratum 1}\]

\[
\mu_x = 6, \; \sigma_x^2 = 6, \; \mu_y = 10, \; \sigma_y^2 = 16, \; \rho_{yx} = 0.8 \quad \text{(A)}
\]

\[
\mu_x = 6.0512, \; \sigma_x^2 = 5.9809, \; \mu_y = 9.9802, \; \sigma_y^2 = 16.0583, \; \rho_{yx} = 0.8129 \quad \text{(B)}
\]
Stratum 2

\[ \mu_x = 5, \sigma_x^2 = 5, \mu_y = 9, \sigma_y^2 = 16, \rho_{yx} = 0.7 \]  \hspace{1cm} (A)

\[ \mu_x = 5.0968, \sigma_x^2 = 5.0155, \mu_y = 9.2189, \sigma_y^2 = 16.2913, \rho_{yx} = 0.6887 \]  \hspace{1cm} (B)

Stratum 3

\[ \mu_x = 4, \sigma_x^2 = 6, \mu_y = 7, \sigma_y^2 = 16, \rho_{yx} = 0.7 \]  \hspace{1cm} (A)

\[ \mu_x = 3.9674, \sigma_x^2 = 5.9354, \mu_y = 6.8833, \sigma_y^2 = 15.5977, \rho_{yx} = 0.7057 \]  \hspace{1cm} (B)

Coding for the simulations was done in R and results are averaged over 5,000 iterations. The empirical MSE of the estimator \( \hat{\mu}_y \) is computed by

\[ \text{MSE}^*(\hat{\mu}_w) = \frac{1}{5000} \sum_{i=1}^{5000} (\hat{\mu}_w - \mu)^2, \]  \hspace{1cm} (V.32)

where \( \hat{\mu}_w = \hat{\mu}_{yw}, \hat{\mu}_{rw}, \hat{\mu}_{pw} \). Here, \( \mu \) is the population mean of the sensitive study variable. The percent relative efficiencies (PREs) of the estimators \( (\hat{\mu}_w) \) with respect to the ordinary mean estimator \( (\hat{\mu}_{yw}) \) is defined as

\[ PRE = \frac{\text{MSE}^*(\hat{\mu}_{yw})}{\text{MSE}^*(\hat{\mu}_w)} \times 100. \]  \hspace{1cm} (V.33)

We will also use the unified measure \( \delta \) of efficiency and privacy as defined in Gupta et al. (2018)[24]. It is given by

\[ \delta = \frac{\text{MSE}^*(\hat{\mu}_w)}{\Delta_{DP}}. \]  \hspace{1cm} (V.34)

In (V.34), MSE is used in place of Var(.) to account for biased estimators. And \( \Delta_{DP} \) is calculated by \( \Delta_{DP} = \sum_{h=1}^{L} \pi_h \Delta_{DP_h} \), where \( \Delta_{DP_h} \) is the privacy level in the \( h^{th} \) stratum.
Table V.1. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under Stratified Random Sampling when Response Rate = 40% , $\sigma_v^2 = \sigma_u^2 = \sigma_p^2 = 1, f = 2$ and $\sigma_s^2 = 0.2*\sigma_x^2$.

<table>
<thead>
<tr>
<th>Est. $\hat{\mu}_{yw}$</th>
<th>W</th>
<th>$\sigma_T^2$</th>
<th>MSE Without ME</th>
<th>MSE With ME</th>
<th>PRE Without ME</th>
<th>PRE With ME</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0408</td>
<td>0.0433</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0371</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0412</td>
<td>0.0437</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0374</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td>0.0864</td>
<td>0.0890</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0910</td>
<td>0.0913</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0016</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td>0.1319</td>
<td>0.1344</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td>0.1259</td>
<td>0.1335</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0415</td>
<td>0.0440</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0377</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.8</td>
<td>0.1142</td>
<td>0.1167</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.8</td>
<td>0.1191</td>
<td>0.1169</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.8</td>
<td>0.1869</td>
<td>0.1894</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.8</td>
<td>0.1859</td>
<td>0.1835</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0016</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.8</td>
<td>0.0420</td>
<td>0.0445</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0381</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0422</td>
<td>0.0432</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0370</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td>0.1325</td>
<td>0.1350</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td>0.1361</td>
<td>0.1337</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td>0.2233</td>
<td>0.2258</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0020</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5</td>
<td>0.2261</td>
<td>0.2230</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0019</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0264</td>
<td>0.0364</td>
<td>154.5455</td>
<td>118.9560</td>
<td>0.0312</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0268</td>
<td>0.0365</td>
<td>153.7313</td>
<td>119.7260</td>
<td>0.0312</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0711</td>
<td>0.0810</td>
<td>121.5190</td>
<td>109.8765</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0693</td>
<td>0.0820</td>
<td>131.3131</td>
<td>111.3415</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.1161</td>
<td>0.1260</td>
<td>113.6090</td>
<td>106.6667</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.1193</td>
<td>0.1219</td>
<td>105.5323</td>
<td>109.5160</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0271</td>
<td>0.0370</td>
<td>153.1365</td>
<td>118.9189</td>
<td>0.0317</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0277</td>
<td>0.0357</td>
<td>151.9856</td>
<td>122.6891</td>
<td>0.0306</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.8</td>
<td>0.0985</td>
<td>0.1084</td>
<td>115.9391</td>
<td>107.6568</td>
<td>0.0019</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.1022</td>
<td>0.1103</td>
<td>116.5362</td>
<td>105.9837</td>
<td>0.0019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1706</td>
<td>0.1806</td>
<td>109.5545</td>
<td>104.8726</td>
<td>0.0016</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1691</td>
<td>0.1723</td>
<td>109.9349</td>
<td>106.5003</td>
<td>0.0015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.0275</td>
<td>0.0375</td>
<td>152.7273</td>
<td>118.6667</td>
<td>0.0321</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0283</td>
<td>0.0369</td>
<td>149.1166</td>
<td>117.0732</td>
<td>0.0316</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1166</td>
<td>0.1266</td>
<td>113.6364</td>
<td>106.6351</td>
<td>0.0022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.1219</td>
<td>0.1303</td>
<td>111.6489</td>
<td>102.6094</td>
<td>0.0023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.2069</td>
<td>0.2168</td>
<td>107.9265</td>
<td>104.1513</td>
<td>0.0019</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1999</td>
<td>0.2122</td>
<td>113.1066</td>
<td>105.0895</td>
<td>0.0019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.0214</td>
<td>0.0269</td>
<td>190.6542</td>
<td>160.9665</td>
<td>0.0230</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0202</td>
<td>0.0259</td>
<td>203.9604</td>
<td>168.7259</td>
<td>0.0222</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.0665</td>
<td>0.0720</td>
<td>129.9248</td>
<td>123.6111</td>
<td>0.0012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.0668</td>
<td>0.0730</td>
<td>136.2275</td>
<td>125.0685</td>
<td>0.0013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1117</td>
<td>0.1173</td>
<td>118.0842</td>
<td>114.5780</td>
<td>0.0010</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1092</td>
<td>0.1147</td>
<td>115.2930</td>
<td>116.3906</td>
<td>0.0010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.0221</td>
<td>0.0275</td>
<td>187.7828</td>
<td>160.0000</td>
<td>0.0235</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0224</td>
<td>0.0275</td>
<td>187.9464</td>
<td>159.2727</td>
<td>0.0235</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.0940</td>
<td>0.0996</td>
<td>121.4894</td>
<td>117.1687</td>
<td>0.0017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.0982</td>
<td>0.1009</td>
<td>121.2831</td>
<td>115.8573</td>
<td>0.0017</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1664</td>
<td>0.1720</td>
<td>112.3197</td>
<td>110.1163</td>
<td>0.0015</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1633</td>
<td>0.1651</td>
<td>113.8396</td>
<td>111.1448</td>
<td>0.0014</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.0225</td>
<td>0.0280</td>
<td>186.6667</td>
<td>158.9286</td>
<td>0.0240</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0229</td>
<td>0.0276</td>
<td>184.2795</td>
<td>156.5217</td>
<td>0.0236</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1123</td>
<td>0.1178</td>
<td>117.9875</td>
<td>114.6010</td>
<td>0.0020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.1174</td>
<td>0.1148</td>
<td>115.9284</td>
<td>116.4634</td>
<td>0.0020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1927</td>
<td>0.2083</td>
<td>115.8796</td>
<td>108.4013</td>
<td>0.0018</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.1929</td>
<td>0.2043</td>
<td>117.2110</td>
<td>109.1532</td>
<td>0.0018</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table V.2. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under Stratified Random Sampling when Response Rate = 40% , \( \sigma_v^2 = \sigma_u^2 = \sigma_p^2 = 1 \), \( f = 2 \) and \( \sigma_s^2 = 0.5 * \sigma_x^2 \).

<table>
<thead>
<tr>
<th>Est. ( \hat{\mu}_{st} )</th>
<th>( W )</th>
<th>( \sigma_T^2 )</th>
<th>MSE Without ME</th>
<th>MSE With ME</th>
<th>PRE Without ME</th>
<th>PRE With ME</th>
<th>( \delta )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.5</td>
<td>0.0425</td>
<td>0.0450</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0158</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0882</td>
<td>0.0907</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1337</td>
<td>0.1362</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0012</td>
<td></td>
</tr>
<tr>
<td>( \hat{\mu}_{yw} )</td>
<td>0.8</td>
<td>0.0443</td>
<td>0.0468</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0165</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.1169</td>
<td>0.1294</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0022</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1896</td>
<td>0.1921</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.0454</td>
<td>0.0479</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0168</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1359</td>
<td>0.1384</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.2268</td>
<td>0.2293</td>
<td>100.0000</td>
<td>100.0000</td>
<td>0.0019</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0281</td>
<td>0.0380</td>
<td>151.2456</td>
<td>118.4211</td>
<td>0.0134</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0728</td>
<td>0.0827</td>
<td>121.1538</td>
<td>109.6735</td>
<td>0.0014</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1178</td>
<td>0.1277</td>
<td>113.4975</td>
<td>106.6562</td>
<td>0.0011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.0297</td>
<td>0.0397</td>
<td>149.1582</td>
<td>117.8841</td>
<td>0.0140</td>
<td></td>
</tr>
<tr>
<td>( \hat{\mu}_{rw} )</td>
<td>0.8</td>
<td>0.1012</td>
<td>0.1111</td>
<td>115.5138</td>
<td>116.4716</td>
<td>0.0019</td>
<td></td>
</tr>
</tbody>
</table>

101
<table>
<thead>
<tr>
<th></th>
<th>( \hat{\mu}_{pw} )</th>
<th>0.5</th>
<th>0.5</th>
<th>0.5</th>
<th>0.5</th>
<th>0.5</th>
<th>0.5</th>
<th>0.5</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\mu}_{pw} )</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>0.8</td>
<td>0.0967 0.1023</td>
<td>120.8893 126.4907</td>
<td>0.0017</td>
<td>0.1005 0.1028</td>
<td>121.7910 126.5564</td>
<td>0.0017</td>
<td>0.1691 0.1747</td>
<td>112.1230 109.9599</td>
<td>0.0015</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0682 0.0737</td>
<td>129.3255 123.0665</td>
<td>0.0012</td>
<td>0.0686 0.0749</td>
<td>135.4227 124.4326</td>
<td>0.0013</td>
<td>0.1134 0.1190</td>
<td>117.9012 114.4538</td>
<td>0.0010</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0967 0.1023</td>
<td>120.8893 126.4907</td>
<td>0.0017</td>
<td>0.1005 0.1028</td>
<td>121.7910 126.5564</td>
<td>0.0017</td>
<td>0.1691 0.1747</td>
<td>112.1230 109.9599</td>
<td>0.0015</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0682 0.0737</td>
<td>129.3255 123.0665</td>
<td>0.0012</td>
<td>0.0686 0.0749</td>
<td>135.4227 124.4326</td>
<td>0.0013</td>
<td>0.1134 0.1190</td>
<td>117.9012 114.4538</td>
<td>0.0010</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0967 0.1023</td>
<td>120.8893 126.4907</td>
<td>0.0017</td>
<td>0.1005 0.1028</td>
<td>121.7910 126.5564</td>
<td>0.0017</td>
<td>0.1691 0.1747</td>
<td>112.1230 109.9599</td>
<td>0.0015</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0682 0.0737</td>
<td>129.3255 123.0665</td>
<td>0.0012</td>
<td>0.0686 0.0749</td>
<td>135.4227 124.4326</td>
<td>0.0013</td>
<td>0.1134 0.1190</td>
<td>117.9012 114.4538</td>
<td>0.0010</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0967 0.1023</td>
<td>120.8893 126.4907</td>
<td>0.0017</td>
<td>0.1005 0.1028</td>
<td>121.7910 126.5564</td>
<td>0.0017</td>
<td>0.1691 0.1747</td>
<td>112.1230 109.9599</td>
<td>0.0015</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0682 0.0737</td>
<td>129.3255 123.0665</td>
<td>0.0012</td>
<td>0.0686 0.0749</td>
<td>135.4227 124.4326</td>
<td>0.0013</td>
<td>0.1134 0.1190</td>
<td>117.9012 114.4538</td>
<td>0.0010</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0967 0.1023</td>
<td>120.8893 126.4907</td>
<td>0.0017</td>
<td>0.1005 0.1028</td>
<td>121.7910 126.5564</td>
<td>0.0017</td>
<td>0.1691 0.1747</td>
<td>112.1230 109.9599</td>
<td>0.0015</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0682 0.0737</td>
<td>129.3255 123.0665</td>
<td>0.0012</td>
<td>0.0686 0.0749</td>
<td>135.4227 124.4326</td>
<td>0.0013</td>
<td>0.1134 0.1190</td>
<td>117.9012 114.4538</td>
<td>0.0010</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0967 0.1023</td>
<td>120.8893 126.4907</td>
<td>0.0017</td>
<td>0.1005 0.1028</td>
<td>121.7910 126.5564</td>
<td>0.0017</td>
<td>0.1691 0.1747</td>
<td>112.1230 109.9599</td>
<td>0.0015</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0682 0.0737</td>
<td>129.3255 123.0665</td>
<td>0.0012</td>
<td>0.0686 0.0749</td>
<td>135.4227 124.4326</td>
<td>0.0013</td>
<td>0.1134 0.1190</td>
<td>117.9012 114.4538</td>
<td>0.0010</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0967 0.1023</td>
<td>120.8893 126.4907</td>
<td>0.0017</td>
<td>0.1005 0.1028</td>
<td>121.7910 126.5564</td>
<td>0.0017</td>
<td>0.1691 0.1747</td>
<td>112.1230 109.9599</td>
<td>0.0015</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0682 0.0737</td>
<td>129.3255 123.0665</td>
<td>0.0012</td>
<td>0.0686 0.0749</td>
<td>135.4227 124.4326</td>
<td>0.0013</td>
<td>0.1134 0.1190</td>
<td>117.9012 114.4538</td>
<td>0.0010</td>
</tr>
</tbody>
</table>
Table V.3. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under Stratified Random Sampling when Response Rate = 40%, $\sigma_v^2 = \sigma_u^2 = \sigma_p^2 = 1$, f = 2 and $\sigma_s^2 = 1^*\sigma_x^2$.

<table>
<thead>
<tr>
<th>Est. $\hat{\mu}_w$</th>
<th>$\sigma_x^2$</th>
<th>MSE</th>
<th>PRE</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W 0.5</td>
<td>$\sigma_T^2$</td>
<td>Without ME</td>
<td>With ME</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.0454</td>
<td>0.0479</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0910</td>
<td>0.0935</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1365</td>
<td>0.1390</td>
<td>100.0000</td>
</tr>
<tr>
<td>$\hat{\mu}_{yw}$ 0.8</td>
<td>0</td>
<td>0.0489</td>
<td>0.0514</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1215</td>
<td>0.1240</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1942</td>
<td>0.1967</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.0512</td>
<td>0.0537</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1416</td>
<td>0.1441</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.2325</td>
<td>0.2350</td>
<td>100.0000</td>
</tr>
<tr>
<td>$\hat{\mu}_{rw}$ 0.8</td>
<td>0</td>
<td>0.0308</td>
<td>0.0408</td>
<td>147.4026</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0756</td>
<td>0.0855</td>
<td>120.3704</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.1206</td>
<td>0.1306</td>
<td>113.1841</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.0341</td>
<td>0.0441</td>
<td>143.4018</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1056</td>
<td>0.1156</td>
<td>115.0568</td>
</tr>
</tbody>
</table>
Tables V.1, V.2, V.3 present the theoretical and empirical MSEs and PREs of the ORRT mean estimators under stratified random sampling when all the variances of measurement errors ($\sigma^2_v, \sigma^2_u$ and $\sigma^2_p$) are set equal to 1 and response rate in Phase I.
is set equal to 40% with different variances of S (0.2*σₙ², 0.5*σₙ², 1*σₙ²), respectively. Also, the mean estimation is less efficient as the variance of S increases in the presence of non-response. This is consistent with the theoretical results. As mentioned earlier, larger variance of S introduces more penalty for using RRT models.

For all three tables, the MSE of the mean estimators increases as W increases under non-response and measurement errors. For example, in Table V.3, the MSE of the generalized mean estimator increased from 0.1218 to 0.2175 as the sensitivity level increased from 0.5 to 1 when variance of T is equal to 1. It indicates that the ORRT model is more efficient when some of the respondents feel the survey question is not sensitive. In addition, as the variance of T increases, the MSE increases while δ decreases with a reasonably small value of σₚ².

Table V.4. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under the Conditions of σᵥ² = σᵤ² = σₚ² = 1, 5, 10 and Stratified Random Sampling when Response Rate = 40%, W = 0.8, and σₚ² = 0.5*σₓ².

<table>
<thead>
<tr>
<th>Est.</th>
<th>f</th>
<th>MSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>0.1301</td>
<td>0.1445</td>
<td>0.1443</td>
</tr>
<tr>
<td>3</td>
<td>0.1770</td>
<td>0.1956</td>
<td>0.2053</td>
</tr>
<tr>
<td>4</td>
<td>0.2363</td>
<td>0.2482</td>
<td>0.2765</td>
</tr>
<tr>
<td>2</td>
<td>0.1130</td>
<td>0.1524</td>
<td>0.1898</td>
</tr>
<tr>
<td>3</td>
<td>0.1692</td>
<td>0.2207</td>
<td>0.2831</td>
</tr>
</tbody>
</table>
Table V.5. Theoretical (bold) and Empirical MSEs/PREs of the ORRT Estimators under the Conditions of Response Rate = 20%, 40%, 60% and Stratified Random Sampling when $\sigma^2_v = \sigma^2_u = \sigma^2_p = 1$, $W = 0.8$, and $\sigma^2_T = 0.5\sigma^2_x$.

<table>
<thead>
<tr>
<th>Response Rate</th>
<th>$f$</th>
<th>$\hat{\mu}^{st}_{pw}$</th>
<th>$\hat{\mu}^{st}_{stw}$</th>
<th>$\hat{\mu}^{st}_{yw}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20%</td>
<td>4</td>
<td>0.2180 0.2895 0.3707</td>
<td>0.2182 0.2855 0.3679</td>
<td>0.1023 0.1288 0.1342</td>
</tr>
<tr>
<td>40%</td>
<td>2</td>
<td>0.1023 0.1288 0.1342</td>
<td>0.1028 0.1282 0.1323</td>
<td>0.1515 0.1743 0.1957</td>
</tr>
<tr>
<td>60%</td>
<td>3</td>
<td>0.1597 0.1478 0.1252</td>
<td>0.1528 0.1770 0.1991</td>
<td>0.2007 0.2298 0.2572</td>
</tr>
<tr>
<td>40%</td>
<td>4</td>
<td>0.1986 0.2329 0.2575</td>
<td>0.1986 0.2329 0.2575</td>
<td>0.1447 0.1111 0.0882</td>
</tr>
<tr>
<td>60%</td>
<td>3</td>
<td>0.2297 0.1747 0.1252</td>
<td>0.2292 0.1770 0.1259</td>
<td>0.3053 0.2300 0.1621</td>
</tr>
<tr>
<td>60%</td>
<td>4</td>
<td>0.3035 0.2363 0.1624</td>
<td>0.3035 0.2363 0.1624</td>
<td>0.1447 0.1111 0.0882</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>20% MSE</th>
<th>40%</th>
<th>60%</th>
<th>20% PRE</th>
<th>40%</th>
<th>60% PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2180 0.2895 0.3707</td>
<td>0.2182 0.2855 0.3679</td>
<td>0.1023 0.1288 0.1342</td>
<td>0.1028 0.1282 0.1323</td>
<td>0.1515 0.1743 0.1957</td>
<td>0.1528 0.1770 0.1991</td>
</tr>
</tbody>
</table>
Table V.4 presents the theoretical and empirical MSEs and PREs of the ORRT mean estimators under stratified random sampling and different variances of measurement errors (1, 5, 10) when the sensitivity level W is equal to 0.8, variance of T is equal to 0.5 and response rate in Phase I is equal to 40%. As the measurement errors increase, the MSE of each mean estimator increases. For example, the MSE of the ratio estimator increased from 0.1645 to 0.2838 as the variance of measurement errors increased from 1 to 10 when the value of f is 3. It is obvious that larger measurement errors have larger negative impact on mean estimation under non-response using stratified random sampling, as was the case with simple random sampling.

Also, from Tables V.1, V.2, V.3 and V.4, it is clear that the generalized mean estimator \(\hat{\mu}_{pw}\) is more efficient than the other two mean estimators even when very large measurement errors are present. However, the ratio estimator \(\hat{\mu}_{rw}\) becomes less efficient than the ordinary mean estimator \(\hat{\mu}_{yw}\) as the measurement errors increase. For example in Table V.4, the MSE of the generalized mean estimator 0.1288 is less than the MSE of the ordinary mean estimator 0.1395 when the variance of measurement errors is 5 and the value of f is 2. However, the MSE of the ratio estimator 0.1509 is larger than other two estimators. Similar to mean estimation under simple random sampling, measurement error on X makes the ratio estimator less efficient than the ordinary mean estimator unless the variance of measurement error on X is small because measurement errors exist only on Y for the ordinary mean estimator while in the ratio estimator measurement errors exist on both X and Y. At the same time, this result shows the superiority of the generalized mean estimator in the presence of measurement errors and non-response using stratified random sampling because it is not affected as poorly as the ratio estimator is by measurement errors on X.
Table V.5 presents the theoretical and empirical MSEs and PREs of the ORRT mean estimators under stratified random sampling and different response rates when the variance of measurement errors is equal to 1, sensitivity level $W$ is equal to 0.8, and variance of $T$ is equal to 0.5. The efficiency of each estimator gets better as the response rate increases. In other words, the larger the sample we collect from the first call in each stratum, the higher is the efficiency of the mean estimation.

In addition, from both Tables V.4 and V.5, the efficiency of each estimator gets better as the value of $f$ decreases. For example, the MSE of the generalized mean estimator decreased from 0.2007 to 0.1023 as the value of $f$ decreased from 4 to 2 when the variance of measurement errors is 1. It is reasonable because smaller $f$ value means we obtain a larger sample from the second call in each stratum and the mean estimation is more efficient when a larger sample is used.

V.4. Concluding Chapter Remarks

The main contribution in this chapter is the re-examination of Chapter IV but under stratified random sampling. Under stratified random sampling, all the conclusions we made under simple random sampling still hold true. From the theoretical conditions (V.27) (V.29) (V.31) and simulation results, the generalized mean estimator is more efficient than the ordinary mean estimator and the ratio estimator when the measurement errors and non-response are present, while the ratio estimator is less efficient than the ordinary mean estimator when the measurement errors on X are large. The reason is the ordinary mean estimator is not affected by the measurement error in X. Even though the generalized mean estimator is also affected by measurement error on X, the use of the regression term was able to overcome the measurement error burden on X.
CHAPTER VI
CONCLUDING REMARKS AND FUTURE DIRECTIONS

Mean estimation of a sensitive variable under measurement errors and non-response using both simple random sampling and stratified random sampling is studied in this dissertation. The empirical results are in good agreement with the corresponding theoretical conclusions. The simple additive RRT is more efficient if we ignore privacy issue, but the general linear combination RRT model is better if we examine the performance of various estimators with respect to the unified measure of efficiency and privacy. The MSE of all mean estimators increases as $W$ increases under all conditions, which shows that the ORRT model leads to better results than non-optional model. The generalized mean estimator always performs better than both the ordinary mean estimator and ratio estimator under measurement errors and non-response. The ratio estimator is more efficient than the ordinary mean estimator when the study variable and the auxiliary variable are highly correlated and the measurement errors on the auxiliary variable are small.

For future studies, one can consider mean estimation of a sensitive variable using different sampling methods, such as unequal probability sampling. Also, instead of estimating mean, one can estimate other parameters like the variance and the distribution function.


APPENDIX A
LIST OF PUBLICATIONS


