SIHM, JEONG SEP, M.A. A Two-Stage Binary Optional Randomized Response Model. (2012)
Directed by Sat Narain Gupta. 84 pp.

Social Desirability Bias (SDB) is the tendency in respondents to answer questions untruthfully in the hope of giving good impression to others. SDB occurs when the survey question is highly sensitive or personal, and responses cause sample statistics to systematically over- or underestimate corresponding population parameters. The Randomized Response Technique (RRT) is one of several methods to get around SDB in surveys involving sensitive questions in a face-to-face interview.

In this thesis, we first review some of the existing binary response RRT models. Then, by combining two existing models, we propose a new model—Two-Stage Binary Optional RRT model. Much of the focus is on estimating π, the prevalence of sensitive characteristic and ω, the sensitivity level of the underlying question. We discuss the asymptotic properties of our estimators and present some simulation results. It turns out that the proposed Two-Stage Binary Optional RRT model is more effective than the Optional RRT model proposed by Gupta 2001 [4].

A TWO-STAGE BINARY OPTIONAL RANDOMIZED RESPONSE MODEL

by

Jeong Sep Sihm

A Thesis Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Greensboro
2012

Approved by

____
Committee Chair

To my wife,

Mi-Young,

as well as to my three delightful children,

Jin Woo, Jane, and Jay Young,

for their support and encouragement.

APPROVAL PAGE

This thesis 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 Narain Gupta

Committee Members ____________________
Scott James Richter

____
Dohyoung Ryang

____
Date of Acceptance by Committee

____
Date of Final Oral Examination

ACKNOWLEDGMENTS

The work for this thesis was undertaken at the suggestion of Professor Sat Narain Gupta, to whom I wish to express my heartfelt appreciation and eternal gratitude for his assistance and helpful criticism. He has guided me so wisely and always knows the answer. I am also indebted to Professor Scott James Richter and Professor Dohyoung Ryang for the assistance and direction they have given me in this undertaking.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Social Desirability Bias

Social Desirability Bias (SDB) is the idiosyncrasy created by respondents in answering sensitive questions unfaithfully in the hope of leaving good impression on others. In addition to this Impression Management component, there also exists Self-Deception component in SDB. Some people just tend to believe that they are not engaged in socially undesirable activities and report to the interviewer accordingly, causing different kind of SDB. Paulhus 1984 [16] recommends that Impression Management, not Self-Deception, be controlled in survey research. SDB can happen when the survey question is highly sensitive or personal. This is one of the many biases which occur during survey sampling. Other typical biases are evasive answer bias, refusal bias, non-response bias, selection bias, voluntary response bias, and so forth. These biases create a problem because they cause sample statistics to systematically over- or underestimate corresponding population parameters.

There are several techniques to promote faithful answers and to avoid Impression Management component of SDB such as the Bogus Pipeline Technique, the Unmatched Count Technique, and the Randomized Response Technique.

1.2 Unmatched Count Technique

This technique has couple of different names; the Item Count Technique and the List Technique. The basic idea of the Unmatched Count Technique (UCT) is very simple. Randomly selected respondents in the control group receive a group of non-sensitive questions, and are asked to report the number of “yes” answers. After one more question—which is sensitive—is added to the existing set of questions, the new set of questions is given to the other group. As members of both groups are randomly selected, we can assume that their proportions of “yes” responses towards the non-sensitive questions would be the same. Thus, we can get the unmatched count from the experimental group. As the respondents are required to simply report the number of “yes” answers, Impression Management component of SDB can be avoided. The population proportion of “yes” answer to the sensitive question can now be deduced statistically.

In many cases, it would be easy for a researcher to implement the UCT. Just paper and pencils are needed and no other complex randomization devices are required. Also, for the participants, the UCT is quite easy to understand and straightforward, providing a strong perceived sense of privacy. Studies such as Coutts and Jann 2011 [2] and Lavender and Anderson 2009 [12] have shown that, in practice, the UCT is more effective than other techniques because the highest perception of anonyminty is found for the UCT among the respondents. However, the theory for the UTC model is not as extensive as is for the RRT models. The RRT models allow many different kinds of improvements which make these models more efficient. These improvements include optional models and two-stage models.

1.3 Bogus Pipeline Method

The term Bogus Pipeline (BPL) was coined by Jones and Sigall 1971 [10] to discribe an imaginary dream device for psychologists, which would provide a direct pipleline to the soul. Thus, they could have access to reliable phychological indicators. Jones and Sigall 1971 [10] proposed that respondents' answers wouldn't be contaminated by many of the biases, including SDB, if they were convinced that the device in front of them was an actual polygraph. Their explanation was that respondents didn't want to be second-guessed by a machine, trying to avoid possible loss of face while believing the true answers would be revealed regardless of their response. Roese and Jamieson 1993 [18] showed that the BPL produced reliable effects consistent with a reduction in SDB after meta-analysing 31 studies that had used the bogus pipeline for their research.

1.4 Randomized Response Technique

The Randomized Response Technique (RRT) was first proposed by Warner 1965 [21]. It is a survey research method specifically designed to ask sensitive questions.

Suppose we need to estimate the proportion of drug abusers in a population in the last 3 months. Let us have a deck of cards where 10% of the cards have the statement "I have used controlled substances without prescription at least once in the last 3 months." The rest of the cards have the statement "I have not used controlled substances without prescription in the last 3 months," written on them. The respondents are expected to give a binary answer—either "yes, this statement is correct," or "no, this statement is not correct"—to the statement on the card which they draw from the deck. Due to the randomization device—10% probability of drawing drug abuse question, the researcher has no idea of what a "yes" answer means individually.

Notice that it is quite important in practice for a respondent to understand that the RRT maintains privacy, as the randomization device is invisible. Some of the respondents might not be able to grasp this probability concept easily. Without this understanding, Impression Management component of SDB cannot be overcome.

Since the RRT method was first introduced in 1965, there are many areas where the RRT models have been used. One of the interesting studies using the RRT in practice is by Schneider 2003 [19]. It was an experimental study to examine whether compensation and stock ownership affect internal auditors' objectivity. In order to elicit truthful responses and overcome SDB from active internal auditors, Schneider adopted the RRT and collected randomized responses from 172 participants. It was found that stock ownership did not affect internal auditors' reporting decisions while compensation tied to stock prices made internal auditors report violations less frequently. In the Netherlands, Lensvelt-Mulders, van der Heijden, Laudy, and van Gils 2006 [13] validated a computer assisted RRT survey to estimate the prevalence of fraud in disability benefits. By the time of Lensvelt-Mulders et al.'s research, the actual survey to estimate the disability fraud in the Netherlands included home interviews by trained interviewers with randomized response questions. Lavender and Anderson 2009 [12] assessed the effect of perceived anonymity on endorsements of eating disorder bahaviors and attitudes among 469 undergraduate women from a university in the Northeastern

United States. They used a standard anonymous true/false survey, the UCT, and the RRT. Then they compared the results generated by those three different survey techniques. In Germany, Ostapczuk, Musch, and Moshagen 2009 [15] studied SDB among the highly educated and the less educated in their attitude towards foreigners, comparing their answers from direct questioning conditions and the RRT conditions. In Hong Kong, Kwan, So, and Tam 2010 [11] showed how truthful answers to sensitive questions about software piracy can be estimated by using the RRT.

1.5 Outline of the Thesis

Chapter I has presented a brief introduction to Social Desirablity Bias and discussed several techniques to promote faithful answers in answering sensitive questions. It also has discussed how those techniques were applied in practice.

Chapter II presents three previous studies and models in the RRT area, which serve as the foundation for the proposed model in this thesis.

Chapter III proposes the Two-Stage Binary Optional RRT model and examines estimators for the two parameters of the model (π—the prevalence of sensitive characteristic and ω—the sensitivity level of the underlying question) and the variances of them. In Section 3.2, $\widehat{\pi}_p$ and $Var(\widehat{\pi}_p)$ are discussed. As for $\widehat{\omega}_p$, the first order approximation of $Var(\widehat{\omega}_p)$ is presented in Section 3.3. Asymptotic normality of $\widehat{\pi}_p$ and $\widehat{\omega}_1$ is discussed in these sections too.

Chapter IV presents optimal sub-sample sizes of n_1 and n_2 to minimize $Var(\widehat{\pi}_p)$ in Theorem IV.1 and to minimize $Var(\widehat{\omega}_1)$ in Theorem IV.2. It also presents optimal value of the Two-Stage Parameter (T).

Chapter V presents how the simulations are set up and discusses the results of simulations of the proposed model.

Chapter VI presents the concluding remarks of this thesis.

Appendix A presents the R program code for the simulations of the proposed model.

CHAPTER II

SOME RANDOMIZED RESPONSE TECHNIQUE MODELS

2.1 First Model

In his groundbreaking paper in 1965, Warner [21] proposed a very interesing idea of how to deal with evasive answer bias, especially when it comes to personal or controversial survey questions. The basic idea is very simple; putting a randomization mechanism between the interviewer and the interviewee, so that the intervewer cannot know what the answer will really mean. By permitting the interviewee to maintain privacy, one can expect increased cooperation and a more truthful answer from the interviewee. Throughout this thesis, we assume that our sample is a simple random sample with replacement.

Warner 1965 [21] proposed a spinner with probability p pointing to the letter A and with probability $(1-p)$ pointing to the letter B. Every respondent belongs to either Group A—the sensitive group, or Group B—the non-sensitive group. The spinner is run without the interviewer's presense and the interviewee is to report a "Yes" or a "No" to indicate whether or not the group the spinner is pointing to is the group he or she actually belongs to.

Let P_y be the probability of a "Yes" response from a respondent. Note that a "Yes" response can be provided in two ways. One is when the respondent belongs to Group A while the spinner points to A. Another is when he or she belongs to Group B while the spinner points to B. Let π be the proportion of a population that belongs to Group A. We want to estimate π.

Then P_y can be expressed as follows.

$$P_y = \pi p + (1-\pi)(1-p) \tag{II.1}$$

Solving for π, we have

$$\pi = \frac{P_y - (1-p)}{2p-1}.$$

Thus, **the Warner's estimate of $\boldsymbol{\pi}$** is given by

$$\widehat{\pi}_w = \frac{\widehat{P_y} - (1-p)}{2p-1}. \tag{II.2}$$

where $\widehat{P_y}$ is the proportion of "Yes" responses in the survey.

$\widehat{P_y}$ is both an unbiased and the Maximum Likelihood Estimator (MLE) of P_y as shown in Chaudhuri 2011 [1]. Taking expected value on both sides of Equation (II.2), we get

$$E\left(\widehat{\pi}_w\right) = \frac{E\left(\widehat{P_y}\right) - (1-p)}{2p-1} = \frac{P_y - (1-p)}{2p-1} = \pi.$$

Thus, $\widehat{\pi}_w$ is an unbiased estimator of π.

Using $Var(\widehat{P_y}) = \frac{P_y(1-P_y)}{n}$, the variance of $\widehat{\pi}_w$ is,

$$Var\left(\widehat{\pi}_w\right) = \frac{1}{(2p-1)^2} Var\left(\widehat{P_y}\right) \tag{II.3}$$

$$= \frac{1}{(2p-1)^2}\left\{\frac{P_y(1-P_y)}{n}\right\}. \tag{II.4}$$

After substituting P_y from Equation (II.1) into Equation (II.4), we have **the variance of the Warner's estimator** as given by

$$Var(\widehat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2} \tag{II.5}$$

with

$$\widehat{Var}(\widehat{\pi}_w) = \frac{\widehat{\pi}_w(1-\widehat{\pi}_w)}{n-1} + \frac{p(1-p)}{(n-1)(2p-1)^2} \tag{II.6}$$

2.2 Two-Stage Model by Mangat and Singh, 1990

In 1990, Mangat and Singh [14] introduced a Two-Stage RRT model by injecting an element of truthful responses into the Warner's randomized response model [21].

In order to have more truthful answers, they put one more randomization device into the original Warner's model. The first randomization device has two options: (1) ‘Do you belong to Group A?’, and (2) ‘Go to the second randomization device,’ And, the second stage—or the second randomization device—is nothing but the Warner's randomization device. The probabilities of (1) and (2) are known to be T and $(1-T)$, respectively. Because the entire process remains unobserved by the interviewer as in the Warner's model, the interviewee can maintain privacy regardless of the answer either from the first randomization device or from the Warner's randomization device.

Let P_y be the probability of a "Yes" response from a respondent under this model. P_y is given by

$$P_y = T\pi + (1-T)\{\pi p + (1-\pi)(1-p)\} = \{T + (2p-1)(1-T)\}\pi + (1-T)(1-p). \tag{II.7}$$

Rewriting this equation for π

$$\pi = \frac{P_y - (1-p)(1-T)}{T + (2p-1)(1-T)} = \frac{P_y - (1-p)(1-T)}{(2p-1) + 2T(1-p)}.$$

This leads to **the Mangat and Singh's estimator for $\boldsymbol{\pi}$**, given by

$$\widehat{\pi}_m = \frac{\widehat{P_y} - (1-p)(1-T)}{(2p-1) + 2T(1-p)}. \tag{II.8}$$

where $\widehat{P_y}$ is the proportion of "Yes" responses in the survey.

As $\widehat{P_y}$ is both an unbiased and the MLE of P_y, $\widehat{\pi}_m$ is unbiased too. This can be seen from the fact that

$$E(\widehat{\pi}_m) = \frac{E\left(\widehat{P_y}\right) - (1-p)(1-T)}{(2p-1) + 2T(1-p)} = \frac{P_y - (1-p)(1-T)}{(2p-1) + 2T(1-p)} = \pi.$$

Also,

$$Var(\widehat{\pi}_m) = \frac{1}{\{(2p-1)+2T(1-p)\}^2} Var(\widehat{P_y}) \tag{II.9}$$

$$= \frac{1}{\{(2p-1)+2T(1-p)\}^2} \left\{ \frac{P_y(1-P_y)}{n} \right\}. \tag{II.10}$$

Using Equation (II.7), this can be rewritten as

$$Var(\widehat{\pi}_m) = \frac{\pi(1-\pi)}{n} + \frac{(1-T)(1-p)\{1-(1-T)(1-p)\}}{n\{(2p-1)+2T(1-p)\}^2} \tag{II.11}$$

with

$$\widehat{Var}(\widehat{\pi}_m) = \frac{\widehat{\pi}_m(1-\widehat{\pi}_m)}{n-1} + \frac{(1-T)(1-p)\{1-(1-T)(1-p)\}}{(n-1)\{(2p-1)+2T(1-p)\}^2} \tag{II.12}$$

Mangat and Singh 1990 [14] showed that

$$Var(\widehat{\pi}_m) < Var(\widehat{\pi}_w) \qquad \text{if} \qquad \frac{1-2p}{1-p} < T \tag{II.13}$$

As $\frac{1-2p}{1-p} < 1$ for $0 < p < 1$, a meaningful value of T can be chosen between $\frac{1-2p}{1-p}$ and 1.

2.3 Optional Randomized Response Model by Gupta, 2001

It is reasonable to assume that some proportion of the population might not feel that the survey question is sensitive and would give candid answers if they get the option to answer truthfully. Instead of injecting an element of truth by the researchers as in the Two-Stage Model by Mangat and Singh, we can incorporate this unknown proportion of truthfulness differently into a new model. In this Optional Model, the respondent has the freedom to choose how to answer the question. If the respondent feels the question is sensitive, he or she can give a scrambled response. If the respondent doesn't feel it's a sensitive quesion, he or she can just give a true answer. This optional randomization process takes place without being observed by the researcher, who has no idea of what method the respondent chose and what a "Yes" response means.

In the Two-Stage Model, parameter T could be chosen by the interviewer, thus was a known constant prior to using the two randomization devices. In this Optional Model, the sensitivity level (ω) of a specific question is defined to be the population proportion of subjects who feel the question is sensitive. Notice that there are two unknown parameters in this model—π and ω. The Optional Randomized Response models were first proposed by Gupta 2001 [4] and Gupta, Gupta, and Singh 2002 [5]. The characteristics of the models have been discussed in great depth by Gupta and Shabbir 2004 [7], Gupta, Thornton, Shabbir, and Singhal 2006 [9], Gupta, Shabbir, and Sehra 2010 [8], and Gupta, Mehta, Shabbir, and Dass 2012 [6].

The probability of a "Yes" response in this model can be expressed as

$$P_y = (1-\omega)\pi + \omega\{\pi p + (1-\pi)(1-p)\} \tag{II.14}$$

Equation (II.14) can be rearranged as

$$P_y - \pi = (p-1)(2\pi - 1)\omega \tag{II.15}$$

As Equation (II.15) includes two parameters—π and ω, it cannot be handled with one set of responses. Assume we have two independent samples with sample sizes n_1 and n_2 respectively ($n_1 + n_2 = n$). Let us also assume that p_1 and p_2 are different probabilities associated with the different Warner's devices used in the two samples.

Using Equation (II.15) for the two independent samples, we have

$$P_{y_1} - \pi = (p_1 - 1)(2\pi - 1)\omega \qquad \text{and} \qquad P_{y_2} - \pi = (p_2 - 1)(2\pi - 1)\omega \tag{II.16}$$

With $\lambda = \frac{(p_1-1)}{(p_2-1)}$ as in Greenberg, Abul-Ela, Simmons, and Horvitz 1969 [3], we have

$$\pi = \frac{\lambda P_{y_2} - P_{y_1}}{\lambda - 1} \tag{II.17}$$

From Equation (II.17), we have **the Gupta estimator for $\boldsymbol{\pi}$** as

$$\widehat{\pi}_g = \frac{\lambda \widehat{P_{y_2}} - \widehat{P_{y_1}}}{\lambda - 1} \tag{II.18}$$

where $\widehat{P_{y_1}}$ and $\widehat{P_{y_2}}$ are the proportions of "Yes" responses in the two samples.

Note that $\widehat{\pi}_g$ is unbiased as shown below.

$$E\left(\widehat{\pi}_g\right) = \frac{\lambda E\left(\widehat{P_{y_2}}\right) - E\left(\widehat{P_{y_1}}\right)}{\lambda - 1} = \frac{\lambda P_{y_2} - P_{y_1}}{\lambda - 1} = \pi \tag{II.19}$$

Using $Var(\widehat{P_{y_1}}) = \frac{P_{y_1}(1-P_{y_1})}{n_1}$ and $Var(\widehat{P_{y_2}}) = \frac{P_{y_2}(1-P_{y_2})}{n_2}$, **the variance of $\widehat{\pi}_g$** is,

$$Var(\widehat{\pi}_g) = \frac{1}{(\lambda-1)^2}\{\lambda^2 Var(\widehat{P_{y_2}}) + Var(\widehat{P_{y_1}})\}$$

$$= \frac{1}{(\lambda-1)^2}\left\{\lambda^2 \frac{P_{y_2}(1-P_{y_2})}{n_2} + \frac{P_{y_1}(1-P_{y_1})}{n_1}\right\} \tag{II.20}$$

Notice that the two samples are independent so that the covariance term does not exist in Equation (II.20).

Using $n_1 = n - n_2$, we can rewrite Equation (II.20) as

$$Var(\widehat{\pi}_g) = \frac{1}{(\lambda-1)^2}\left\{\lambda^2 \frac{P_{y_2}(1-P_{y_2})}{n_2} + \frac{P_{y_1}(1-P_{y_1})}{n-n_2}\right\} \tag{II.21}$$

After taking partial derivative on both sides of Equation (II.21), the optimal ratio of $\frac{n_1}{n_2}$—which gives the minimum variance—is obtained.

$$\frac{\partial Var(\widehat{\pi}_g)}{\partial n_2} = \frac{1}{(\lambda-1)^2}\left\{-\lambda^2 \frac{P_{y_2}(1-P_{y_2})}{n_2^2} + \frac{P_{y_1}(1-P_{y_1})}{(n-n_2)^2}\right\} = 0 \tag{II.22}$$

Solving Equation (II.22) for $\frac{n_1}{n_2}$, we have **the optimal ratio of $\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_g)}$** as follows.

$$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_g)} = \frac{1}{\lambda}\sqrt{\frac{P_{y_1}(1-P_{y_1})}{P_{y_2}(1-P_{y_2})}} = \frac{(1-p_2)}{(1-p_1)}\sqrt{\frac{P_{y_1}(1-P_{y_1})}{P_{y_2}(1-P_{y_2})}} \tag{II.23}$$

Now, let us solve Equations (II.16) for ω. Note that

$$P_{y_1} - P_{y_2} = (p_1 - p_2)(2\pi - 1)\omega \tag{II.24}$$

Solving Equation (II.24) for ω and substituting $\pi = \frac{\lambda P_{y_2} - P_{y_1}}{\lambda - 1}$ and $\lambda = \frac{(p_1-1)}{(p_2-1)}$ from Equations (II.17), we have,

$$\omega = \frac{P_{y_1} - P_{y_2}}{2P_{y_1}(1-p_2) - 2P_{y_2}(1-p_1) - (p_1 - p_2)}. \tag{II.25}$$

By replacing P_{y_1} and P_{y_2} with their unbiased MLEs, **the Gupta estimator for $\boldsymbol{\omega}$** is

$$\widehat{\omega}_g = \frac{\widehat{P_{y_1}} - \widehat{P_{y_2}}}{2\widehat{P_{y_1}}(1-p_2) - 2\widehat{P_{y_2}}(1-p_1) - (p_1 - p_2)} \tag{II.26}$$

Given that $\widehat{\omega}_g$ is a ratio of combinations of two random variables, calculation of its mean and variance will require some approximation, as we will discuss in the next chapter.

CHAPTER III

PROPOSED MODEL: TWO-STAGE BINARY OPTIONAL RRT MODEL

3.1 Model Setup

The proposed model in this thesis is the combination of the Two-Stage Model in Section 2.2 and the Optional Model in Section 2.3. The first randomization device the interviewee encounters in the proposed model has two options (1) 'Do you belong to the sensitive group?' and (2) 'Go to the second randomization device.' The second stage—or the second randomization device—is nothing but the Optional RRT Model by Gupta 2001 [4] in Section 2.3. The interviewee is not observed during the entire process of applying this model like the previous models explained in Chapter II, in order for him or her to maintain privacy.

Let P_y be the probability of "Yes" response from a respondent under this model, T be the probability of asking 'Do you belong to the sensitive group?' in the first randomized device, π be the proportion of the population that belongs to the sensitive group, p be the probability of the spinner pointing to the sensitive group, and ω be the level of sensitivity of the survery question in the population. We have,

$$P_y = T\pi + (1-T)\left\{(1-\omega)\pi + \omega\{\pi p + (1-\pi)(1-p)\}\right\}. \tag{III.1}$$

Equation (III.1) can be re-arranged as

$$P_y = T\pi + \pi - T\pi + (1 - T)\left\{-\pi + \pi p + (1 - \pi)(1 - p)\right\}\omega$$

This leads to

$$P_y - \pi = (1 - T)(p - 1)(2\pi - 1)\omega \tag{III.2}$$

Equation (III.2) cannot be handled directly, because it has two unknown parameters—π and ω—in it. T is assumed known. Assume also we have two independent samples with sample sizes n_1 and n_2 respectively $(n_1 + n_2 = n)$. Let us assume that p_1 and p_2 are different probabilites associated with the different Warner's devices used in the two samples with this background.

Using Equation (III.2) for the two independent samples, we have

$$P_{y_1} - \pi = (1 - T)(p_1 - 1)(2\pi - 1)\omega \tag{III.3}$$

$$P_{y_2} - \pi = (1 - T)(p_2 - 1)(2\pi - 1)\omega \tag{III.4}$$

With $\lambda = \frac{(p_1 - 1)}{(p_2 - 1)}$, we get

$$\pi = \frac{\lambda P_{y_2} - P_{y_1}}{\lambda - 1} \tag{III.5}$$

3.2 $\widehat{\pi}_p$ and $Var\left(\widehat{\pi}_p\right)$

Equation (III.5) leads to the estimator

$$\widehat{\pi}_p = \frac{\lambda\widehat{P_{y_2}} - \widehat{P_{y_1}}}{\lambda - 1} \tag{III.6}$$

where $\widehat{P_{y_1}}$ and $\widehat{P_{y_2}}$ are the proportions of "Yes" responses in the two samples.

Theorem III.1. *$\widehat{\pi}_p \sim AN(\pi, V_\pi)$, where $V_\pi = \frac{1}{(\lambda-1)^2}\left\{\lambda^2\frac{P_{y_2}(1-P_{y_2})}{n_2} + \frac{P_{y_1}(1-P_{y_1})}{n_1}\right\}$.*

Proof. According to Equation (III.6), $\widehat{\pi}_p$ is a linear combination of $\widehat{P_{y_2}}$ and $\widehat{P_{y_1}}$. As $\widehat{P_{y_2}}$ and $\widehat{P_{y_1}}$ are independent and have asymptotically normal distributions, the linear combination is also asymptotically normal. It may be noted that as the total sample size n goes to infinity, so will the sub-sample sizes n_1 and n_2, although at different rates.

Using $E(\widehat{P_{y_2}}) = P_{y_2}$ and $E(\widehat{P_{y_1}}) = P_{y_1}$, the expected value of $\widehat{\pi}_p$ is

$$E\left(\widehat{\pi}_p\right) = \frac{\lambda E\left(\widehat{P_{y_2}}\right) - E\left(\widehat{P_{y_1}}\right)}{\lambda - 1} = \frac{\lambda P_{y_2} - P_{y_1}}{\lambda - 1} = \pi.$$

Thus, $\widehat{\pi}_p$ is an unbiased estimator for π.

Using $Var(\widehat{P_{y_1}}) = \frac{P_{y_1}(1-P_{y_1})}{n_1}$, $Var(\widehat{P_{y_2}}) = \frac{P_{y_2}(1-P_{y_2})}{n_2}$, and $Cov(\widehat{P_{y_1}}, \widehat{P_{y_2}}) = 0$ as the two samples are independent, the variance of $\widehat{\pi}_p$ is given by

$$Var(\widehat{\pi}_p) = \frac{1}{(\lambda-1)^2}\{\lambda^2 Var(\widehat{P_{y_2}}) + Var(\widehat{P_{y_1}})\}$$

$$= \frac{1}{(\lambda-1)^2}\left\{\lambda^2 \frac{P_{y_2}(1-P_{y_2})}{n_2} + \frac{P_{y_1}(1-P_{y_1})}{n_1}\right\} = V_\pi \tag{III.7}$$

$\square$

Note that Equation (III.6) will be used to simulate the mean and variance of $\widehat{\pi}_p$ and Equation (III.7) will be used to calculate the theoretical variance of $\widehat{\pi}_p$ later in Section 5.1.

3.3 $\widehat{\omega}_p$ and $Var(\widehat{\omega}_p)$ with 1st Order Approximation

Subtracting Equation (III.4) from Equation (III.3), we have

$$P_{y_1} - P_{y_2} = (1-T)(p_1 - p_2)(2\pi - 1)\omega. \tag{III.8}$$

Solving for ω, we have

$$\omega = \left\{\frac{P_{y_1} - P_{y_2}}{(1-T)(p_1 - p_2)}\right\}\frac{1}{(2\pi - 1)}. \tag{III.9}$$

$(2\pi - 1)$ is expressed as follows

$$
\begin{aligned}
2\pi - 1 &= \frac{2(\lambda P_{y_2} - P_{y_1})}{\lambda - 1} - 1 \\
&= \frac{2\lambda P_{y_2} - 2P_{y_1} - \lambda + 1}{\lambda - 1} \\
&= \frac{\lambda(2P_{y_2} - 1) - (2P_{y_1} - 1)}{\lambda - 1} \\
&= \frac{\{\frac{1-p_1}{1-p_2}(2P_{y_2} - 1) - (2P_{y_1} - 1)}{\frac{1-p_1}{1-p_2} - 1} \\
&= \frac{1 - p_2}{p_2 - p_1}\left\{\frac{1 - p_1}{1 - p_2}(2P_{y_2} - 1) - (2P_{y_1} - 1)\right\} \\
&= \frac{(1 - p_1)(2P_{y_2} - 1) - (1 - p_2)(2P_{y_1} - 1)}{p_2 - p_1} \qquad \text{(III.10)}
\end{aligned}
$$

$$\text{If} \quad \theta_1 = 2P_{y_1} - 1 \quad \text{and} \quad \theta_2 = 2P_{y_2} - 1, \qquad \text{(III.11)}$$

then, $(2\pi - 1)$ is given by

$$2\pi - 1 = \frac{(1 - p_1)\theta_2 - (1 - p_2)\theta_1}{p_2 - p_1}. \qquad \text{(III.12)}$$

Now from Equation (III.9), ω is given by

$$\omega = \left\{ \frac{P_{y_1} - P_{y_2}}{(1-T)(p_1 - p_2)} \right\} \frac{1}{(2\pi - 1)}$$

$$= \frac{P_{y_1} - P_{y_2}}{(1-T)\{(1-p_2)\theta_1 - (1-p_1)\theta_2\}}$$

$$= \frac{(2P_{y_1} - 1) - (2P_{y_2} - 1)}{2(1-T)\{(1-p_2)\theta_1 - (1-p_1)\theta_2\}}$$

$$= \frac{\theta_1 - \theta_2}{2(1-T)\{(1-p_2)\theta_1 - (1-p_1)\theta_2\}} \qquad \text{(III.13)}$$

By replacing θ_i with its unbiased MLE ($\widehat{\theta}_i = 2\widehat{P_{y_i}} - 1$) in (III.13), **the estimator for $\boldsymbol{\omega}$** can be expressed as follows.

$$\widehat{\omega}_p = \frac{\widehat{\theta}_1 - \widehat{\theta}_2}{2(1-T)\{(1-p_2)\widehat{\theta}_1 - (1-p_1)\widehat{\theta}_2\}} \qquad \text{(III.14)}$$

Note that Equation (III.14) will be used to simulate the mean and variance of $\widehat{\omega}_p$ later in Section 5.2.

Also note that $\widehat{\omega}_p$ is a bivariate function of $\widehat{\theta}_1$ and $\widehat{\theta}_2$. Thus, Equation (III.14) can be written as

$$\widehat{\omega}_p(\widehat{\theta}_1, \widehat{\theta}_2) = \frac{\widehat{\theta}_1 - \widehat{\theta}_2}{2(1-T)\{(1-p_2)\widehat{\theta}_1 - (1-p_1)\widehat{\theta}_2\}} \qquad \text{(III.15)}$$

Taylor's expansion for a bivariate function $g(x,y)$ is given by, as in Spiegel 1991 [20, p.9],

$$\begin{aligned} g(x,y) &= g(a,b) + g_x(a,b)(x-a) + g_y(a,b)(y-b) \\ &+ \frac{g_{xx}(a,b)}{2!}(x-a)^2 + \frac{g_{yy}(a,b)}{2!}(y-b)^2 + 2\frac{g_{xy}(a,b)}{2!}(x-a)(y-b) \\ &+ \cdots\cdots \end{aligned} \tag{III.16}$$

So, using the first order Taylor's expansion, $\widehat{\omega}_p$ becomes,

$$\widehat{\omega}_p \approx \widehat{\omega}_p(\theta_1,\theta_2) + \left.\frac{\partial\widehat{\omega}_p(\widehat{\theta_1},\widehat{\theta_2})}{\partial\widehat{\theta_1}}\right|_{\theta_1,\theta_2}(\widehat{\theta_1}-\theta_1) + \left.\frac{\partial\widehat{\omega}_p(\widehat{\theta_1},\widehat{\theta_2})}{\partial\widehat{\theta_2}}\right|_{\theta_1,\theta_2}(\widehat{\theta_2}-\theta_2) = \widehat{\omega}_1 \tag{III.17}$$

Note the following first order partial derivatives

$$\begin{aligned} \frac{\partial\widehat{\omega}_p(\widehat{\theta_1},\widehat{\theta_2})}{\partial\widehat{\theta_1}} &= \frac{1}{2(1-T)}\frac{\partial}{\partial\widehat{\theta_1}}\left\{\frac{\widehat{\theta_1}-\widehat{\theta_2}}{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}}\right\} \\ &= \frac{1}{2(1-T)}\left\{\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}^{-1} - (\widehat{\theta_1}-\widehat{\theta_2})\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}^{-2}(1-p_2)\right\} \\ &= \frac{1}{2(1-T)}\left\{\frac{\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}-(\widehat{\theta_1}-\widehat{\theta_2})(1-p_2)}{\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}^2}\right\} \\ &= \frac{1}{2(1-T)}\left\{\frac{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}-(1-p_2)\widehat{\theta_1}+(1-p_2)\widehat{\theta_2}}{\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}^2}\right\} \\ &= \frac{1}{2(1-T)}\left\{\frac{-(1-p_1)\widehat{\theta_2}+(1-p_2)\widehat{\theta_2}}{\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}^2}\right\} \\ &= \frac{1}{2(1-T)}\left\{\frac{(p_1-p_2)\widehat{\theta_2}}{\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}^2}\right\} \end{aligned} \tag{III.18}$$

and,

$$\begin{aligned}\frac{\partial\widehat{\omega}_p(\widehat{\theta_1},\widehat{\theta_2})}{\partial\widehat{\theta_2}} &= \frac{1}{2(1-T)}\frac{\partial}{\partial\widehat{\theta_2}}\left\{\frac{\widehat{\theta_1}-\widehat{\theta_2}}{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}}\right\} \\
&= \frac{1}{2(1-T)}\left\{-\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}^{-1}+(\widehat{\theta_1}-\widehat{\theta_2})\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}^{-2}(1-p_1)\right\} \\
&= \frac{1}{2(1-T)}\left\{\frac{(\widehat{\theta_1}-\widehat{\theta_2})(1-p_1)-\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}}{\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}^2}\right\} \\
&= \frac{1}{2(1-T)}\left\{\frac{(1-p_1)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}-(1-p_2)\widehat{\theta_1}+(1-p_1)\widehat{\theta_2}}{\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}^2}\right\} \\
&= \frac{1}{2(1-T)}\left\{\frac{(1-p_1)\widehat{\theta_1}-(1-p_2)\widehat{\theta_1}}{\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}^2}\right\} \\
&= \frac{1}{2(1-T)}\left\{\frac{-(p_1-p_2)\widehat{\theta_1}}{\{(1-p_2)\widehat{\theta_1}-(1-p_1)\widehat{\theta_2}\}^2}\right\}. \qquad \text{(III.19)}\end{aligned}$$

Thus, the first order approximation of $\widehat{\omega}_p$ is

$$\widehat{\omega}_1 = \frac{1}{2(1-T)}\left\{\frac{\theta_1-\theta_2}{(1-p_2)\theta_1-(1-p_1)\theta_2}+\frac{(p_1-p_2)\theta_2(\widehat{\theta_1}-\theta_1)}{\{(1-p_2)\theta_1-(1-p_1)\theta_2\}^2}-\frac{(p_1-p_2)\theta_1(\widehat{\theta_2}-\theta_2)}{\{(1-p_2)\theta_1-(1-p_1)\theta_2\}^2}\right\} \qquad \text{(III.20)}$$

Note that Equation (III.20) will be used to simulate the mean and variance of $\widehat{\omega}_1$ later in Section 5.2.

Theorem III.2. $\widehat{\omega}_1 \sim AN(\omega, V_\omega)$,

$$\textit{where } V_\omega = \frac{(p_1-p_2)^2\left\{(2P_{y_2}-1)^2\left\{\frac{P_{y_1}(1-P_{y_1})}{n_1}\right\}+(2P_{y_1}-1)^2\left\{\frac{P_{y_2}(1-P_{y_2})}{n_2}\right\}\right\}}{(1-T)^2\left\{(1-p_2)(2P_{y_1}-1)-(1-p_1)(2P_{y_2}-1)\right\}^4}.$$

Proof. According to Equation (III.20), $\widehat{\omega}_1$ is a linear combination of $\widehat{\theta}_1$ and $\widehat{\theta}_2$, and hence of $\widehat{P_{y_2}}$ and $\widehat{P_{y_1}}$. As $\widehat{P_{y_2}}$ and $\widehat{P_{y_1}}$ are independent and have asymptotically normal distributions, the linear combination is also asymptotically normal.

Applying expected value on both sides of (III.20) and using $E(\widehat{\theta}_i) = \theta_i$, we get

$$E\left(\widehat{\omega}_1\right) = \frac{1}{2(1-T)}\left\{\frac{\theta_1-\theta_2}{(1-p_2)\theta_1-(1-p_1)\theta_2} + \frac{(p_1-p_2)\theta_2\left(E\left(\widehat{\theta}_1\right)-\theta_1\right)}{\{(1-p_2)\theta_1-(1-p_1)\theta_2\}^2} - \frac{(p_1-p_2)\theta_1\left(E\left(\widehat{\theta}_2\right)-\theta_2\right)}{\{(1-p_2)\theta_1-(1-p_1)\theta_2\}^2}\right\}$$

$$= \frac{1}{2(1-T)}\left\{\frac{\theta_1-\theta_2}{(1-p_2)\theta_1-(1-p_1)\theta_2}\right\}$$

$$= \frac{P_{y_1}-P_{y_2}}{(1-T)\{(1-p_2)(2P_{y_1}-1)-(1-p_1)(2P_{y_2}-1)\}}$$

$$= \omega$$

Thus, $\widehat{\omega}_1$ is an unbiased estimator for ω.

Taking variance on both sides of Equation (III.20), and using $Var(\widehat{\theta}_1) = 4Var(\widehat{P_{y_1}}) = \frac{4P_{y_1}(1-P_{y_1})}{n_1}$, $Var(\widehat{\theta}_2) = 4Var(\widehat{P_{y_2}}) = \frac{4P_{y_2}(1-P_{y_2})}{n_2}$, and $Cov(\widehat{P_{y_1}}, \widehat{P_{y_2}}) = 0$ (as the two samples are independent), the variance of $\widehat{\omega}_1$ is,

$$Var\left(\widehat{\omega}_1\right) = \frac{1}{4(1-T)^2}\left\{\frac{(p_1-p_2)^2\theta_2^2 Var\left(\widehat{\theta}_1\right)}{\{(1-p_2)\theta_1-(1-p_1)\theta_2\}^4} + \frac{(p_1-p_2)^2\theta_1^2 Var\left(\widehat{\theta}_2\right)}{\{(1-p_2)\theta_1-(1-p_1)\theta_2\}^4}\right\}$$

$$= \frac{(p_1-p_2)^2}{4(1-T)^2\{(1-p_2)\theta_1-(1-p_1)\theta_2\}^4}\left\{\theta_2^2 Var\left(\widehat{\theta}_1\right) + \theta_1^2 Var\left(\widehat{\theta}_2\right)\right\}$$

$$= \frac{(p_1-p_2)^2\left\{(2P_{y_2}-1)^2\left\{\frac{P_{y_1}(1-P_{y_1})}{n_1}\right\} + (2P_{y_1}-1)^2\left\{\frac{P_{y_2}(1-P_{y_2})}{n_2}\right\}\right\}}{(1-T)^2\{(1-p_2)(2P_{y_1}-1)-(1-p_1)(2P_{y_2}-1)\}^4} = V_\omega \qquad \text{(III.21)}$$

□

Note that Equation (III.21) will be used to calculate the theoretical variance of $\widehat{\omega}_1$ later in Section 5.2.

CHAPTER IV

OPTIMALITY ISSUES

4.1 Optimal Sub-Sample Sizes

Theorem IV.1. *The optimal value of* $\frac{n_1}{n_2}$ *that minimizes* $Var\left(\widehat{\pi}_p\right)$ *is given by*

$$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)} = \frac{1}{\lambda}\sqrt{\frac{P_{y_1}(1-P_{y_1})}{P_{y_2}(1-P_{y_2})}} = \frac{(1-p_2)}{(1-p_1)}\sqrt{\frac{P_{y_1}(1-P_{y_1})}{P_{y_2}(1-P_{y_2})}}, \tag{IV.1}$$

where n_1 *and* n_2 *are sub-sample sizes of* n *with* $n = n_1 + n_2$*.*

Proof. Using $n_1 = n - n_2$, we can rewrite Equation (III.7) as follows.

$$Var(\widehat{\pi}_p) = \frac{1}{(\lambda-1)^2}\left\{\lambda^2\frac{P_{y_2}(1-P_{y_2})}{n_2} + \frac{P_{y_1}(1-P_{y_1})}{n-n_2}\right\} \tag{IV.2}$$

After taking partial derivative on both sides of Equation (IV.2), we get

$$\frac{\partial Var(\widehat{\pi}_p)}{\partial n_2} = \frac{1}{(\lambda-1)^2}\left\{-\lambda^2\frac{P_{y_2}(1-P_{y_2})}{n_2^2} + \frac{P_{y_1}(1-P_{y_1})}{(n-n_2)^2}\right\} = 0 \tag{IV.3}$$

$$\frac{\partial^2 Var(\widehat{\pi}_p)}{\partial n_2^2} = \frac{1}{(\lambda-1)^2}\left\{2\lambda^2\frac{P_{y_2}(1-P_{y_2})}{n_2^3} + 2\frac{P_{y_1}(1-P_{y_1})}{(n-n_2)^3}\right\} > 0 \tag{IV.4}$$

Note that the second derivative of $Var(\widehat{\pi}_p)$ is always positive. Thus, $Var(\widehat{\pi}_p)$ is convex and we have the minimum variance when $\frac{\partial Var(\widehat{\pi}_p)}{\partial n_2} = 0$.

Solving Equation (IV.3) for $\left(\frac{n_1}{n_2}\right)$, we have

$$\lambda^2 \frac{P_{y_2}(1-P_{y_2})}{n_2^2} = \frac{P_{y_1}(1-P_{y_1})}{(n-n_2)^2} \qquad \Longleftrightarrow \qquad \frac{(n-n_2)^2}{n_2^2} = \frac{1}{\lambda^2}\left\{\frac{P_{y_1}(1-P_{y_1})}{P_{y_2}(1-P_{y_2})}\right\}$$

$$\Longleftrightarrow \qquad \frac{n_1^2}{n_2^2} = \frac{1}{\lambda^2}\left\{\frac{P_{y_1}(1-P_{y_1})}{P_{y_2}(1-P_{y_2})}\right\}$$

$$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)} = \frac{1}{\lambda}\sqrt{\frac{P_{y_1}(1-P_{y_1})}{P_{y_2}(1-P_{y_2})}} = \frac{(1-p_2)}{(1-p_1)}\sqrt{\frac{P_{y_1}(1-P_{y_1})}{P_{y_2}(1-P_{y_2})}}.$$

□

Notice that $\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_g)}$ from the Gupta 2001 [4] model in Section 2.3 and $\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$ in Equation (IV.1) look similar, but they are not identical because $(1-T)$ is included in P_{y_2} and P_{y_1} for $\widehat{\pi}_p$ in our model. (See Equations (III.3) and (III.4).)

Theorem IV.2. *The optimal value of* $\frac{n_1}{n_2}$ *that minimizes* $Var(\widehat{\omega}_1)$ *is given by*

$$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)} = \left|\frac{2P_{y_2}-1}{2P_{y_1}-1}\right|\sqrt{\frac{P_{y_1}(1-P_{y_1})}{P_{y_2}(1-P_{y_2})}}, \tag{IV.5}$$

when n_1 *and* n_2 *are sub-sample sizes of* n*, where* $n = n_1 + n_2$ *and the 1st order Taylor approximation of* $\widehat{\omega}_p$ *is used.*

Proof. Using $n_1 = n - n_2$, we can rewrite (III.21) as follows.

$$Var(\widehat{\omega}_1) = \frac{(p_1-p_2)^2\left\{(2P_{y_2}-1)^2\left\{\frac{P_{y_1}(1-P_{y_1})}{n_1}\right\} + (2P_{y_1}-1)^2\left\{\frac{P_{y_2}(1-P_{y_2})}{n-n_1}\right\}\right\}}{(1-T)^2\{(1-p_2)(2P_{y_1}-1)-(1-p_1)(2P_{y_2}-1)\}^4} \tag{IV.6}$$

After taking partial derivative on both sides of Equation (IV.6) with respect to n_1, we have the optimal $\frac{n_1}{n_2}$ ratio which gives the minimum variance of $\widehat{\omega}_1$.

$$\frac{\partial Var\left(\widehat{\omega}_1\right)}{\partial n_1} = -(2P_{y_2}-1)^2\left\{\frac{P_{y_1}(1-P_{y_1})}{n_1^2}\right\} + (2P_{y_1}-1)^2\left\{\frac{P_{y_2}(1-P_{y_2})}{(n-n_1)^2}\right\} = 0 \tag{IV.7}$$

$$\frac{\partial^2 Var\left(\widehat{\omega}_1\right)}{\partial n_1^2} = 2(2P_{y_2}-1)^2\left\{\frac{P_{y_1}(1-P_{y_1})}{n_1^3}\right\} + 2(2P_{y_1}-1)^2\left\{\frac{P_{y_2}(1-P_{y_2})}{(n-n_1)^3}\right\} > 0 \tag{IV.8}$$

Note that the second derivative of $Var\left(\widehat{\omega}_1\right)$ is always positive. Thus, $Var\left(\widehat{\omega}_1\right)$ is convex and we have the minimum variance when $\frac{\partial Var(\widehat{\omega}_1)}{\partial n_1} = 0$.

Solving Equation (IV.7) for $\frac{n_1}{n_2}$, we have

$$(2P_{y_1}-1)^2\left\{\frac{P_{y_2}(1-P_{y_2})}{(n-n_1)^2}\right\} = (2P_{y_2}-1)^2\left\{\frac{P_{y_1}(1-P_{y_1})}{n_1^2}\right\}$$

$$\frac{n_1^2}{(n-n_1)^2} = \frac{(2P_{y_2}-1)^2}{(2P_{y_1}-1)^2}\left\{\frac{P_{y_1}(1-P_{y_1})}{P_{y_2}(1-P_{y_2})}\right\}$$

$$\frac{n_1^2}{n_2^2} = \frac{(2P_{y_2}-1)^2}{(2P_{y_1}-1)^2}\left\{\frac{P_{y_1}(1-P_{y_1})}{P_{y_2}(1-P_{y_2})}\right\}$$

$$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)} = \left|\frac{2P_{y_2}-1}{2P_{y_1}-1}\right|\sqrt{\frac{P_{y_1}(1-P_{y_1})}{P_{y_2}(1-P_{y_2})}}$$

$\square$

Now, let's compare two optimal values of $\frac{n_1}{n_2}$ to see how they behave in a given setting. From Equation (IV.5), we have $\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$ as

$$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)} = \left|\frac{1-2(1-T)(1-p_1)\omega}{1-2(1-T)(1-p_2)\omega}\right| \sqrt{\frac{P_{y_1}(1-P_{y_1})}{P_{y_2}(1-P_{y_2})}} \tag{IV.9}$$

Here in Equation (IV.9), p_1 is in its numerator and p_2 is in the denominator, whereas p_1 and p_2 are located in the opposite way in Equation (IV.1). Thus, these two optimal values are quite opposite of each other and behave like reciprocal of each other. This will be shown later in Section 5.3 in detail. The estimation of $\widehat{\pi}$ and its variance are the most important task because they are directly related to the survey question. Throughout Chapter V, the optimal values of $\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$ were chosen first. Then n_1 and n_2 were determined accordingly and used in each simulation.

4.2 Optimal Value of the Two-Stage Parameter (T)

In this section, we study how $Var(\widehat{\pi}_p)$ behaves with respect to T in the proposed model.

Theorem IV.3. *$Var(\widehat{\pi}_p)$ is maximum, when $T = T^*$,*

$$\textit{where} \quad T^* = \frac{n_1\lambda\left\{2(1-p_2)\omega - 1\right\} + n_2\left\{2(1-p_1)\omega - 1\right\}}{2\omega n(1-p_1)}.$$

Proof. Expanding the variance formula (III.7) with P_{y_1} and P_{y_2} from Equations (III.3) and (III.4), we have,

$$Var(\widehat{\pi}_p) = \frac{1}{(\lambda-1)^2}\left\{\lambda^2\frac{P_{y_2}(1-P_{y_2})}{n_2} + \frac{P_{y_1}(1-P_{y_1})}{n_1}\right\}$$

$$= \frac{1}{(\lambda-1)^2}\frac{\lambda^2}{n_2}\left\{(1-T)(p_2-1)(2\pi-1)\omega+\pi\right\}\left\{1-(1-T)(p_2-1)(2\pi-1)\omega-\pi\right\}$$

$$+\frac{1}{(\lambda-1)^2}\frac{1}{n_1}\left\{(1-T)(p_1-1)(2\pi-1)\omega+\pi\right\}\left\{1-(1-T)(p_1-1)(2\pi-1)\omega-\pi\right\} \quad \text{(IV.10)}$$

The coefficient of T^2 is negative.

$$-\frac{1}{(\lambda-1)^2}\left\{\frac{\lambda^2}{n_2^2}(1-p_2)^2(2\pi-1)^2\omega^2+\frac{1}{n_1^2}(1-p_1)^2(2\pi-1)^2\omega^2\right\}<0 \quad \text{(IV.11)}$$

Thus, $\left(\frac{\partial^2 Var(\widehat{\pi}_p)}{\partial T^2}\right)_{T=T^*} < 0$ for the quadratic equation of T. And the right-hand side of Equation (IV.10) is concave. So T^*, which satisfies $\left(\frac{\partial Var(\widehat{\pi}_p)}{\partial T}\right)_{T=T^*} = 0$, gives the maximum value.

Setting the first partial derivative of $Var(\widehat{\pi}_p)$ with respect to T in Equation (IV.10) equal to zero, we get

$$
\begin{aligned}
\frac{\partial Var(\widehat{\pi})}{\partial T} &= \frac{1}{(\lambda-1)^2}\frac{\lambda^2}{n_2}\left\{-(p_2-1)(2\pi-1)\omega\right\}\left\{1-(1-T)(p_2-1)(2\pi-1)\omega-\pi\right\} \\
&+\frac{1}{(\lambda-1)^2}\frac{\lambda^2}{n_2}\left\{(1-T)(p_2-1)(2\pi-1)\omega+\pi\right\}\left\{(p_2-1)(2\pi-1)\omega\right\} \\
&+\frac{1}{(\lambda-1)^2}\frac{1}{n_1}\left\{-(p_1-1)(2\pi-1)\omega\right\}\left\{1-(1-T)(p_1-1)(2\pi-1)\omega-\pi\right\} \\
&+\frac{1}{(\lambda-1)^2}\frac{1}{n_1}\left\{(1-T)(p_1-1)(2\pi-1)\omega+\pi\right\}\left\{(p_1-1)(2\pi-1)\omega\right\}=0 \\
\iff \quad & \frac{\lambda^2(p_2-1)}{n_2}\left\{-1+2(1-T)(p_2-1)(2\pi-1)\omega+2\pi\right\} \\
&+\frac{(p_1-1)}{n_1}\left\{-1+2(1-T)(p_1-1)(2\pi-1)\omega+2\pi\right\}=0 \\
\iff \quad & \frac{\lambda^2(p_2-1)}{n_2}\left\{2(1-T)(p_2-1)(2\pi-1)\omega+(2\pi-1)\right\} \\
&+\frac{(p_1-1)}{n_1}\left\{2(1-T)(p_1-1)(2\pi-1)\omega+(2\pi-1)\right\}=0 \\
\iff \quad & \frac{\lambda^2(p_2-1)}{n_2}(2\pi-1)\left\{2(1-T)(p_2-1)\omega+1\right\}+\frac{(p_1-1)}{n_1}(2\pi-1)\left\{2(1-T)(p_1-1)\omega+1\right\}=0 \\
\iff \quad & \frac{\lambda^2(p_2-1)}{n_2}\left\{2(1-T)(p_2-1)\omega+1\right\}+\frac{(p_1-1)}{n_1}\left\{2(1-T)(p_1-1)\omega+1\right\}=0 \\
\iff \quad & T=\frac{\frac{\lambda}{n_2}+\frac{1}{n_1}}{2\omega\left\{\frac{\lambda}{n_2}(p_2-1)+\frac{1}{n_1}(p_1-1)\right\}}+1 \quad = \quad T^* \qquad \text{(IV.12)}
\end{aligned}
$$

Simplifying further, we have T^* given by

$$T^* = \frac{\frac{\lambda}{n_2} + \frac{1}{n_1}}{2\omega\left\{\frac{\lambda}{n_2}(p_2-1) + \frac{1}{n_1}(p_1-1)\right\}} + 1 = \frac{\frac{\lambda}{n_2} + \frac{1}{n_1}}{2\omega\left\{\frac{1}{n_2}\frac{(p_1-1)}{(p_2-1)}(p_2-1) + \frac{1}{n_1}(p_1-1)\right\}} + 1$$

$$= \frac{\frac{\lambda}{n_2} + \frac{1}{n_1}}{2\omega\left\{\frac{1}{n_2}(p_1-1) + \frac{1}{n_1}(p_1-1)\right\}} + 1 = \frac{\frac{\lambda}{n_2} + \frac{1}{n_1}}{2\omega\left\{\frac{1}{n_2} + \frac{1}{n_1}\right\}(p_1-1)} + 1$$

$$= \frac{n_1\lambda + n_2}{2\omega\left\{n_1 + n_2\right\}(p_1-1)} + 1 = \frac{n_1\lambda + n_2 + 2\omega n_1(p_1-1) + 2\omega n_2(p_1-1)}{2\omega n(p_1-1)}$$

$$= \frac{n_1\lambda + n_2 + 2\omega n_1\lambda(p_2-1) + 2\omega n_2(p_1-1)}{2\omega n(p_1-1)} = \frac{n_1\lambda\left\{2(p_2-1)\omega+1\right\} + n_2\left\{2(p_1-1)\omega+1\right\}}{2\omega n(p_1-1)}$$

$$= \frac{n_1\lambda\left\{2(1-p_2)\omega-1\right\} + n_2\left\{2(1-p_1)\omega-1\right\}}{2\omega n(1-p_1)} \tag{IV.13}$$

□

Note that

$$T^* \geq 0 \qquad \text{when} \quad \left\{n_1\lambda\left\{2(1-p_2)\omega-1\right\} + n_2\left\{2(1-p_1)\omega-1\right\}\right\} \geq 0 \tag{IV.14}$$

$$T^* < 0 \qquad \text{when} \quad \left\{n_1\lambda\left\{2(1-p_2)\omega-1\right\} + n_2\left\{2(1-p_1)\omega-1\right\}\right\} < 0. \tag{IV.15}$$

Lemma IV.4. *If* $Var(\widehat{\pi}_p) = Var(\widehat{\pi}_g)$,

$$T = 0 \qquad or \qquad T = \frac{n_1\lambda\big\{2(1-p_2)\omega-1\big\} + n_2\big\{2(1-p_1)\omega-1\big\}}{\omega n(1-p_1)} = T_b$$

Proof. Notice that $Var(\widehat{\pi}_g) = Var(\widehat{\pi}_p)|_{T=0}$ because the proposed model is nothing but the Gupta 2001 [4] model when $T = 0$.

Also, from Equations (II.20) and (IV.2), we have,

$$Var(\widehat{\pi}_g) = \frac{1}{(\lambda-1)^2}\left\{\lambda^2\frac{P_{y_2|g}(1-P_{y_2|g})}{n_2} + \frac{P_{y_1|g}(1-P_{y_1|g})}{n_1}\right\}$$

$$Var(\widehat{\pi}_p) = \frac{1}{(\lambda-1)^2}\left\{\lambda^2\frac{P_{y_2|p}(1-P_{y_2|p})}{n_2} + \frac{P_{y_1|p}(1-P_{y_1|p})}{n_1}\right\}.$$

From Equations (II.16), we have the following for $Var(\widehat{\pi}_g)$.

$$P_{y_1|g} = (p_1-1)(2\pi-1)\omega+\pi, \qquad \text{and} \qquad P_{y_2|g} = (p_2-1)(2\pi-1)\omega+\pi.$$

From Equations (III.3) and (III.4), we have the following for $Var(\widehat{\pi}_p)$.

$$P_{y_1|p} = (1-T)(p_1-1)(2\pi-1)\omega+\pi, \quad \text{and} \quad P_{y_2|p} = (1-T)(p_2-1)(2\pi-1)\omega+\pi.$$

$$\text{If} \quad \beta_1 = (p_1-1)(2\pi-1)\omega \quad \text{and} \quad \beta_2 = (p_2-1)(2\pi-1)\omega, \tag{IV.16}$$

the variances $Var(\widehat{\pi}_g)$ and $Var(\widehat{\pi}_p)$ from Equations (II.20) and (IV.2) can be expressed as follows.

$$Var(\widehat{\pi}_g) = Var(\widehat{\pi}_p)|_{T=0} = \frac{1}{(\lambda-1)^2}\left\{\frac{\lambda^2}{n_2}(\pi+\beta_2)(1-\pi-\beta_2) + \frac{1}{n_1}(\pi+\beta_1)(1-\pi-\beta_1)\right\} \tag{IV.17}$$

$$Var(\widehat{\pi}_p) = \frac{1}{(\lambda-1)^2}\left\{\frac{\lambda^2(\pi+(1-T)\beta_2)(1-\pi-(1-T)\beta_2)}{n_2} + \frac{(\pi+(1-T)\beta_1)(1-\pi-(1-T)\beta_1)}{n_1}\right\} \tag{IV.18}$$

Plugging in Equations (IV.17) and (IV.18) into $Var(\widehat{\pi}_p) = Var(\widehat{\pi}_g)$, we have

$$\frac{1}{(\lambda-1)^2}\left\{\frac{\lambda^2(\pi+(1-T)\beta_2)(1-\pi-(1-T)\beta_2)}{n_2}+\frac{(\pi+(1-T)\beta_1)(1-\pi-(1-T)\beta_1)}{n_1}\right\}$$

$$= \quad \frac{1}{(\lambda-1)^2}\left\{\frac{\lambda^2}{n_2}(\pi+\beta_2)(1-\pi-\beta_2)+\frac{1}{n_1}(\pi+\beta_1)(1-\pi-\beta_1)\right\}.$$

Reorganizing this, we have

$$\frac{\lambda^2(\pi+(1-T)\beta_2)(1-\pi-(1-T)\beta_2)}{n_2}+\frac{(\pi+(1-T)\beta_1)(1-\pi-(1-T)\beta_1)}{n_1}$$

$$= \quad \frac{\lambda^2}{n_2}(\pi+\beta_2)(1-\pi-\beta_2)+\frac{1}{n_1}(\pi+\beta_1)(1-\pi-\beta_1)$$

$$\Longleftrightarrow \lambda^2 n_1\big(\pi+(1-T)\beta_2\big)\big(1-\pi-(1-T)\beta_2\big)+n_2\big(\pi+(1-T)\beta_1\big)\big(1-\pi-(1-T)\beta_1\big)$$

$$= \quad \lambda^2 n_1(\pi+\beta_2)(1-\pi-\beta_2)+n_2(\pi+\beta_1)(1-\pi-\beta_1)$$

$$\Longleftrightarrow \lambda^2 n_1(\pi+\beta_2-\beta_2T)(1-\pi-\beta_2+\beta_2T)+n_2(\pi+\beta_1-\beta_1T)(1-\pi-\beta_1+\beta_1T)$$

$$= \quad \lambda^2 n_1(\pi+\beta_2)(1-\pi-\beta_2)+n_2(\pi+\beta_1)(1-\pi-\beta_1)$$

$$\Longleftrightarrow \lambda^2 n_1\big((\pi+\beta_2)(1-\pi-\beta_2)+\beta_2(\pi+\beta_2)T+\beta_2(\pi+\beta_2-1)T-\beta_2^2T^2\big)$$

$$+ \quad n_2\big((\pi+\beta_1)(1-\pi-\beta_1)+\beta_1(\pi+\beta_1)T+\beta_1(\pi+\beta_1-1)T-\beta_1^2T^2\big)$$

$$= \quad \lambda^2 n_1(\pi+\beta_2)(1-\pi-\beta_2)+n_2(\pi+\beta_1)(1-\pi-\beta_1)$$

$$\Longleftrightarrow \lambda^2 n_1\big(\beta_2(\pi+\beta_2)T+\beta_2(\pi+\beta_2-1)T-\beta_2^2T^2\big)+n_2\big(\beta_1(\pi+\beta_1)T+\beta_1(\pi+\beta_1-1)T-\beta_1^2T^2\big)=0$$

Thus, the equation becomes

$$\lambda^2 n_1\big(\beta_2(2\pi + 2\beta_2 - 1)T - \beta_2^2 T^2\big) + n_2\big(\beta_1(2\pi + 2\beta_1 - 1)T - \beta_1^2 T^2\big) = 0. \tag{IV.19}$$

Factoring the left hand side of (IV.19), we have

$$T\left\{(\lambda^2 n_1 \beta_2^2 + n_2\beta_1^2)T - \lambda^2 n_1\beta_2(2\pi + 2\beta_2 - 1) - n_2\beta_1(2\pi + 2\beta_1 - 1)\right\} = 0 \tag{IV.20}$$

The solutions are

$$T = 0 \qquad \text{or,} \tag{IV.21}$$

$$T = \frac{\lambda^2 n_1\beta_2(2\pi + 2\beta_2 - 1) + n_2\beta_1(2\pi + 2\beta_1 - 1)}{\lambda^2 n_1\beta_2^2 + n_2\beta_1^2} = T_b \tag{IV.22}$$

Plugging in β_1 and β_1 from Equations (IV.16) into Equation (IV.22), T_b will be,

$$T_b = \frac{\lambda^2 n_1(p_2-1)(2\pi-1)\omega\Big(2(p_2-1)(2\pi-1)\omega + 2\pi - 1\Big) + n_2(p_1-1)(2\pi-1)\omega\Big(2(p_1-1)(2\pi-1)\omega + 2\pi - 1\Big)}{\lambda^2 n_1(p_2-1)^2(2\pi-1)^2\omega^2 + n_2(p_1-1)^2(2\pi-1)^2\omega^2}.$$

Dividing out the common factor of $(2\pi - 1)^2\omega$, we get

$$\begin{aligned} T_b &= \frac{\lambda^2 n_1(p_2-1)\{2(p_2-1)(2\pi-1)\omega + 2\pi - 1\} + n_2(p_1-1)\{2(p_1-1)(2\pi-1)\omega + 2\pi - 1\}}{\lambda^2 n_1(p_2-1)^2(2\pi-1)\omega + n_2(p_1-1)^2(2\pi-1)\omega} \\ &= \frac{\lambda^2 n_1(p_2-1)(2\pi-1)\{2(p_2-1)\omega + 1\} + n_2(p_1-1)(2\pi-1)\{2(p_1-1)\omega + 1\}}{\lambda^2 n_1(p_2-1)^2(2\pi-1)\omega + n_2(p_1-1)^2(2\pi-1)\omega} \\ &= \frac{\lambda^2 n_1(p_2-1)\{2(p_2-1)\omega + 1\} + n_2(p_1-1)\{2(p_1-1)\omega + 1\}}{\lambda^2 n_1(p_2-1)^2\omega + n_2(p_1-1)^2\omega}. \end{aligned}$$

Using $\lambda = \frac{(p_1-1)}{(p_2-1)}$, it can be further simplified to

$$T_b = \frac{\lambda \frac{(p_1-1)}{(p_2-1)} n_1(p_2-1)\big\{2(p_2-1)\omega+1\big\} + n_2(p_1-1)\big\{2(p_1-1)\omega+1\big\}}{\frac{(p_1-1)^2}{(p_2-1)^2} n_1(p_2-1)^2\omega + n_2(p_1-1)^2\omega}$$

$$= \frac{\lambda n_1(p_1-1)\big\{2(p_2-1)\omega+1\big\} + n_2(p_1-1)\big\{2(p_1-1)\omega+1\big\}}{\omega\big\{n_1+n_2\big\}(p_1-1)^2}$$

Thus, we have

$$T_b = \frac{n_1\lambda\big\{2(1-p_2)\omega-1\big\} + n_2\big\{2(1-p_1)\omega-1\big\}}{\omega n(1-p_1)}. \tag{IV.23}$$

□

Lemma IV.5.

$$2T^* = T_b$$

Proof. From Equations (IV.13) and (IV.23), we have

$$2T^* = 2\frac{n_1\lambda\big\{2(1-p_2)\omega-1\big\} + n_2\big\{2(1-p_1)\omega-1\big\}}{2\omega n(1-p_1)}$$

$$= \frac{n_1\lambda\big\{2(1-p_2)\omega-1\big\} + n_2\big\{2(1-p_1)\omega-1\big\}}{\omega n(1-p_1)}$$

$$= T_b. \tag{IV.24}$$

□

Lemma IV.6. *$T_b < 1$,*

when $0 < p_1 < 1$, $0 < p_2 < 1$, $0 < \omega < 1$, $0 < \frac{n_1}{n} < 1$, and $0 < \frac{n_2}{n} < 1$.

Proof. From Equation (IV.23), T_b can be rewritten as

$$
\begin{aligned}
T_b &= \frac{n_1\lambda\{2(p_2-1)\omega+1\}+n_2\{2(p_1-1)\omega+1\}}{\omega n(p_1-1)} \\
&= \frac{2n_1\lambda(p_2-1)\omega+n_1\lambda+2n_2(p_1-1)\omega+n_2}{\omega n(p_1-1)} \\
&= \frac{2n_1}{n}+\frac{n_1}{n(p_2-1)\omega}+\frac{2n_2}{n}+\frac{n_2}{n(p_1-1)\omega} \\
&= 2+\frac{n_1}{n(p_2-1)\omega}+\frac{n_2}{n(p_1-1)\omega} \\
&= 2-\frac{1}{\omega}\left\{\frac{n_1}{n(1-p_2)}+\frac{n_2}{n(1-p_1)}\right\} \\
&= 2-\frac{1}{\omega}\left\{\frac{n_1/n}{(1-p_2)}+\frac{1-n_1/n}{(1-p_1)}\right\} \\
&= 2-\frac{1}{\omega}\left\{\frac{n_1/n}{(1-p_2)}-\frac{n_1/n}{(1-p_1)}+\frac{1}{(1-p_1)}\right\} \\
&= 2-\frac{1}{\omega}\left\{\frac{(p_2-p_1)}{(1-p_2)(1-p_1)}\left(\frac{n_1}{n}\right)+\frac{1}{(1-p_1)}\right\}. \qquad \text{(IV.25)}
\end{aligned}
$$

When the quantity inside of the bracket in Equation (IV.25) has the smallest value, we have the maximum value of T_b, given by

$$
\mathrm{Max}\left[T_b\right] = 2-\mathrm{Min}\left[\frac{1}{\omega}\left\{\frac{(p_2-p_1)(n_1/n)}{(1-p_2)(1-p_1)}+\frac{1}{(1-p_1)}\right\}\right] = 2-1=1 \qquad \text{(IV.26)}
$$

□

A slightly different but essentially the same proof can be found in Theorem 4 of Gupta et al 2012 [6].

By Equations (IV.14) and (IV.15) and Lemma IV.5, $T_b \geq 0$ or $T_b < 0$.

(1) If $T_b \geq 0$, using Lemma IV.6, we have $0 \leq T_b < 1$. Thus, we can choose a T such that $T_b < T < 1$. This is the case with the black parabola in Figure 1. When $T_b < T$, $Var(\widehat{\pi}_p) < Var(\widehat{\pi}_g)$.

(2) If $T_b < 0$, then $T_b < 0 < T < 1$. Thus, $Var(\widehat{\pi}_p) < Var(\widehat{\pi}_g)$ holds true for all T, $0 < T < 1$. This is the case with the left-side red parabola in Figure 1.

In either case, for every value of T_b, we can find a T such that T satisfies $Var(\widehat{\pi}_p) < Var(\widehat{\pi}_g)$. Thus, for all T_b, there is T, which is $0 < T < 1$, that satisfies $T_b < T < 1$ and $Var(\widehat{\pi}_p) < Var(\widehat{\pi}_g)$.

Notice that, in Figure 1, T^* is shown to be positive or negative and $Var(\widehat{\pi}_p)|_{T=T^*}$ is the maximum value as explained in Theorem IV.3 and Equations (IV.14) and (IV.15). The two solutions—0 and T_b—satisfying $Var(\widehat{\pi}_p) = Var(\widehat{\pi}_g)$ are illustrated accordingly as proved in Lemma IV.4. As $Var(\widehat{\pi}_p) = Var(\widehat{\pi}_g)$ leads to a quadratic equation, $2T^* = T_b$ can be easily verified in Figure 1. Notice also that $T_b < 1$ in Figure 1 as proved in Lemma IV.6. Thus, as explained above, a meaningful value of T which satisfies $(T_b < T < 1)$ can always be chosen. A slightly different but essentially the same proof—but for a different model—can be found in Theorem 5 of Gupta et al 2012 [6].

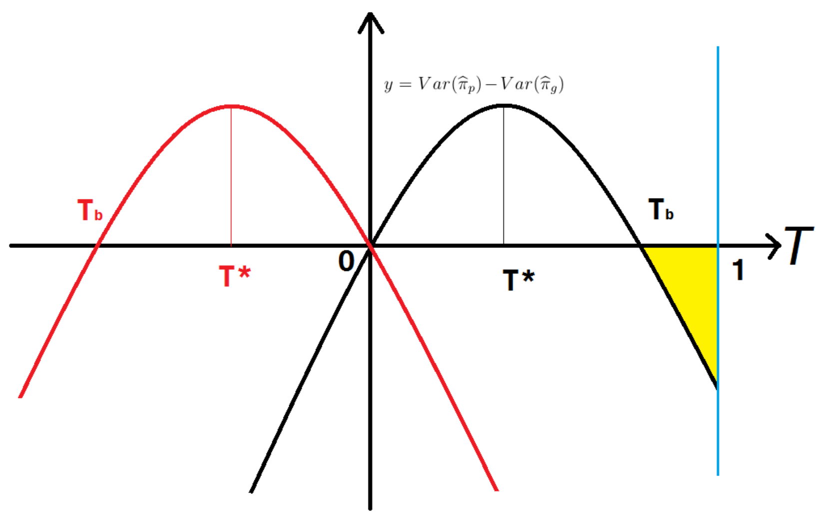

Figure 1: $(T = 0)$, $(2T^{*} = T_b)$ and $(T_b < 1)$

CHAPTER V

SIMULATION RESULTS

In Sections 5.1 and 5.2, simulation results are presented for the two main parameters (π and ω). The optimal values of $\frac{n_1}{n_2}$ are presented in Section 5.3. For the theorerical values of $Var\left(\widehat{\omega}_p\right)$, the first order Talyor's approximation was used. All the simulations in this thesis were conducted in the R programming language (R Development Core Team, 2012 [17]). Three parameters—T, π, and ω—were allowed to vary while all the other variables were fixed; $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and 1000 trials per simulation. Then, n_1 and n_2 were chosen to minimize the variance of $\widehat{\pi}_p$ for each case. An R code for the proposed model is attached in Appendix A.

The first thing to verify is that simulated values for $\widehat{\pi}_p$ and $\widehat{\omega}_p$ provide good estimates for the respective parameters. Second, simulated variances of $\widehat{\pi}_p$ and $\widehat{\omega}_p$ are examined to see if they are close to the theoretical variances. Third, with the help of statistical software simulation, we examine normality of $\widehat{\pi}_p$, $\widehat{\omega}_p$, and $\widehat{\omega}_1$, as we already proved asymptotic normality of $\widehat{\pi}_p$ and $\widehat{\omega}_1$ in Theorems III.1 and III.2. Lastly, the optimal values of $\frac{n_1}{n_2}$ with respect to $\widehat{\pi}_p$ and $\widehat{\omega}_1$ are presented in Table 9.

5.1 Simulation of $E\left(\widehat{\pi}_p\right)$ and $Var\left(\widehat{\pi}_p\right)$

For the simulation of $E\left(\widehat{\pi}_p\right)$ and $Var\left(\widehat{\pi}_p\right)$, the value of T varied from 0.1 to 0.5 in steps of 0.1, while ω varied from 0.1 to 0.9 in increments of 0.2. π values were selected to be 0.1, 0.2, 0.3, and 0.8. Table 1 is for $\pi = 0.1$, Table 2 for $\pi = 0.2$, Table 3 for

$\pi = 0.3$, and Table 4 for $\pi = 0.8$. In practice, π must be quite small because it is the proportion of a population which belongs to a specific group that entails certain degree of sensitivity. Nonetheless, Table 4 for $\pi = 0.8$ is included for the sake of mathematical model.

Simulated values of $E(\widehat{\pi}_p)$ and $Var(\widehat{\pi}_p)$ are very close to corresponding true parameter value of π and theoretical value of $Var(\widehat{\pi}_p)$. This is clearly shown in Tables 1, 2, 3, and 4.

As for the normality of $\widehat{\pi}_p$, there is little evidence that the samples of $\widehat{\pi}_p$ are from non-normal distributions. All the p-values are greater than 0.01 in Tables 1, 2, 3, and 4 and we have only a few p-values between 0.01 and 0.05. As we proved that $\widehat{\pi}_p$ is asymptotically nomal in Theorem III.1, the number of p-values that are less than 0.05 will decrease as n increases further. Out of 100 runs, only 6 cases have a p-value of less than 0.05. And there is no p-value less than 0.01 for $\widehat{\pi}_p$. Although we used a total sample size of 1000 for our simulation study, we did experiment with smaller total sample sizes of 800 and 600, and the results of the normality of $\widehat{\pi}_p$ were similar.

Notice that n_1 and n_2 are calculated for each case to have minimum variance of $\widehat{\pi}_p$. Thus, n_1 and n_2 are all different for each case.

Table 1: Simulation Results for $\widehat{\pi}_p$ ($\pi = 0.1$)

$\pi = 0.1$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.1$	0.1	0.2	0.3	0.4	0.5
0.1	Simulated $Mean\,(\widehat{\pi}_p)$	0.1000	0.1000	0.1010	0.1005	0.0993
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.5036	0.0193	0.5553	0.6697	0.3289
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0148	0.0142	0.0150	0.0141	0.0137
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0145	0.0144	0.0143	0.0142	0.0141
	Optimal n_1	829	830	832	834	837
	Optimal n_2	171	170	168	166	163
0.3	Simulated $Mean\,(\widehat{\pi}_p)$	0.1000	0.1001	0.1000	0.1000	0.1001
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.8416	0.6815	0.5242	0.6726	0.3472
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0165	0.0160	0.0160	0.0153	0.0146
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0161	0.0158	0.0156	0.0154	0.0151
	Optimal n_1	810	812	814	817	820
	Optimal n_2	190	188	186	183	180
0.5	Simulated $Mean\,(\widehat{\pi}_p)$	0.1006	0.1006	0.0986	0.0994	0.0999
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.0847	0.3429	0.2622	0.1888	0.1083
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0179	0.0170	0.0168	0.0167	0.0161
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0172	0.0169	0.0166	0.0163	0.0159
	Optimal n_1	806	806	807	809	811
	Optimal n_2	194	194	193	191	189

(Continued on next page.)

$\pi = 0.1$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.1$	0.1	0.2	0.3	0.4	0.5
0.7	Simulated $Mean\,(\widehat{\pi}_p)$	0.1000	0.0999	0.1004	0.1007	0.1009
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.4427	0.2335	0.2330	0.4422	0.4753
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0178	0.0176	0.0174	0.0172	0.0162
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0180	0.0177	0.0174	0.0170	0.0166
	Optimal n_1	812	809	807	806	807
	Optimal n_2	188	191	193	194	193
0.9	Simulated $Mean\,(\widehat{\pi}_p)$	0.0999	0.1011	0.0992	0.1002	0.0995
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.7997	0.2142	0.6837	0.2039	0.3914
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0184	0.0183	0.0178	0.0181	0.0167
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0185	0.0183	0.0180	0.0176	0.0172
	Optimal n_1	825	818	812	808	806
	Optimal n_2	175	182	188	192	194

Table 2: Simulation Results for $\widehat{\pi}_p$ ($\pi = 0.2$)

$\pi = 0.2$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.2$	0.1	0.2	0.3	0.4	0.5
0.1	Simulated $Mean\,(\widehat{\pi}_p)$	0.1999	0.1991	0.1994	0.2006	0.1993
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.2117	0.5547	0.3494	0.5250	0.7446
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0188	0.0184	0.0191	0.0189	0.0185
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0185	0.0185	0.0184	0.0184	0.0183
	Optimal n_1	842	843	844	845	845
	Optimal n_2	158	157	156	155	155
0.3	Simulated $Mean\,(\widehat{\pi}_p)$	0.2004	0.2010	0.1994	0.2000	0.2005
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.5224	0.9052	0.5283	0.8546	0.3197
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0191	0.0193	0.0186	0.0190	0.0186
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0192	0.0191	0.0190	0.0189	0.0188
	Optimal n_1	833	834	836	837	838
	Optimal n_2	167	166	164	163	162
0.5	Simulated $Mean\,(\widehat{\pi}_p)$	0.1990	0.2002	0.2002	0.1995	0.2001
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.8803	0.1244	0.4427	0.6523	0.1496
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0197	0.0198	0.0200	0.0198	0.0193
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0198	0.0196	0.0195	0.0193	0.0192
	Optimal n_1	830	831	831	832	834
	Optimal n_2	170	169	169	168	166

(Continued on next page.)

$\pi = 0.2$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

ω	$\pi = 0.2$	T: 0.1	0.2	0.3	0.4	0.5
0.7	Simulated $Mean\,(\widehat{\pi}_p)$	0.1996	0.1991	0.1998	0.2004	0.2001
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.9666	0.7086	0.5608	0.8158	0.8659
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0198	0.0198	0.0196	0.0195	0.0203
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0202	0.0200	0.0199	0.0197	0.0195
	Optimal n_1	832	831	830	830	831
	Optimal n_2	168	169	170	170	169
0.9	Simulated $Mean\,(\widehat{\pi}_p)$	0.2002	0.2007	0.1987	0.2004	0.1994
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.0113	0.0683	0.4936	0.9697	0.8176
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0204	0.0201	0.0197	0.0198	0.0197
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0204	0.0203	0.0202	0.0200	0.0198
	Optimal n_1	838	835	832	831	830
	Optimal n_2	162	165	168	169	170

Table 3: Simulation Results for $\widehat{\pi}_p$ ($\pi = 0.3$)

$\pi = 0.3$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.3$	0.1	0.2	0.3	0.4	0.5
0.1	Simulated $Mean\,(\widehat{\pi}_p)$	0.3010	0.3003	0.2999	0.3002	0.3000
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.0798	0.3527	0.5406	0.4402	0.6443
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0211	0.0206	0.0217	0.0206	0.0204
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0209	0.0209	0.0208	0.0208	0.0208
	Optimal n_1	847	848	848	848	848
	Optimal n_2	153	152	152	152	152
0.3	Simulated $Mean\,(\widehat{\pi}_p)$	0.3004	0.2990	0.2987	0.2994	0.2989
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.7040	0.5951	0.1879	0.5741	0.2431
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0214	0.0217	0.0212	0.0218	0.0211
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0212	0.0211	0.0211	0.0210	0.0210
	Optimal n_1	844	844	845	845	846
	Optimal n_2	156	156	155	155	154
0.5	Simulated $Mean\,(\widehat{\pi}_p)$	0.2999	0.3008	0.3003	0.2995	0.2987
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.0465	0.7464	0.7525	0.7084	0.8284
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0225	0.0204	0.0215	0.0218	0.0211
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0214	0.0213	0.0213	0.0212	0.0211
	Optimal n_1	842	842	843	843	844
	Optimal n_2	158	158	157	157	156

(Continued on next page.)

$\pi = 0.3$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.3$	0.1	0.2	0.3	0.4	0.5
0.7	Simulated $Mean\,(\widehat{\pi}_p)$	0.3002	0.3005	0.2997	0.2997	0.3002
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.8269	0.5882	0.7657	0.4592	0.3040
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0212	0.0219	0.0220	0.0215	0.0215
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0215	0.0215	0.0214	0.0214	0.0213
	Optimal n_1	843	842	842	842	843
	Optimal n_2	157	158	158	158	157
0.9	Simulated $Mean\,(\widehat{\pi}_p)$	0.3003	0.2996	0.2995	0.3008	0.3016
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.8752	0.8666	0.0976	0.5584	0.1954
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0214	0.0204	0.0222	0.0216	0.0223
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0216	0.0216	0.0215	0.0215	0.0214
	Optimal n_1	845	844	843	842	842
	Optimal n_2	155	156	157	158	158

Table 4: Simulation Results for $\widehat{\pi}_p$ ($\pi = 0.8$)

$\pi = 0.8$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.8$	0.1	0.2	0.3	0.4	0.5
0.1	Simulated $Mean\,(\widehat{\pi}_p)$	0.7997	0.7999	0.7994	0.8002	0.8005
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.6968	0.9073	0.2319	0.0969	0.2437
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0186	0.0184	0.0185	0.0178	0.0183
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0185	0.0185	0.0184	0.0184	0.0183
	Optimal n_1	842	843	844	845	845
	Optimal n_2	158	157	156	155	155
0.3	Simulated $Mean\,(\widehat{\pi}_p)$	0.7995	0.8002	0.8000	0.8000	0.7997
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.0210	0.8930	0.2113	0.1912	0.9710
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0192	0.0186	0.0185	0.0193	0.0196
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0192	0.0191	0.0190	0.0189	0.0188
	Optimal n_1	833	834	836	837	838
	Optimal n_2	167	166	164	163	162
0.5	Simulated $Mean\,(\widehat{\pi}_p)$	0.8002	0.8000	0.7997	0.7999	0.8006
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.6400	0.2635	0.9466	0.7093	0.4467
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0204	0.0198	0.0200	0.0192	0.0188
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0198	0.0196	0.0195	0.0193	0.0192
	Optimal n_1	830	831	831	832	834
	Optimal n_2	170	169	169	168	166

(Continued on next page.)

$\pi = 0.8$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.8$	0.1	0.2	0.3	0.4	0.5
0.7	Simulated $Mean\,(\widehat{\pi}_p)$	0.8010	0.8004	0.7997	0.8003	0.7998
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.8112	0.5191	0.1500	0.2790	0.0430
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0205	0.0200	0.0199	0.0191	0.0192
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0202	0.0200	0.0199	0.0197	0.0195
	Optimal n_1	832	831	830	830	831
	Optimal n_2	168	169	170	170	169
0.9	Simulated $Mean\,(\widehat{\pi}_p)$	0.7992	0.7998	0.7997	0.7995	0.8006
	p-value of Shapiro-Wilk Normality Test on $\widehat{\pi}_p$	0.6447	0.5576	0.5268	0.0140	0.9122
	Simulated $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0200	0.0204	0.0198	0.0211	0.0197
	Theoretical $\sqrt{Var\,(\widehat{\pi}_p)}$	0.0204	0.0203	0.0202	0.0200	0.0198
	Optimal n_1	838	835	832	831	830
	Optimal n_2	162	165	168	169	170

5.2 Simulation of $E(\widehat{\omega}_p)$ and $Var(\widehat{\omega}_p)$

In this section, the setup for simulation is the same as in Secion 5.1. Three parameters—T, π, and ω—were allowed to vary while all the other variables were fixed; $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and 1000 trials per simulation. Then, n_1 and n_2 were chosen to minimize the variance of $\widehat{\pi}_p$ for each case. The theoretical values of the variance of $\widehat{\omega}_p$ are estimated by the first order Taylor approximation.

For the simulation of $E(\widehat{\omega}_p)$, $E(\widehat{\omega}_1)$, $Var(\widehat{\omega}_p)$, and $Var(\widehat{\omega}_1)$, the value of T varied from 0.1 to 0.5 in steps of 0.1, while ω varying from 0.1 to 0.9 in increments of 0.2. π values are 0.1, 0.2, 0.3 and 0.8. Table 5 is for $\pi = 0.1$, Table 6 for $\pi = 0.2$, Table 7 for $\pi = 0.3$, and Table 8 for $\pi = 0.8$.

As indicated in Section 3.3, empirical means are very close to corresponding true parameter value of ω. Simulated values of $E(\widehat{\omega}_p)$ and $E(\widehat{\omega}_1)$ are very close to ω as shown in Tables 5, 6, 7, and 8. It can be shown that T causes bigger bias for $E(\widehat{\omega}_p)$ and $E(\widehat{\omega}_1)$ as T approaches 1 further. However, in practice T will remain small. Theoretical values of $Var(\widehat{\omega}_1)$ are in agreement with simulated values of $Var(\widehat{\omega}_p)$ and $Var(\widehat{\omega}_1)$. One may note that $Var(\widehat{\omega}_1)$ increases as T increases, but that is only natural since the sample pool which provides randomized responses shrinks as T increases.

In Theorems III.2, we proved that $\widehat{\omega}_1$ is asymptotically normal. In Tables 5, 6, 7, and 8, only 4 out of 100 simulations display p-values less than 0.05 for the Shapiro-Wilk normality test on $\widehat{\omega}_1$. Also notice that we don't have any p-value less than 0.01 for $\widehat{\omega}_1$, indicating there is not much evidence that the samples of $\widehat{\omega}_1$ are from non-normal distributions. For $\widehat{\omega}_p$, the p-values in Tables 5, 6, 7, and 8 show strong evidence of non-normality as π approaches 0.5 and T approaches 1. However, there is no longer evidence of non-normality of $\widehat{\omega}_p$ if we assign large number to n ($n = 10^5$ will suffice.).

Also notice that n_1 and n_2 are calculated to have minimum variance of $\widehat{\pi}_p$—not $\widehat{\omega}_1$—as estimating $\widehat{\pi}_p$ is of greater importance than estimating any other parameters.

Table 5: Simulation Results for $\widehat{\omega}_p$ ($\pi = 0.1$)

$\pi = 0.1$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.1$	0.1	0.2	0.3	0.4	0.5
0.1	Simulated $Mean\,(\widehat{\omega}_p)$	0.0973	0.0988	0.0930	0.0926	0.0981
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.2773	0.6990	0.8563	0.1831	0.4735
	Simulated $\sqrt{Var\,(\widehat{\omega}_p)}$	0.0590	0.0642	0.0770	0.0812	0.0994
	Simulated $Mean\,(\widehat{\omega}_1)$	0.0989	0.1003	0.0951	0.0946	0.1004
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.8101	0.2138	0.3047	0.8762	0.2350
	Simulated $\sqrt{Var\,(\widehat{\omega}_1)}$	0.0587	0.0641	0.0765	0.0806	0.0993
	Theoretical $\sqrt{Var\,(\widehat{\omega}_1)}$	0.0573	0.0639	0.0725	0.0839	0.1001
0.3	Simulated $Mean\,(\widehat{\omega}_p)$	0.2992	0.3030	0.3000	0.3032	0.2976
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.5105	0.3564	0.7671	0.6585	0.8058
	Simulated $\sqrt{Var\,(\widehat{\omega}_p)}$	0.0618	0.0689	0.0804	0.0906	0.1046
	Simulated $Mean\,(\widehat{\omega}_1)$	0.3007	0.3048	0.3021	0.3054	0.3001
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.0313	0.0832	0.2003	0.9224	0.3071
	Simulated $\sqrt{Var\,(\widehat{\omega}_1)}$	0.0618	0.0689	0.0804	0.0904	0.1045
	Theoretical $\sqrt{Var\,(\widehat{\omega}_1)}$	0.0610	0.0683	0.0775	0.0897	0.1064
0.5	Simulated $Mean\,(\widehat{\omega}_p)$	0.4954	0.4957	0.5067	0.4991	0.4941
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.7299	0.3150	0.0994	0.8581	0.8645

(Continued on next page.)

$\pi = 0.1$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.1$	0.1	0.2	0.3	0.4	0.5
	Simulated $\sqrt{Var\left(\widehat{\omega}_p\right)}$	0.0627	0.0703	0.0770	0.0904	0.1081
	Simulated $Mean\left(\widehat{\omega}_1\right)$	0.4969	0.4974	0.5086	0.5012	0.4969
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.8923	0.1945	0.7300	0.2590	0.0555
	Simulated $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0622	0.0700	0.0772	0.0907	0.1082
	Theoretical $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0608	0.0687	0.0788	0.0918	0.1094
0.7	Simulated $Mean\left(\widehat{\omega}_p\right)$	0.7017	0.7000	0.6962	0.6920	0.6924
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.4762	0.0102	0.0461	0.1887	0.0659
	Simulated $\sqrt{Var\left(\widehat{\omega}_p\right)}$	0.0561	0.0664	0.0811	0.0939	0.1104
	Simulated $Mean\left(\widehat{\omega}_1\right)$	0.7026	0.7014	0.6979	0.6943	0.6951
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.5993	0.1886	0.3647	0.9175	0.6092
	Simulated $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0560	0.0661	0.0807	0.0932	0.1094
	Theoretical $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0586	0.0672	0.0778	0.0915	0.1103
0.9	Simulated $Mean\left(\widehat{\omega}_p\right)$	0.8960	0.8983	0.8994	0.9026	0.9040
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.2880	0.5640	0.0244	0.0054	0.8989
	Simulated $\sqrt{Var\left(\widehat{\omega}_p\right)}$	0.0560	0.0656	0.0792	0.0947	0.1086
	Simulated $Mean\left(\widehat{\omega}_1\right)$	0.8967	0.8994	0.9008	0.9048	0.9063
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.7018	0.5112	0.6784	0.4416	0.6700
	Simulated $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0558	0.0652	0.0787	0.0939	0.1087

(Continued on next page.)

$\pi = 0.1$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.1$	0.1	0.2	0.3	0.4	0.5
	Theoretical $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0551	0.0642	0.0754	0.0899	0.1094

Table 6: Simulation Results for $\widehat{\omega}_p$ ($\pi = 0.2$)

$\pi = 0.2$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.2$	0.1	0.2	0.3	0.4	0.5
0.1	Simulated $Mean\,(\widehat{\omega}_p)$	0.0967	0.0992	0.1025	0.0927	0.0975
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.2267	0.0734	0.0315	0.1178	0.0302
	Simulated $\sqrt{Var\,(\widehat{\omega}_p)}$	0.0949	0.1016	0.1272	0.1488	0.1720
	Simulated $Mean\,(\widehat{\omega}_1)$	0.1007	0.1034	0.1083	0.0995	0.1050
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.6371	0.6830	0.4509	0.0916	0.1503
	Simulated $\sqrt{Var\,(\widehat{\omega}_1)}$	0.0942	0.1011	0.1266	0.1475	0.1715
	Theoretical $\sqrt{Var\,(\widehat{\omega}_1)}$	0.0936	0.1054	0.1206	0.1408	0.1687
0.3	Simulated $Mean\,(\widehat{\omega}_p)$	0.2976	0.2936	0.2987	0.2988	0.2892
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.1666	0.0082	0.0011	0.0378	0.1905
	Simulated $\sqrt{Var\,(\widehat{\omega}_p)}$	0.0901	0.1014	0.1158	0.1441	0.1723
	Simulated $Mean\,(\widehat{\omega}_1)$	0.3010	0.2977	0.3032	0.3049	0.2965
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.7142	0.8480	0.4990	0.9836	0.3822
	Simulated $\sqrt{Var\,(\widehat{\omega}_1)}$	0.0892	0.0997	0.1148	0.1427	0.1707
	Theoretical $\sqrt{Var\,(\widehat{\omega}_1)}$	0.0913	0.1033	0.1190	0.1393	0.1675
0.5	Simulated $Mean\,(\widehat{\omega}_p)$	0.5010	0.4978	0.4949	0.4991	0.4957
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.7304	0.0127	0.0001	0.0000	0.0102

(Continued on next page.)

$\pi = 0.2$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.2$	0.1	0.2	0.3	0.4	0.5
	Simulated $\sqrt{Var\left(\widehat{\omega}_p\right)}$	0.0894	0.1041	0.1132	0.1423	0.1730
	Simulated $Mean\left(\widehat{\omega}_1\right)$	0.5040	0.5015	0.4993	0.5048	0.5027
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.1959	0.9715	0.3543	0.3487	0.7544
	Simulated $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0892	0.1028	0.1116	0.1404	0.1703
	Theoretical $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0874	0.0998	0.1154	0.1361	0.1652
0.7	Simulated $Mean\left(\widehat{\omega}_p\right)$	0.6956	0.6951	0.6969	0.6872	0.6848
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.0103	0.0083	0.0608	0.0023	0.0310
	Simulated $\sqrt{Var\left(\widehat{\omega}_p\right)}$	0.0833	0.0961	0.1066	0.1369	0.1694
	Simulated $Mean\left(\widehat{\omega}_1\right)$	0.6980	0.6979	0.7000	0.6922	0.6920
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.3789	0.2298	0.6144	0.3732	0.6654
	Simulated $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0823	0.0955	0.1060	0.1347	0.1674
	Theoretical $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0827	0.0952	0.1110	0.1321	0.1615
0.9	Simulated $Mean\left(\widehat{\omega}_p\right)$	0.8972	0.8951	0.8976	0.8989	0.8920
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.2114	0.7450	0.0235	0.0017	0.0074
	Simulated $\sqrt{Var\left(\widehat{\omega}_p\right)}$	0.0808	0.0899	0.1077	0.1269	0.1560
	Simulated $Mean\left(\widehat{\omega}_1\right)$	0.8984	0.8972	0.9003	0.9024	0.8972
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.5852	0.4135	0.8261	0.2664	0.5849
	Simulated $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0801	0.0890	0.1071	0.1252	0.1547

(Continued on next page.)

$\pi = 0.2$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.2$	0.1	0.2	0.3	0.4	0.5
	Theoretical $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0782	0.0906	0.1064	0.1278	0.1573

Table 7: Simulation Results for $\widehat{\omega}_p$ ($\pi = 0.3$)

$\pi = 0.3$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.3$	0.1	0.2	0.3	0.4	0.5
0.1	Simulated $Mean\,(\widehat{\omega}_p)$	0.0856	0.0878	0.0873	0.0847	0.0701
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.0000	0.0016	0.0000	0.0000	0.0001
	Simulated $\sqrt{Var\,(\widehat{\omega}_p)}$	0.1701	0.1708	0.2079	0.2322	0.2842
	Simulated $Mean\,(\widehat{\omega}_1)$	0.0984	0.1000	0.1032	0.1012	0.0904
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.7258	0.2640	0.7515	0.9179	0.5745
	Simulated $\sqrt{Var\,(\widehat{\omega}_1)}$	0.1627	0.1668	0.2012	0.2257	0.2770
	Theoretical $\sqrt{Var\,(\widehat{\omega}_1)}$	0.1562	0.1767	0.2024	0.2368	0.2849
0.3	Simulated $Mean\,(\widehat{\omega}_p)$	0.2933	0.2876	0.3011	0.2793	0.3066
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.0000	0.0000	0.0015	0.0049	0.0000
	Simulated $\sqrt{Var\,(\widehat{\omega}_p)}$	0.1522	0.1666	0.1838	0.2268	0.2835
	Simulated $Mean\,(\widehat{\omega}_1)$	0.3034	0.2986	0.3135	0.2954	0.3256
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.6215	0.1317	0.4477	0.0734	0.8956
	Simulated $\sqrt{Var\,(\widehat{\omega}_1)}$	0.1477	0.1627	0.1827	0.2227	0.2770
	Theoretical $\sqrt{Var\,(\widehat{\omega}_1)}$	0.1466	0.1666	0.1928	0.2270	0.2756
0.5	Simulated $Mean\,(\widehat{\omega}_p)$	0.5003	0.4904	0.4944	0.4841	0.5000
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.0002	0.0000	0.0025	0.0000	0.0038

(Continued on next page.)

$\pi = 0.3$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.3$	0.1	0.2	0.3	0.4	0.5
	Simulated $\sqrt{Var\left(\widehat{\omega}_p\right)}$	0.1400	0.1688	0.1899	0.2204	0.2771
	Simulated $Mean\left(\widehat{\omega}_1\right)$	0.5089	0.4997	0.5055	0.4983	0.5181
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.6729	0.7136	0.6895	0.2229	0.0370
	Simulated $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.1360	0.1623	0.1859	0.2137	0.2753
	Theoretical $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.1368	0.1567	0.1829	0.2171	0.2657
0.7	Simulated $Mean\left(\widehat{\omega}_p\right)$	0.6995	0.6880	0.7032	0.6859	0.6771
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.0008	0.1381	0.0000	0.0000	0.0000
	Simulated $\sqrt{Var\left(\widehat{\omega}_p\right)}$	0.1231	0.1538	0.1801	0.2162	0.2593
	Simulated $Mean\left(\widehat{\omega}_1\right)$	0.7037	0.6956	0.7135	0.6984	0.6921
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.6318	0.0942	0.4382	0.5706	0.9906
	Simulated $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.1202	0.1501	0.1753	0.2098	0.2516
	Theoretical $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.1285	0.1478	0.1733	0.2074	0.2560
0.9	Simulated $Mean\left(\widehat{\omega}_p\right)$	0.9007	0.9001	0.8853	0.8738	0.8745
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.0000	0.0100	0.0346	0.0001	0.0000
	Simulated $\sqrt{Var\left(\widehat{\omega}_p\right)}$	0.1193	0.1472	0.1688	0.2035	0.2657
	Simulated $Mean\left(\widehat{\omega}_1\right)$	0.9030	0.9047	0.8932	0.8835	0.8902
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.0328	0.4356	0.6157	0.1560	0.4850
	Simulated $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.1164	0.1436	0.1644	0.1964	0.2517

(Continued on next page.)

$\pi = 0.3$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.3$	0.1	0.2	0.3	0.4	0.5
	Theoretical $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.1218	0.1405	0.1652	0.1985	0.2462

Table 8: Simulation Results for $\widehat{\omega}_p$ ($\pi = 0.8$)

$\pi = 0.8$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.8$	0.1	0.2	0.3	0.4	0.5
0.1	Simulated $Mean\,(\widehat{\omega}_p)$	0.0955	0.0941	0.0900	0.0957	0.0919
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.4862	0.2306	0.0264	0.0295	0.1136
	Simulated $\sqrt{Var\,(\widehat{\omega}_p)}$	0.0953	0.1093	0.1237	0.1438	0.1674
	Simulated $Mean\,(\widehat{\omega}_1)$	0.0997	0.0988	0.0953	0.1017	0.0986
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.4785	0.9118	0.4188	0.5492	0.9310
	Simulated $\sqrt{Var\,(\widehat{\omega}_1)}$	0.0946	0.1083	0.1216	0.1422	0.1660
	Theoretical $\sqrt{Var\,(\widehat{\omega}_1)}$	0.0936	0.1054	0.1206	0.1408	0.1687
0.3	Simulated $Mean\,(\widehat{\omega}_p)$	0.2943	0.2975	0.2959	0.2912	0.2938
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.0218	0.5220	0.0098	0.0001	0.0204
	Simulated $\sqrt{Var\,(\widehat{\omega}_p)}$	0.0914	0.1000	0.1197	0.1431	0.1690
	Simulated $Mean\,(\widehat{\omega}_1)$	0.2980	0.3012	0.3004	0.2972	0.3014
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.4003	0.1983	0.8416	0.4635	0.6186
	Simulated $\sqrt{Var\,(\widehat{\omega}_1)}$	0.0900	0.0996	0.1180	0.1404	0.1671
	Theoretical $\sqrt{Var\,(\widehat{\omega}_1)}$	0.0913	0.1033	0.1190	0.1393	0.1675
0.5	Simulated $Mean\,(\widehat{\omega}_p)$	0.4963	0.4949	0.4944	0.4960	0.4952
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.0002	0.0000	0.0267	0.1210	0.6562

(Continued on next page.)

$\pi = 0.8$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

ω	$\pi = 0.8$	T = 0.1	0.2	0.3	0.4	0.5
	Simulated $\sqrt{Var\left(\widehat{\omega}_p\right)}$	0.0918	0.1009	0.1181	0.1358	0.1654
	Simulated $Mean\left(\widehat{\omega}_1\right)$	0.4997	0.4986	0.4986	0.5012	0.5014
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.3245	0.0683	0.6687	0.4085	0.2826
	Simulated $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0905	0.0991	0.1169	0.1344	0.1651
	Theoretical $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0874	0.0998	0.1154	0.1361	0.1652
0.7	Simulated $Mean\left(\widehat{\omega}_p\right)$	0.7003	0.6990	0.6964	0.6951	0.6955
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.0097	0.0001	0.0000	0.5440	0.2388
	Simulated $\sqrt{Var\left(\widehat{\omega}_p\right)}$	0.0816	0.0986	0.1150	0.1328	0.1652
	Simulated $Mean\left(\widehat{\omega}_1\right)$	0.7024	0.7021	0.6999	0.6992	0.7013
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.2366	0.0638	0.0234	0.6874	0.7710
	Simulated $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0811	0.0977	0.1132	0.1321	0.1640
	Theoretical $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0827	0.0952	0.1110	0.1321	0.1615
0.9	Simulated $Mean\left(\widehat{\omega}_p\right)$	0.9019	0.8962	0.8977	0.8934	0.8969
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_p$	0.6381	0.0293	0.0062	0.0000	0.6961
	Simulated $\sqrt{Var\left(\widehat{\omega}_p\right)}$	0.0778	0.0914	0.1071	0.1344	0.1597
	Simulated $Mean\left(\widehat{\omega}_1\right)$	0.9029	0.8983	0.9005	0.8983	0.9019
	p-value of Shapiro-Wilk Normality Test on $\widehat{\omega}_1$	0.9229	0.4182	0.4801	0.0537	0.3503
	Simulated $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0774	0.0903	0.1057	0.1322	0.1594

(Continued on next page.)

$\pi = 0.8$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

		T				
ω	$\pi = 0.8$	0.1	0.2	0.3	0.4	0.5
	Theoretical $\sqrt{Var\left(\widehat{\omega}_1\right)}$	0.0782	0.0906	0.1064	0.1278	0.1573

5.3 Calculation of $\frac{n_1}{n_2}$

In this section, the setup for calculating the optimal sample ratios is the same as in Secions 5.1 and 5.2. The value of T varied from 0.1 to 0.5 in steps of 0.1, while ω varied from 0.1 to 0.9 in increments of 0.2. π values are 0.1, 0.2, 0.3 and 0.8. All the optimal values of $\frac{n_1}{n_2}$ are shown in Table 9.

Notice that there are two optimal values of $\frac{n_1}{n_2}$ for each set of T, π, and ω values. One is to minimize $Var(\widehat{\pi}_p)$ and another is to minimize $Var\left(\widehat{\omega}_1\right)$. For $\widehat{\pi}_p$, the ratio is usually in the range between 4.1 and 5.6. For $\widehat{\omega}_1$, the ratio ranges from 0.04 to 0.92. As discussed in Section 4.1 in more detail, these two optimal values behave in an opposite way to each other.

As estimating $\widehat{\pi}_p$ is the most important task in this analysis, the optimal ratios to minimize $Var(\widehat{\pi}_p)$ were chosen first. Then, n_1 and n_2 were calculated and used accordingly in each simulation in Secions 5.1 and 5.2.

Table 9: Calculation of Optimal Values of $\frac{n_1}{n_2}$

$\pi = (.1, .2, .3 \; \& \; .8)$, $n = 1000$, $p_1 = 0.85$, $p_2 = 0.15$, and $trials = 1000$

			T				
π	ω		0.1	0.2	0.3	0.4	0.5
0.1	0.1	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	4.837	4.899	4.967	5.040	5.121
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.743	0.765	0.789	0.813	0.839
	0.3	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	4.261	4.310	4.374	4.453	4.553
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.443	0.485	0.530	0.577	0.627
	0.5	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	4.165	4.163	4.181	4.223	4.293
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.200	0.267	0.334	0.401	0.471
	0.7	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	4.319	4.232	4.180	4.162	4.181
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.067	0.043	0.144	0.240	0.334
	0.9	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	4.730	4.486	4.319	4.214	4.165
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.416	0.226	0.067	0.073	0.200
0.2	0.1	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	5.342	5.372	5.402	5.435	5.469
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.821	0.839	0.858	0.877	0.897
	0.3	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	4.998	5.035	5.080	5.131	5.192

(Continued on next page.)

π	ω		T 0.1	0.2	0.3	0.4	0.5
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.519	0.567	0.615	0.664	0.715
	0.5	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	4.893	4.904	4.928	4.967	5.022
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.235	0.315	0.394	0.472	0.551
	0.7	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	4.955	4.912	4.893	4.898	4.928
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.077	0.050	0.169	0.283	0.394
	0.9	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	5.183	5.046	4.955	4.904	4.893
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.455	0.254	0.077	0.085	0.235
0.3	0.1	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	5.548	5.560	5.572	5.584	5.597
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.852	0.869	0.885	0.901	0.917
	0.3	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	5.398	5.416	5.436	5.460	5.486
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.561	0.610	0.658	0.707	0.755
	0.5	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	5.340	5.348	5.362	5.382	5.410
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.256	0.343	0.428	0.511	0.593
	0.7	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	5.362	5.345	5.339	5.344	5.362
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.083	0.054	0.184	0.309	0.428
	0.9	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	5.463	5.403	5.362	5.342	5.340

(Continued on next page.)

π	ω		T 0.1	0.2	0.3	0.4	0.5
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.480	0.272	0.083	0.092	0.256
0.8	0.1	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	5.342	5.372	5.402	5.435	5.469
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.821	0.839	0.858	0.877	0.897
	0.3	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	4.998	5.035	5.080	5.131	5.192
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.519	0.567	0.615	0.664	0.715
	0.5	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	4.893	4.904	4.928	4.967	5.022
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.235	0.315	0.394	0.472	0.551
	0.7	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	4.955	4.912	4.893	4.898	4.928
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.077	0.050	0.169	0.283	0.394
	0.9	$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\pi}_p)}$	5.183	5.046	4.955	4.904	4.893
		$\left(\frac{n_1}{n_2}\right)_{opt(\widehat{\omega}_1)}$	0.455	0.254	0.077	0.085	0.235

CHAPTER VI

CONCLUDING REMARKS

In this thesis, a new RRT model is proposed by combining the two previous models from Mangat and Singh 1990 [14], which introduced the Two-Stage (or Partial) RRT models, and Gupta 2001 [4], which introduced the concept of optional scrambling. Partial RRT models are expected to be more efficient because they introduce an element of truthful reporting in the survey in a random fashion. Optional models provide greater efficiency by incorporating truthful reporting from those respondents who do not consider the underlying question to be sensitive and are willing to provide a truthful response.

In Chapter III, we derive estimators for the prevalence of the sensitive characteristic (π) and the optionality parameter (ω). We show that $\widehat{\pi}_p$ is unbiased and has asymptotically normal distribution. We also discuss in detail the properties of $\widehat{\omega}_p$ and show that it too is unbiased and has asymptotically normal distribution if we use first order approximation of $\widehat{\omega}_p$.

The main focus in Chapter IV is on showing that introduction of truth element T in a binary optional RRT model may not always produce greater efficiency, as shown in Gupta et al 2012 [6] in the quatitative setting. In an optional RRT model, introduction of truth element T has to be weighed against the shrinking pool of respondents who provide a scrambled response. We discuss in detail how to select an optimal value of the truth parameter T so that the proposed Two-Stage Binary Optional RRT model performs better than the corresponding One-Stage model.

In Chapter V, we present results of an extensive simulation study and show that empirical mean and variance of $\widehat{\pi}_p$ are in good agreement with the corresponding theoretical values. Asymptotic normality of $\widehat{\pi}_p$ is also demonstrated. It is also shown that first order approximation of $\widehat{\omega}_p$ works very well. The theoretical approximate variance of $\widehat{\omega}_1$ was very close to the corresponding simulated variance while simulated means of $\widehat{\omega}_p$ and $\widehat{\omega}_1$ were also close to the true parameter value of ω. Asymptotic normality of $\widehat{\omega}_1$ was also clear in our simulations with $n = 1000$ in Section 5.2. However, normality of $\widehat{\omega}_p$ couldn't be observed unless the sample size is very large. In general, T tends to cause bigger bias in $\widehat{\omega}_p$ as $T \to 1$ and π tends to cause bigger bias as $\pi \to 0.5$. We used optimal n_1 and n_2 to have minimum variance of $\widehat{\pi}_p$ in each simulation. In every RRT model, the utmost importance should be given to the estimation of $\widehat{\pi}$ and its variance because they are directly related to the survey question.

In summary, the proposed Two-Stage Binary Optional RRT model will be more effective research tool than the Gupta 2001 [4] model because the variance of $\widehat{\pi}_p$ in the proposed model can be made smaller than the variance of the Gupta 2001 [4] model.

REFERENCES

[1] A. Chaudhuri. *Randomized Response and Indirect Questioning Techniques in Surveys*. Chapman & Hall/CRC, 2011.

[2] E. Coutts and B. Jann. Sensitive questions in online surveys: Experimental results for the RRT and the UCT. *Sociology Methods & Research*, 40(1–4):169–193, 2011.

[3] B. G. Greenberg, A. L. A. Abul-Ela, W. R. Simmons, and D. G. Horvitz. The unrelated question randomized response model-theoretical framework. *Journal of American Statistical Association*, 64(326):520–539, 1969.

[4] S. N. Gupta. Qualifying the sensitivity level of binary response personal interview survey questions. *Journal of Combinatorics, Information & System Sciences*, 26 (1–4):101–109, 2001.

[5] S. N. Gupta, B. C. Gupta, and S. Singh. Estimation of sensitivity level of personal interview survey questions. *Journal of Statistical Planning and Inference*, 100(2): 239–247, 2002.

[6] S. N. Gupta, S. Mehta, J. Shabbir, and B. K. Dass. Generalized scrambling in quantitative optional randomized response models. Forthcoming in *Communications in Statistics—Theory and Methods*, 2012.

[7] S. N. Gupta and J. Shabbir. Sensitivity estimation for personal interview survey questions. *Statistica*, 64(3):643–653, 2004.

[8] S. N. Gupta, J. Shabbir, and S. Sehra. A comparison of muliplicative and additive optional RRT models. *Journal of Statistical Planning and Inference*, 140(10): 2870–2874, 2010.

[9] S. N. Gupta, B. Thornton, J. Shabbir, and S. Singhal. A comparison of muliplicative and additive optional RRT models. *Journal of Statistical Theory and Applications*, 64:226–239, 2006.

[10] E. Jones and H. Sigall. The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, 76(5):349–364, 1971.

[11] S. S. K. Kwan, M. K. P. So, and K. Y. Tam. Applying the randomized response technique to elicit truthful responses to sensitive questions in is research: The case of software piracy behavior. *Information Systems Research*, 21(4):941–959, 2010.

[12] J. M. Lavender and D. A. Anderson. Effect of perceived anonymity in assessments of eating disorderd behaviors and attitudes. *International Journal of Eating Disorders*, 42(6):546–551, 2009.

[13] G. J. L. M. Lensvelt-Mulders, P. G. M. van der Heijden, O. Laudy, and G. van Gils. A validation of a computer-assisted randomized response survey to estimate the prevalence of fraud in social security. *Journal of the Royal Statistical Society. Series A*, 169(2):305–318, 2006.

[14] N. S. Mangat and R. Singh. An alternative randomized response procedure. *Biometrika*, 77(2):439–442, 1990.

[15] M. Ostrapczuk, J. Musch, and M. Moshagen. A randomized-response investigation of the education effect in attitudes towards foreigners. *European Journal of Social Psychology*, 39(6):920–931, 2009.

[16] D. L. Paulhus. Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3):598–609, 1984.

[17] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

[18] N. J. Roese and D. W. Jamieson. Twenty years of bogus pipeline research: A critical review and meta-analysis. *Psychological Bulletin*, 114(2):363–375, 1993.

[19] A. Schneider. An examination of whether incentive compensation and stock ownership affect internal auditor objectivity. *Journal of Managerial Issues*, 15(4): 486–497, 2003.

[20] M. Spiegel. *Theory and Problems of Advanced Calculus*. Schaum's Outline Series. McGraw Hill, 1971.

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

APPENDIX A

[R-CODE] PROPOSED MODEL

```
########################################################################
#                                                                      #
# TITLE:          TWO-STAGE BINARY OPTIONAL RRT MODEL                  #
# DATE:           2012-06-15                                           #
# AUTHOR:         J. SIHM                                              #
# DESCRIPTION:    APPENDIX TO SIHM'S MASTER'S                          #
#                 THESIS AT UNC-GREENSBORO, SUMMER 2012                #
# THESIS ADVISOR: DR. SAT N. GUPTA                                     #
#                 DEPARTMENT OF MATHEMATICS & STATISTICS               #
#                 UNIVERSITY OF NORTH CAROLINA AT GREENSBORO           #
#                                                                      #
########################################################################
#                                                                      #
# R VERSION:      2.15.0 (2012-03-30)                                  #
# PLATFORM:       x86-64-pc-mingw32/x64 (64-bit)                       #
# PACKAGES USED:  "nortest"                                            #
#                 BY JUERGEN GROSS, 2012-04-24                         #
#                                                                      #
# Please install & include "nortest" package before running this code. #
# > install.packages("nortest")                                        #
# > library(nortest)                                                   #
#                                                                      #
# REMARKS:        NOTICE THAT THE OUTPUT OF 2ND ORDER APPROXIMATION OF #
#                 W.hat WAS NOT PRESENTED IN TABLES 5 6 7 AND 8.       #
#                 ALSO NOTICE THAT A SEPERATE PYTHON PROGRAM TO CONVERT #
#                 THE OUTPUT OF THIS PROGRAM INTO LaTeX SCRIPTS FOR    #
#                 TABLES 1 ~ 9 EXISTS AND WILL BE INCLUDED IN AUTHOR'S #
#                 FUTURE PHD PROGRAMMING PROJECT AT UNCG               #
#                                                                      #
########################################################################
```

```
p1 <- 0.85; p2 <- 0.15;

n <- 1000; trials <- 1000
lambda <- (1-p1)/(1-p2)

# nt = 9 from 0.1 to 0.9
nt <- 9
# npi = 9 from 0.1 to 0.9
npi <- 9
# nw = 9 from 0.1 to 0.9
nw <- 9

# Create a Matrix with nrow = nt*npi*nw & ncol = 29
# (1) t, (2) pi, (3) mean(pi.hat), (4) var(pi.hat),
# (5) Theoretical Var[pi.hat], (6) n, (7) n1, (8) n2,
# (9) trials, (10) p1, (11) p2, (12) lambda
# (13) py1, (14) mean(py1.hat), (15) var(py1.hat),
# (16) py2, (17) mean(py2.hat), (18) var(py2.hat)
# (19) w, (20) mean(w.hat), (21) var(w.hat),
# (22) opt_ratio_pi.hat
# 6 Columns Added on Feb 9, 2012: thus, 22 + 6 = 28
# (23) Theoretical E[pi.hat], (24) theta1, (25) theta2
# (26) Theoretical E[w.hat],
# (27) Theoretical 1st order Taylor Var[w.hat],
# (28) opt-ratio_w.hat
# 1 Column Added on Mar 29, 2012: thus, 28 + 1 = 29
# (29) Theoretical  2nd order Taylor Var[w.hat]
# 12 Columns Added on Jun 1, 2012: thus, 29 + 12 = 41
# (30) mean(theta1.hat), (31) var(theta1.hat)
# (32) mean(theta2.hat), (33) var(theta2.hat)
# (34) mean(w.hat1), (35) var(w.hat1)
# (36) mean(w.hat2), (37) var(w.hat2)
# (38) p-value for Normality Test sf.test on pi.hat
# (39) p-value for Normality Test sf.test on w.hat
```

```
# (40) p-value for Normality Test sf.test on w.hat1
# (41) p-value for Normality Test sf.test on w.hat2
# 41 columns, we have.
ncolumns <- 41

mdat <- matrix(rep(0,nt*npi*nw*ncolumns), nrow=nt*npi*nw,
ncol=ncolumns)
i <- j <- k <- l <-0
ni <- nt; nj <- npi; nk <- nw; nl <- trials

for (i in 1:ni)
{
for (j in 1:nj)
{
for (k in 1:nk)
{
mdat[k+nk*(nj*(i-1)+(j-1)),1] <- (1/(nt+1))*i      # t
mdat[k+nk*(nj*(i-1)+(j-1)),2] <- (1/(npi+1))*j     # pi
mdat[k+nk*(nj*(i-1)+(j-1)),6] <- n
## mdat[k+nk*(nj*(i-1)+(j-1)),7] <- n1
## mdat[k+nk*(nj*(i-1)+(j-1)),8] <- n2
mdat[k+nk*(nj*(i-1)+(j-1)),9] <- trials
mdat[k+nk*(nj*(i-1)+(j-1)),10] <- p1
mdat[k+nk*(nj*(i-1)+(j-1)),11] <- p2
mdat[k+nk*(nj*(i-1)+(j-1)),12] <- lambda
mdat[k+nk*(nj*(i-1)+(j-1)),19] <- (1/(nw+1))*k     # w

# (13) py1 <- t*pi+(1-t)*((1-w)*pi+w*(pi*p1+(1-pi)*(1-p1)))


mdat[k+nk*(nj*(i-1)+(j-1)),13] <- mdat[k+nk*(nj*(i-1)+(j-1)),1]*
mdat[k+nk*(nj*(i-1)+(j-1)),2]+
(1-mdat[k+nk*(nj*(i-1)+(j-1)),1])*((1-mdat[k+nk*(nj*(i-1)+(j-1)),
19])*mdat[k+nk*(nj*(i-1)+(j-1)),2]+
mdat[k+nk*(nj*(i-1)+(j-1)),19]*(mdat[k+nk*(nj*(i-1)+(j-1)),2]*
```

```
mdat[k+nk*(nj*(i-1)+(j-1)),10]+
(1-mdat[k+nk*(nj*(i-1)+(j-1)),2])*(1-mdat[k+nk*(nj*(i-1)+
(j-1)),10])))

# (16) py2 <- t*pi+(1-t)*((1-w)*pi+w*(pi*p2+(1-pi)*(1-p2)))

mdat[k+nk*(nj*(i-1)+(j-1)),16] <- mdat[k+nk*(nj*(i-1)+(j-1)),
1]*mdat[k+nk*(nj*(i-1)+(j-1)),2]+
(1-mdat[k+nk*(nj*(i-1)+(j-1)),1])*((1-mdat[k+nk*(nj*(i-1)+(j-
1)),19])*mdat[k+nk*(nj*(i-1)+(j-1)),2]+
mdat[k+nk*(nj*(i-1)+(j-1)),19]*(mdat[k+nk*(nj*(i-1)+(j-1)),
2]*mdat[k+nk*(nj*(i-1)+(j-1)),11]+
(1-mdat[k+nk*(nj*(i-1)+(j-1)),2])*(1-mdat[k+nk*(nj*(i-1)+
(j-1)),11])))

# (24) theta1 <- (2py1 - 1)

mdat[k+nk*(nj*(i-1)+(j-1)),24] <- 2*mdat[k+nk*(nj*(i-1)+
(j-1)),13] - 1

# (25) theta2 <- (2py2 - 1)

mdat[k+nk*(nj*(i-1)+(j-1)),25] <- 2*mdat[k+nk*(nj*(i-1)+
(j-1)),16] - 1

# Theoretical Part #

# (22) opt_ratio_pi.hat <- (1/lambda)*sqrt(py1*(1-py1)/
# (py2*(1-py2)))

mdat[k+nk*(nj*(i-1)+(j-1)),22] <- (1/mdat[k+nk*(nj*(i-1)+
(j-1)),12])*
sqrt(mdat[k+nk*(nj*(i-1)+(j-1)),13]*(1-mdat[k+nk*(nj*(i-
1)+(j-1)),13])/
```

```
(mdat[k+nk*(nj*(i-1)+(j-1)),16]*(1-mdat[k+nk*(nj*(i-
1)+(j-1)),16])))

# (8) n2 <- n/( ratio + 1 )

mdat[k+nk*(nj*(i-1)+(j-1)),8] <-
round(mdat[k+nk*(nj*(i-1)+(j-1)),6]/
( mdat[k+nk*(nj*(i-1)+(j-1)),22] + 1 ),0)

# (7) n1 <- n - n2

mdat[k+nk*(nj*(i-1)+(j-1)),7] <-
mdat[k+nk*(nj*(i-1)+(j-1)),6] - mdat[k+nk*(nj*(i-1)+(j-1)),8]

# (5) Var(pi.hat)

mdat[k+nk*(nj*(i-1)+(j-1)),5] <- (mdat[k+nk*(nj*(i-1)+(j-1)),
12]-1)^(-2)*
(mdat[k+nk*(nj*(i-1)+(j-1)),12]^2*(mdat[k+nk*(nj*(i-1)+
(j-1)),16]*
(1-mdat[k+nk*(nj*(i-1)+(j-1)),16])/mdat[k+nk*(nj*(i-1)+
(j-1)),8]) +
mdat[k+nk*(nj*(i-1)+(j-1)),13]*(1-mdat[k+nk*(nj*(i-1)+
(j-1)),13])/mdat[k+nk*(nj*(i-1)+(j-1)),7])

# (23) Theoretical E(pi.hat) <- (lambda*Py2 - Py1)/(lambda-1)

mdat[k+nk*(nj*(i-1)+(j-1)),23] <- (mdat[k+nk*(nj*(i-1)+(j-1)),12]*
mdat[k+nk*(nj*(i-1)+(j-1)),16] -
mdat[k+nk*(nj*(i-1)+(j-1)),13])/(mdat[k+nk*(nj*(i-1)+(j-1)),12]-1)

# (26) Theoretical E(w.hat) <- (theta1-theta2)/(2*(1-T)*
# ((1-p2)*theta1-(1-p1)*theta2))
```

```
mdat[k+nk*(nj*(i-1)+(j-1)),26] <- (mdat[k+nk*(nj*(i-1)+(j-
1)),24]-mdat[k+nk*(nj*(i-1)+(j-1)),25])/
(2*(1-mdat[k+nk*(nj*(i-1)+(j-1)),1])*
((1-mdat[k+nk*(nj*(i-1)+(j-1)),11])*mdat[k+nk*(nj*(i-
1)+(j-1)),24]-
(1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*mdat[k+nk*(nj*(i-
1)+(j-1)),25]))

# (27) Theoretical 1st Var(w.hat) <- (p1-p2)^2*
# ( theta2^2*(Py1*(1-Py1)/n1) + theta1^2*(Py2*(1-Py2)/n2) )/
# ( (1-t)^2*( (1-p2)*theta1 - (1-p1)*theta2 )^4 )

mdat[k+nk*(nj*(i-1)+(j-1)),27] <- (mdat[k+nk*(nj*(i-1)+(j-
1)),10]-mdat[k+nk*(nj*(i-1)+(j-1)),11])^2*
( mdat[k+nk*(nj*(i-1)+(j-1)),25]^2*(mdat[k+nk*(nj*(i-1)+(j-
1)),13]*(1-mdat[k+nk*(nj*(i-1)+(j-1)),13])/
mdat[k+nk*(nj*(i-1)+(j-1)),7]) + mdat[k+nk*(nj*(i-1)+(j-
1)),24]^2*(mdat[k+nk*(nj*(i-1)+(j-1)),16]*
(1-mdat[k+nk*(nj*(i-1)+(j-1)),16])/mdat[k+nk*(nj*(i-1)+(j-
1)),8]) )/
( (1-mdat[k+nk*(nj*(i-1)+(j-1)),1])^2*( (1-mdat[k+nk*(nj*(i-
1)+(j-1)),11])*
mdat[k+nk*(nj*(i-1)+(j-1)),24] - (1-mdat[k+nk*(nj*(i-1)+(j-
1)),10])*mdat[k+nk*(nj*(i-1)+(j-1)),25] )^4 )

# (28) opt_ratio_w.hat <- abs((2*Py2-1)/(2*Py1-1))*
# sqrt((Py1*(1-Py1))/(Py2*(1-Py2)))

mdat[k+nk*(nj*(i-1)+(j-1)),28] <- abs((2*mdat[k+nk*
(nj*(i-1)+(j-1)),16]-1)/
(2*mdat[k+nk*(nj*(i-1)+(j-1)),13]-1))*sqrt((mdat[k+nk*
(nj*(i-1)+(j-1)),13]*
(1-mdat[k+nk*(nj*(i-1)+(j-1)),13]))/(mdat[k+nk*(nj*
(i-1)+(j-1)),16]*(1-mdat[k+nk*(nj*(i-1)+(j-1)),16])))
```

```
# (29) Theoretical 2nd Var(w.hat) <- (p1-p2)^2*
# ( theta2^2*(Py1*(1-Py1)/n1) +
# theta1^2*(Py2*(1-Py2)/n2) +
#  ((1-p2)^2*theta2^2*(8*Py1^2*(1-Py1)^2)/n1^2)/( (1-
# p2)*theta1 - (1-p1)*theta2 )^2 +
#  ((1-p1)^2*theta1^2*(8*Py2^2*(1-Py2)^2)/n2^2)/( (1-
# p2)*theta1 - (1-p1)*theta2 )^2 +
#  ((1-p2)*theta1+(1-p1)*theta2)^2 *(4*Py1*Py2*(1-
# Py1)*(1-Py2)/(n1*n2))/( (1-p2)*theta1 - (1-p1)*theta2 )^2  )/
# ( (1-t)^2*( (1-p2)*theta1 - (1-p1)*theta2 )^4 )

mdat[k+nk*(nj*(i-1)+(j-1)),29] <- (mdat[k+nk*(nj*(i-1)+
(j-1)),10]-mdat[k+nk*(nj*(i-1)+(j-1)),11])^2*
(  mdat[k+nk*(nj*(i-1)+(j-1)),25]^2*(mdat[k+nk*(nj*(i-1)+
(j-1)),13]*(1-mdat[k+nk*(nj*(i-1)+(j-1)),13])/mdat[k+nk*(nj*(i-1)+(j-1)),7]) +
mdat[k+nk*(nj*(i-1)+(j-1)),24]^2*(mdat[k+nk*(nj*(i-1)+
(j-1)),16]*(1-mdat[k+nk*(nj*(i-1)+(j-1)),16])/mdat[k+nk*(nj*(i-1)+(j-1)),8]) +
( 8*(1-mdat[k+nk*(nj*(i-1)+(j-1)),11])^2*mdat[k+nk*(nj*(i-
1)+(j-1)),25]^2*mdat[k+nk*(nj*(i-1)+(j-1)),13]^2*(1-
mdat[k+nk*(nj*(i-1)+(j-1)),13])^2/mdat[k+nk*(nj*(i-1)+(j-1)),7]^2 +
8*(1-mdat[k+nk*(nj*(i-1)+(j-1)),10])^2*mdat[k+nk*(nj*
(i-1)+(j-1)),24]^2*mdat[k+nk*(nj*(i-1)+(j-1)),16]^2*(1-
mdat[k+nk*(nj*(i-1)+(j-1)),16])^2/mdat[k+nk*(nj*(i-1)+(j-1)),8]^2 +
4*(  (1-mdat[k+nk*(nj*(i-1)+(j-1)),11])*mdat[k+nk*(nj*
(i-1)+(j-1)),24] + (1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*
mdat[k+nk*(nj*(i-1)+(j-1)),25] )^2*mdat[k+nk*(nj*(i-1)+(j-1)),13]*
mdat[k+nk*(nj*(i-1)+(j-1)),16]*(1-mdat[k+nk*(nj*(i-1)+(j-1)),13])*
(1-mdat[k+nk*(nj*(i-1)+(j-1)),16])/(mdat[k+nk*(nj*(i-1)+(j-1)),7]*
mdat[k+nk*(nj*(i-1)+(j-1)),8]) )/
(  (1-mdat[k+nk*(nj*(i-1)+(j-1)),11])*mdat[k+nk*(nj*(i-1)+(j-1)),24] -
 (1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*mdat[k+nk*(nj*(i-1)+(j-1)),25]   )^2 )/
( (1-mdat[k+nk*(nj*(i-1)+(j-1)),1])^2*( (1-mdat[k+nk*(nj*(i-1)+(j-1)),11])*
mdat[k+nk*(nj*(i-1)+(j-1)),24] - (1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*
mdat[k+nk*(nj*(i-1)+(j-1)),25] )^4 )
```

```
}
}
}

pi.hat <-  w.hat <- w.hat1 <- w.hat2 <-
py1.hat <- py2.hat <- theta1.hat <- theta2.hat <-
a <- b1 <- b2 <- c <- d1 <- d2 <- e <- numeric(trials)

set.seed(76)

## for (i in 1:ni)
for (i in c(1,2,3,4,5))
{
## for (j in 1:nj)
for (j in c(1,2,3,8))
{
## for (k in 1:nk)
for (k in c(1,3,5,7,9))
{



for (l in 1:nl)
{
# Group 1

a <- rbinom(mdat[k+nk*(nj*(i-1)+(j-1)),7],1,mdat[k+nk*(nj*(i-1)+(j-1)),2])   # pi
b1 <- rbinom(mdat[k+nk*(nj*(i-1)+(j-1)),7],1,mdat[k+nk*(nj*(i-1)+(j-1)),10])  # p1
c <- rbinom(mdat[k+nk*(nj*(i-1)+(j-1)),7],1,mdat[k+nk*(nj*(i-1)+(j-1)),19])  # w
d1 <- rbinom(mdat[k+nk*(nj*(i-1)+(j-1)),7],1,mdat[k+nk*(nj*(i-1)+(j-1)),1])   # t

e <- d1*a+(1-d1)*((1-c)*a + c*(a*b1 + (1-a)*(1-b1)))

# Py1.hat <- Sum(e)/n1
```

```
py1.hat[l] <- sum(e)/mdat[k+nk*(nj*(i-1)+(j-1)),7]

# theta1.hat <- 2*Py1.hat - 1

theta1.hat[l] <- 2*py1.hat[l] - 1

a <- c <- e <- c(rep(NA,trials))

# Group 2

a <- rbinom(mdat[k+nk*(nj*(i-1)+(j-1)),8],1,mdat[k+nk*(nj*(i-1)+(j-1)),2])   # pi
b2 <- rbinom(mdat[k+nk*(nj*(i-1)+(j-1)),8],1,mdat[k+nk*(nj*(i-1)+(j-1)),11])  # p2
c <- rbinom(mdat[k+nk*(nj*(i-1)+(j-1)),8],1,mdat[k+nk*(nj*(i-1)+(j-1)),19])  # w
d2 <- rbinom(mdat[k+nk*(nj*(i-1)+(j-1)),8],1,mdat[k+nk*(nj*(i-1)+(j-1)),1])   # t

e <- d2*a+(1-d2)*((1-c)*a + c*(a*b2 + (1-a)*(1-b2)))

# Py2.hat <- Sum(e)/n2

py2.hat[l] <- sum(e)/mdat[k+nk*(nj*(i-1)+(j-1)),8]

# theta2.hat <- 2*Py2.hat - 1

theta2.hat[l] <- 2*py2.hat[l] - 1

a <- c <- e <- b1 <- b2 <- d1 <- d2 <- c(rep(NA,trials))

# Calculation of Estimators
# pi.hat & w.hat #

pi.hat[l] <- (mdat[k+nk*(nj*(i-1)+(j-1)),12]*py2.hat[l] -
py1.hat[l])/(mdat[k+nk*(nj*(i-1)+(j-1)),12]-1)

w.hat[l] <- (1/(2*pi.hat[l]-1))*(py1.hat[l]-py2.hat[l])/
```

```
((1-mdat[k+nk*(nj*(i-1)+(j-1)),1])*
(mdat[k+nk*(nj*(i-1)+(j-1)),10]-mdat[k+nk*(nj*(i-1)+(j-1)),11]))

# w.hat1: First Order Approximation #
# w.hat1 <- (1/(2*(1-t)))*(
# (theta1-theta2)/( (1-p2)*theta1 - (1-p1)*theta2 )
# + ((p1-p2)*theta2*(theta1.hat-theta1))/( (1-p2)*theta1 - (1-p1)*theta2 )^2
# - ((p1-p2)*theta1*(theta2.hat-theta2))/( (1-p2)*theta1 - (1-p1)*theta2 )^2
# )

w.hat1[l] <- (1/(2*(1-mdat[k+nk*(nj*(i-1)+(j-1)),1])))*
((mdat[k+nk*(nj*(i-1)+(j-1)),24]-mdat[k+nk*(nj*(i-1)+(j-1)),25])/
( (1-mdat[k+nk*(nj*(i-1)+(j-1)),11])*mdat[k+nk*(nj*(i-1)+(j-1)),24] -
(1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*mdat[k+nk*(nj*(i-1)+(j-1)),25] ) +
((mdat[k+nk*(nj*(i-1)+(j-1)),10]-mdat[k+nk*(nj*(i-1)+(j-1)),11])*
mdat[k+nk*(nj*(i-1)+(j-1)),25]*(theta1.hat[l]-mdat[k+nk*(nj*(i-1)+(j-1)),24]))/( (1-
mdat[k+nk*(nj*(i-1)+(j-1)),11])*mdat[k+nk*(nj*(i-1)+(j-1)),24] -
(1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*mdat[k+nk*(nj*(i-1)+(j-1)),25] )^2 -
((mdat[k+nk*(nj*(i-1)+(j-1)),10]-mdat[k+nk*(nj*(i-1)+(j-1)),11])*
mdat[k+nk*(nj*(i-1)+(j-1)),24]*(theta2.hat[l]-mdat[k+nk*(nj*(i-1)+(j-1)),25]))/( (1-
mdat[k+nk*(nj*(i-1)+(j-1)),11])*mdat[k+nk*(nj*(i-1)+(j-1)),24] -
(1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*mdat[k+nk*(nj*(i-1)+(j-1)),25] )^2 )

# w.hat2: Second Order Approximation #
# w.hat2 <- (1/(2*(1-t)))*(
# (theta1-theta2)/( (1-p2)*theta1 - (1-p1)*theta2 )
# + ((p1-p2)*theta2*(theta1.hat-theta1))/( (1-p2)*theta1 - (1-p1)*theta2 )^2
# - ((p1-p2)*theta1*(theta2.hat-theta2))/( (1-p2)*theta1 - (1-p1)*theta2 )^2
# - ((1-p2)*(p1-p2)*theta2*(theta1.hat-theta1)^2)/( (1-p2)*theta1 - (1-p1)*theta2 )^3
# - ((1-p1)*(p1-p2)*theta1*(theta2.hat-theta2)^2)/( (1-p2)*theta1 - (1-p1)*theta2 )^3
# + ((p1-p2)*((1-p2)*theta1+(1-p1)*theta2)*(theta1.hat-theta1)*(theta2.hat-theta2))
# /( (1-p2)*theta1 - (1-p1)*theta2 )^3
# )
```

```
w.hat2[l] <- (1/(2*(1-mdat[k+nk*(nj*(i-1)+(j-1)),1])))*
((mdat[k+nk*(nj*(i-1)+(j-1)),24]-mdat[k+nk*(nj*(i-1)+(j-1)),25])/
( (1-mdat[k+nk*(nj*(i-1)+(j-1)),11])*mdat[k+nk*(nj*(i-1)+(j-1)),24] -
(1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*mdat[k+nk*(nj*(i-1)+(j-1)),25] ) +
((mdat[k+nk*(nj*(i-1)+(j-1)),10]-mdat[k+nk*(nj*(i-1)+(j-1)),11])*
mdat[k+nk*(nj*(i-1)+(j-1)),25]*(theta1.hat[l]-mdat[k+nk*(nj*(i-1)+(j-1)),24]))/( (1-
mdat[k+nk*(nj*(i-1)+(j-1)),11])*mdat[k+nk*(nj*(i-1)+(j-1)),24] -
(1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*mdat[k+nk*(nj*(i-1)+(j-1)),25] )^2 -
((mdat[k+nk*(nj*(i-1)+(j-1)),10]-mdat[k+nk*(nj*(i-1)+(j-1)),11])*
mdat[k+nk*(nj*(i-1)+(j-1)),24]*(theta2.hat[l]-mdat[k+nk*(nj*(i-1)+(j-1)),25]))/( (1-
mdat[k+nk*(nj*(i-1)+(j-1)),11])*mdat[k+nk*(nj*(i-1)+(j-1)),24] -
(1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*mdat[k+nk*(nj*(i-1)+(j-1)),25] )^2 -
((1-mdat[k+nk*(nj*(i-1)+(j-1)),11])*(mdat[k+nk*(nj*(i-1)+(j-1)),10]-
mdat[k+nk*(nj*(i-1)+(j-1)),11])*mdat[k+nk*(nj*(i-1)+(j-1)),25]*(theta1.hat[l]-
mdat[k+nk*(nj*(i-1)+(j-1)),24])^2)/( (1-mdat[k+nk*(nj*(i-1)+(j-1)),11])*
mdat[k+nk*(nj*(i-1)+(j-1)),24] - (1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*
mdat[k+nk*(nj*(i-1)+(j-1)),25] )^3 -
((1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*(mdat[k+nk*(nj*(i-1)+(j-1)),10]-
mdat[k+nk*(nj*(i-1)+(j-1)),11])*mdat[k+nk*(nj*(i-1)+(j-1)),24]*(theta2.hat[l]-
mdat[k+nk*(nj*(i-1)+(j-1)),25])^2)/( (1-mdat[k+nk*(nj*(i-1)+(j-1)),11])*
mdat[k+nk*(nj*(i-1)+(j-1)),24] - (1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*
mdat[k+nk*(nj*(i-1)+(j-1)),25] )^3 +
((mdat[k+nk*(nj*(i-1)+(j-1)),10]-mdat[k+nk*(nj*(i-1)+(j-1)),11])*((1-
mdat[k+nk*(nj*(i-1)+(j-1)),11])*mdat[k+nk*(nj*(i-1)+(j-1)),24]+(1-
mdat[k+nk*(nj*(i-1)+(j-1)),10])*mdat[k+nk*(nj*(i-1)+(j-1)),25])*(theta1.hat[l]-
mdat[k+nk*(nj*(i-1)+(j-1)),24])*(theta2.hat[l]-mdat[k+nk*(nj*(i-1)+(j-1)),25]))/
( (1-mdat[k+nk*(nj*(i-1)+(j-1)),11])*mdat[k+nk*(nj*(i-1)+(j-1)),24] -
(1-mdat[k+nk*(nj*(i-1)+(j-1)),10])*mdat[k+nk*(nj*(i-1)+(j-1)),25] )^3 )

}

# Simulation Part #

mdat[k+nk*(nj*(i-1)+(j-1)),14] <- mean(py1.hat)
```

```
mdat[k+nk*(nj*(i-1)+(j-1)),15] <- var(py1.hat)

mdat[k+nk*(nj*(i-1)+(j-1)),17] <- mean(py2.hat)
mdat[k+nk*(nj*(i-1)+(j-1)),18] <- var(py2.hat)

mdat[k+nk*(nj*(i-1)+(j-1)),30] <- mean(theta1.hat)
mdat[k+nk*(nj*(i-1)+(j-1)),31] <- var(theta1.hat)

mdat[k+nk*(nj*(i-1)+(j-1)),32] <- mean(theta2.hat)
mdat[k+nk*(nj*(i-1)+(j-1)),33] <- var(theta2.hat)

mdat[k+nk*(nj*(i-1)+(j-1)),3] <- mean(pi.hat)
mdat[k+nk*(nj*(i-1)+(j-1)),4] <- var(pi.hat)

mdat[k+nk*(nj*(i-1)+(j-1)),20] <- mean(w.hat)
mdat[k+nk*(nj*(i-1)+(j-1)),21] <- var(w.hat)

mdat[k+nk*(nj*(i-1)+(j-1)),34] <- mean(w.hat1)
mdat[k+nk*(nj*(i-1)+(j-1)),35] <- var(w.hat1)

mdat[k+nk*(nj*(i-1)+(j-1)),36] <- mean(w.hat2)
mdat[k+nk*(nj*(i-1)+(j-1)),37] <- var(w.hat2)

# Normality Test
#    Anderson-Darling Test:     ad.test(x)     (# of (x) > 7)
#    Cramer-von Mises test:     cvm.test(x)    (# of (x) > 7)
#    Kolmogorov-Smirnov Test:   lillie.test(x)
#    Shapiro-Francia Test:      sf.test(x)     (Only for 5-5000)

mdat[k+nk*(nj*(i-1)+(j-1)),38] <- round(sf.test(pi.hat)$p.value,6)
mdat[k+nk*(nj*(i-1)+(j-1)),39] <- round(sf.test(w.hat)$p.value,6)
mdat[k+nk*(nj*(i-1)+(j-1)),40] <- round(sf.test(w.hat1)$p.value,6)
mdat[k+nk*(nj*(i-1)+(j-1)),41] <- round(sf.test(w.hat2)$p.value,6)
```

```
pi.hat <- py1.hat <- py2.hat <- w.hat <- c(rep(NA,trials))
a <- b1 <- b2 <- c <- d1 <- d2 <- e <- c(rep(NA,trials))

}
}
}

# ------------------------------------------------------------
# Export the matrix of "mdat" as csv format
write.csv(mdat, file="2012-JUN-15.SIHM.RRT.csv")
```