# Propensity score adjusted method for missing data

by

Minsun Kim Riddles

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:

Jae Kwang Kim, Co-major Professor

Sarah M. Nusser, Co-major Professor

Song X. Chen

Ranjan Maitra

Cindy L. Yu

Iowa State University

Ames, Iowa

2013

# DEDICATION

I dedicate this dissertation to God Almighty, who has kept me through the journey of completing this work. I also dedicate this work to my husband, John Riddles, my parents, Soonrim Sim and Namdoo Kim, my sister, Youngsun Kim, my parents-in-law, Deanna and Ron Riddles, and their family, Ronnie, Christine, and Angela Riddles, each of which have been very supportive of me throughout my dissertation.

# TABLE OF CONTENTS

# ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my major professors, Jae-kwang Kim and Sarah Nusser, who continually conveyed encouragement and support in regard to research and scholarship. Without their expertise, guidance, and persistent help, this dissertation would not have been possible. Also, I would like to thank my committee members, Song Chen, Ranjan Maitra, and Cindy Yu, for their feedback on my research.

I owe a huge thanks to Graham Kalton, David Mortganstein, and Jill Montaquila at Westat for giving me continuing advice and support. I would also like to thank Reka Howard, Eunice Kim, Jongho Im, Yeunsook Lee, Sixia Chen, and Stephanie Zimmer for being good friends and providing comfort throughout this dissertation.

# ABSTRACT

Propensity score adjustment is a popular technique for handling unit nonresponse in sample surveys. When the response probability does not depend on the study variable that is subject to missingness, conditional on the auxiliary variables that are observed throughout the sample, the response mechanism is often called missing at random (MAR) or ignorable, and the propensity score can be computed using the auxiliary variables. On the other hand, if the response probability depends on the study variable that is subject to missingness, the response mechanism is often called not missing at random (NMAR) or nonignorable, and estimating the response probability requires additional distributional assumptions about the study variable. In this dissertation, we investigate the propensity-score-adjustment method and the asymptotic properties of the estimators under two different assumptions, MAR and NMAR.

We discuss some asymptotic properties of propensity-score-adjusted(PSA) estimators and derive optimal estimators based on a regression model for the finite population under MAR. An optimal propensity-score-adjusted estimator can be implemented using an augmented propensity model. Variance estimation is discussed, and the results from two simulation studies are presented. We also consider the NMAR case with an explicit parametric model for response probability and propose a parameter estimation method for the response model that is based on the distributional assumptions of the observed part of the sample instead of making fully parametric assumptions about the population distribution. The proposed method has the advantage that the model for the observed part of the sample can be verified from the data, which leads to an estimator that is less sensitive to model assumptions. Under NMAR, asymptotic properties of PSA estimators

are presented, variance estimation is discussed, and results from two limited simulation studies are presented to compare the performance of the proposed method with the existing methods.

# CHAPTER 1. INTRODUCTION

Throughout the dissertation, we consider a population of three random variables $(\mathbf{X}, Y, \delta)$ and $n$ independent realization of the random variables, $(\mathbf{x}_i, y_i, \delta_i)$, $i = 1, 2, \cdots, n$. The auxiliary variable $\mathbf{x}_i$ is observed for $i = 1, 2, \cdots, n$, while $y_i$ is observed if and only if $\delta_i = 1$, and $\delta_i$ is the response indicator, which is dichotomous, taking values of 1 or 0.

The response mechanism is the distribution of $\delta$, which is crucial in estimation with missing data. The response mechanism is missing completely at random (MCAR) if the response indicator $\delta$ is independent of the study variable $Y$, which is subject to missingness, and the auxiliary variable $\mathbf{X}$ that is observed throughout the sample. A weaker condition for the response mechanism is missing at random (MAR). The response mechanism is MAR if the response indicator $\delta$ is independent of the study variable $Y$, where $Y$ has missingness, conditional on the auxiliary variable $\mathbf{X}$ that is observed throughout the sample [Rubin (1976)]. When the response mechanism is missing at random, the response mechanism is also called ignorable [Little and Rubin (2002)]. Lastly, the response mechanism is not missing at random (NMAR) or nonignorable if the response indicator $\delta$ depends on the study variable $Y$ even conditional on the auxiliary variable $\mathbf{X}$.

The parameter of interest is $\theta_0$, which is determined by solving $E\{U(\theta; \mathbf{X}, Y)\} = 0$. Under complete response, a consistent estimator of $\theta_0$ can be found by solving the estimating equation,

$$\sum_{i=1}^{n} U(\theta; \mathbf{x}_i, y_i) = 0. \tag{1.1}$$

In the presence of missingness, the estimating equation (1.1) cannot be computed. Instead, one can consider an estimator that is based on the complete cases as $\sum_{i=1}^{n} \delta_i U(\theta; \mathbf{x}_i, y_i) =$

0. While the estimation based on the complete cases results in unbiased estimation under MCAR, the method results in biased estimation if MCAR does not hold. Also, this method does not utilize the auxiliary variable $\mathbf{x}_i$ information for $\delta_i = 0$ cases, which can be used to improve efficiency.

If the true response probability $\pi_{i*}$ were known, bias could be adjusted by incorporating the response probability in the estimating equation $\sum_{i=1}^{n} \delta_i \pi_{i*}^{-1} U(\theta; \mathbf{x}_i, y_i) = 0$. However, the true response probabilities are generally unknown. Instead, we can consider the estimating equation

$$\sum_{i=1}^{n} \delta_i \pi_i^{-1} U(\theta; \mathbf{x}_i, y_i) = 0, \tag{1.2}$$

where $\pi_i = Pr(\delta_i = 1 | \mathbf{x}_i, y_i)$ is the conditional response probability. The conditional response probability $\pi_i = Pr(\delta_i = 1 | \mathbf{x}_i, y_i)$ is often called the propensity score (PS) [Rosenbaum and Rubin (1983)].

Since the conditional response probability $\pi_i = Pr(\delta_i = 1 | \mathbf{x}_i, y_i)$ is also unknown in general, the conditional response probability still needs to be estimated. One can consider a parametric model for the response probability as $\pi_i = \pi(\boldsymbol{\phi}_0; \mathbf{x}_i, y_i)$ for some $\boldsymbol{\phi}_0 \in \Omega$. Henceforth, this will be referred to as the PS model. If the parameter $\boldsymbol{\phi}_0$ can be estimated consistently by $\hat{\boldsymbol{\phi}}$, an estimator for $\theta_0$ can be found by solving

$$U_{PSA}(\theta; \hat{\boldsymbol{\phi}}) = \sum_{i=1}^{n} \delta_i \hat{\pi}_i^{-1} U(\theta; \mathbf{x}_i, y_i) = 0, \tag{1.3}$$

where $\hat{\pi}_i = \pi(\hat{\boldsymbol{\phi}}; \mathbf{x}_i, y_i)$. The estimator that is found by solving (1.3) for $\theta$ is called the propensity-score-adjusted (PSA) estimator.

Before considering the PS model under nonignorable nonresponse, we will first examine the PS adjustment under ignorable nonresponse. Much research has been conducted on the PSA estimator for reducing nonresponse bias under MAR [Fuller et al. (1994); Rizzo et al. (1996)]. Rosenbaum and Rubin (1983) proposed using the PSA approach to estimate the treatment effects in observational studies. Little (1988) reviewed the PSA methods for handling unit nonresponse in survey sampling, Duncan and Stasny (2001) used the PSA approach to control coverage bias in telephone surveys, Lee (2006) applied the PSA method to a volunteer panel web survey, and Durrant and Skinner (2006) used the PSA approach to address measurement error.

In addition to applications, it is important to understand the asymptotic properties of PSA estimators. Kim and Kim (2007) used a Taylor expansion to obtain the asymptotic mean and variance of PSA estimators and discussed variance estimation. Da Silva and Opsomer (2006) and Da Silva and Opsomer (2009) considered nonparametric methods to obtain PSA estimators and presented their asymptotic properties. Otherwise, despite the popularity of PSA estimators, asymptotic properties of PSA estimators have received little attention in literature.

While much of the existing work on PSA estimators for unit nonresponse assumes that the response mechanism is MAR and the auxiliary variables for the PS model are observed throughout the sample, PSA estimators under nonignorable nonresponse has been receiving more attention in recent research.

As previously mentioned, the response mechanism is crucial in PS estimation since it determines how the parameter in the model for the response probability is estimated. Since, under nonignorable nonresponse, the PS model $Pr(\delta_i = 1|\mathbf{x}_i, y_i) = \pi(\boldsymbol{\phi}_0; \mathbf{x}_i, y_i)$ involves $y_i$, which is partially missing, while under ignorable nonresponse, the PS model $Pr(\delta_i = 1|\mathbf{x}_i, y_i) = \pi(\boldsymbol{\phi}_0; \mathbf{x}_i)$ does not involve missing part, estimation under nonignorable nonresponse is more complicated than estimation under MAR or ignorable nonresponse. In addition, the methodologies that are developed for ignorable nonresponse cannot be simply applied or extended to estimation under nonignorable nonresponse.

In order to estimate the parameters in the response probability model consistently, assumptions on the distribution of the study variable are added. Fully parametric approaches, which make parametric assumptions on the population distribution of the study variable in additional to the response probability model, are considered in Greenlees et al. (1982), Baker and Laird (1988), and Ibrahim et al. (1999). Parameter estimation without parametric assumptions on the population distribution of the study variable has been developed in recent works such as Chang and Kott (2008), Kott and Chang (2010), and Wang et al. (2013).

Efficient or optimal estimation of the parameters in the response probability model and inference with the estimated PS model are also addressed. Kim and Kim (2007) showed that the maximum likelihood estimation for the response probability model does not necessarily lead to optimal PSA estimator. Recent works such as Tan (2006) and Cao et al. (2009) also addressed this issue under ignorable nonresponse, but extensions to nonignorable nonresponse model are not well developed.

Our goal is to examine the asymptotic properties of PSA estimators and develop efficient estimation with PS adjustment. Chapter 2 is devoted to PSA estimation under ignorable nonresponse, and Chapter 3 is for PSA estimation under nonignorable nonresponse. In Chapter 2, we discuss asymptotic properties of PSA estimators and derive optimal estimators based on a regression model for the finite population under ignorable nonresponse. An optimal PSA estimator is implemented using an augmented propensity model. Variance estimation is discussed, and the results from two simulation studies are presented. In Chapter 3, we propose a new approach that is based on the distributional assumptions of the observed part of the sample instead of making fully parametric assumptions on the overall population distribution and the response mechanism under nonignorable nonresponse. Asymptotic properties of the resulting PSA estimator are presented. Also, to improve the efficiency of the PSA estimator, we incorporate the auxiliary variable $\mathbf{X}$ information by using the generalized method of moment. Variance estimation for each estimator is discussed, and results from two limited simulation studies are presented. Concluding remarks are made in Chapter 4.

# CHAPTER 2. SOME THEORY FOR PROPENSITY SCORE ADJUSTMENT ESTIMATORS IN SURVEY SAMPLING

Published in Survey Methodology, Volume 38, 157-165

Jae Kwang Kim and Minsun Kim Riddles

## Abstract

The propensity-scoring-adjustment approach is commonly used to handle selection bias in survey sampling applications, including unit nonresponse and undercoverage. The propensity score is computed using auxiliary variables observed throughout the sample. We discuss some asymptotic properties of propensity-score-adjusted estimators and derive optimal estimators based on a regression model for the finite population. An optimal propensity-score-adjusted estimator can be implemented using an augmented propensity model. Variance estimation is discussed and the results from two simulation studies are presented.

*Key words:* Calibration, Missing data, Nonresponse, Weighting.

## 2.1 Introduction

Consider a finite population of size $N$, where $N$ is known. For each unit $i$, $y_i$ is the study variable and $\mathbf{x}_i$ is the $q$-dimensional vector of auxiliary variables. The parameter of interest is the finite population mean of the study variable, $\theta = N^{-1}\sum_{i=1}^{N} y_i$. The finite population $\mathcal{F}_N = \{(\mathbf{x}_1', y_1), (\mathbf{x}_2', y_2), \cdots, (\mathbf{x}_N', y_N)\}$ is assumed to be a random sample of size $N$ from a superpopulation distribution $F(\mathbf{x}, y)$. Suppose a sample of size $n$ is drawn from the finite population according to a probability sampling design. Let $w_i = \pi_i^{-1}$ be the design weight, where $\pi_i$ is the first-order inclusion probability of unit $i$ obtained from the probability sampling design. Under complete response, the finite population mean can be estimated by the Horvitz-Thompson (HT) estimator, $\hat{\theta}_{HT} = N^{-1}\sum_{i\in A} w_i y_i$, where $A$ is the set of indices appearing in the sample.

In the presence of missing data, the HT estimator $\hat{\theta}_{HT}$ cannot be computed. Let $r$ be the response indicator variable that takes the value one if $y$ is observed and takes the value zero otherwise. Conceptually, as discussed by Fay (1992), Shao and Steel (1999), and Kim and Rao (2009), the response indicator can be extended to the entire population as $\mathcal{R}_N = \{r_1, r_2, \cdots, r_N\}$, where $r_i$ is a realization of the random variable $r$. In this case, the complete-case (CC) estimator $\hat{\theta}_{CC} = \sum_{i\in A} w_i r_i y_i / \sum_{i\in A} w_i r_i$ converges in probability to $E(Y|r=1)$. Unless the response mechanism is missing completely at random in the sense that $E\left(Y|r=1\right) = E\left(Y\right)$, the CC estimator is biased. To correct for the bias of the CC estimator, if the response probability

$$p(\mathbf{x}, y) = Pr(r = 1|\mathbf{x}, y) \tag{2.1}$$

is known, then the weighted CC estimator $\hat{\theta}_{WCC} = N^{-1}\sum_{i\in A} w_i r_i y_i / p(\mathbf{x}_i, y_i)$ can be used to estimate $\theta$. Note that $\hat{\theta}_{WCC}$ is unbiased because

$$E\{\sum_{i\in A} w_i r_i y_i / p(\mathbf{x}_i, y_i)|\mathcal{F}_N\} = E\{\sum_{i=1}^{N} r_i y_i / p(\mathbf{x}_i, y_i)|\mathcal{F}_N\} = \sum_{i=1}^{N} y_i.$$

If the response probability (2.1) is unknown, one can postulate a parametric model for the response probability $p(\mathbf{x}, y; \boldsymbol{\phi})$ indexed by $\boldsymbol{\phi} \in \Omega$ such that $p(\mathbf{x}, y) = p(\mathbf{x}, y; \boldsymbol{\phi}_0)$

for some $\boldsymbol{\phi}_0 \in \Omega$. We assume that there exists a $\sqrt{n}$-consistent estimator $\hat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}_0$ such that

$$\sqrt{n}\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0\right) = O_p\left(1\right), \tag{2.2}$$

where $g_n = O_p(1)$ indicates $g_n$ is bounded in probability. Using $\hat{\boldsymbol{\phi}}$, we can obtain the estimated response probability by $\hat{p}_i = p(\mathbf{x}_i, y_i; \hat{\boldsymbol{\phi}})$, which is often called the propensity score Rosenbaum and Rubin (1983). The propensity-score-adjusted (PSA) estimator can be constructed as

$$\hat{\theta}_{PSA} = \frac{1}{N}\sum_{i \in A} w_i \frac{r_i}{\hat{p}_i} y_i. \tag{2.3}$$

The PSA estimator (2.3) is widely used. Many surveys use the PSA estimator to reduce nonresponse bias [Fuller et al. (1994); Rizzo et al. (1996)]. Rosenbaum and Rubin (1983) and Rosenbaum (1987) proposed using the PSA approach to estimate the treatment effects in observational studies. Little (1988) reviewed the PSA methods for handling unit nonresponse in survey sampling. Duncan and Stasny (2001) used the PSA approach to control coverage bias in telephone surveys. Folsom (1991) and Iannacchione et al. (1991) used a logistic regression model for the response probability estimation. Lee (2006) applied the PSA method to a volunteer panel web survey. Durrant and Skinner (2006) used the PSA approach to address measurement error.

Despite the popularity of PSA estimators, asymptotic properties of PSA estimators have not received much attention in survey sampling literature. Kim and Kim (2007) used a Taylor expansion to obtain the asymptotic mean and variance of PSA estimators and discussed variance estimation. Da Silva and Opsomer (2006) and Da Silva and Opsomer (2009) considered nonparametric methods to obtain PSA estimators.

In this paper, we discuss optimal PSA estimators in the class of PSA estimators of the form (2.3) that use a $\sqrt{n}$-consistent estimator $\hat{\boldsymbol{\phi}}$. Such estimators are asymptotically unbiased for $\theta$. Finding minimum variance PSA estimators among this particular class of PSA estimators is a topic of major interest in this paper.

Section 2.2 presents the main results. An optimal PSA estimator using an augmented propensity score model is proposed in Section 2.3. In Section 2.4, variance estimation of the proposed estimator is discussed. Results from two simulation studies can be found in Section 2.5 and concluding remarks are made in Section 2.6.

## 2.2 Main Results

In this section, we discuss some asymptotic properties of PSA estimators. We assume that the response mechanism does not depend on $y$. Thus, we assume that

$$Pr(r = 1|\mathbf{x}, y) = Pr(r = 1|\mathbf{x}) = p(\mathbf{x}; \boldsymbol{\phi}_0) \tag{2.4}$$

for some unknown vector $\boldsymbol{\phi}_0$. The first equality implies that the data are missing-at-random (MAR), as we always observe $\mathbf{x}$ in the sample. Note that the MAR condition is assumed in the population model. In the second equality, we further assume that the response mechanism is known up to an unknown parameter $\boldsymbol{\phi}_0$. The response mechanism is slightly different from that of Kim and Kim (2007), where the response mechanism is assumed to be under the classical two-phase sampling setup and depends on the realized sample:

$$Pr(r = 1|\mathbf{x}, y, I = 1) = Pr(r = 1|\mathbf{x}, I = 1) = p(\mathbf{x}; \boldsymbol{\phi}_A^0). \tag{2.5}$$

Here, $I$ is the sampling indicator function defined throughout the population. That is, $I_i = 1$ if $i \in A$ and $I_i = 0$ otherwise. Unless the sampling design is non-informative in the sense that the sample selection probabilities are correlated with the response indicator even after conditioning on auxiliary variables Pfeffermann and Sverchkov (1999), the two response mechanisms, (2.4) and (2.5), are different. In survey sampling, assumption (2.4) is more appropriate because an individual's decision on whether or not to respond to a survey is at his or her own discretion. Here, the response indicator variable $r_i$ is defined throughout the population, as discussed in Section 2.1.

We consider a class of $\sqrt{n}$-consistent estimators of $\boldsymbol{\phi}_0$ in (2.4). In particular, we consider a class of estimators which can be written as a solution to

$$\hat{\mathbf{U}}_h(\boldsymbol{\phi}) \equiv \sum_{i \in A} w_i \left\{ r_i - p_i(\boldsymbol{\phi}) \right\} \mathbf{h}_i(\boldsymbol{\phi}) = \mathbf{0}, \tag{2.6}$$

where $p_i(\boldsymbol{\phi}) = p(\mathbf{x}_i; \boldsymbol{\phi})$ for some function $\mathbf{h}_i(\boldsymbol{\phi}) = \mathbf{h}(\mathbf{x}_i; \boldsymbol{\phi})$, a smooth function of $\mathbf{x}_i$ and parameter $\boldsymbol{\phi}$. Thus, the solution to (2.6) can be written as $\hat{\boldsymbol{\phi}}_h$, which depends on the choice of $\mathbf{h}_i(\boldsymbol{\phi})$. Any solution $\hat{\boldsymbol{\phi}}_h$ to (2.6) is consistent for $\boldsymbol{\phi}_0$ in (2.4) because $E\{\hat{\mathbf{U}}_h(\boldsymbol{\phi}_0) | \mathcal{F}_N\} = E\left[\sum_{i=1}^{N} \{r_i - p_i(\boldsymbol{\phi}_0)\} \mathbf{h}_i(\boldsymbol{\phi}_0) | \mathcal{F}_N\right]$ is zero under the response mechanism in (2.4). If we drop the sampling weights $w_i$ in (2.6), the estimated parameter $\hat{\boldsymbol{\phi}}_h$ is consistent for $\boldsymbol{\phi}_A^0$ in (2.5) and the resulting PSA estimator is consistent only when the sampling design is non-informative. The PSA estimators obtained from (2.6) using the sampling weights are consistent regardless of whether the sampling design is non-informative or not. According to Chamberlain (1987), any $\sqrt{n}$-consistent estimator of $\boldsymbol{\phi}_0$ in (2.4) can be written as a solution to (2.6). Thus, the choice of $\mathbf{h}_i(\boldsymbol{\phi})$ in (2.6) determines the efficiency of the resulting PSA estimator.

Let $\hat{\theta}_{PSA,h}$ be the PSA estimator in (2.3) using $\hat{p}_i = p_i(\hat{\boldsymbol{\phi}}_h)$ with $\hat{\boldsymbol{\phi}}_h$ being the solution to (2.6). To discuss the asymptotic properties of $\hat{\theta}_{PSA,h}$, assume a sequence of finite populations and samples, as in Isaki and Fuller (1982), such that $\sum_{i \in A} w_i \mathbf{u}_i - \sum_{i=1}^{N} \mathbf{u}_i = O_p\left(n^{-1/2} N\right)$ for any population characteristics $\mathbf{u}_i$ with bounded fourth moments. We also assume that the sampling weights are uniformly bounded. That is, $K_1 < N^{-1} n w_i < K_2$ for all $i$ uniformly in $n$, where $K_1$ and $K_2$ are fixed constants. In addition, we assume the following regularity conditions:

[C1] The response mechanism satisfies (2.4), where $p(\mathbf{x}; \boldsymbol{\phi})$ is continuous in $\boldsymbol{\phi}$ with continuous first and second derivatives in an open set containing $\boldsymbol{\phi}_0$. The responses are independent in the sense that $Cov(r_i, r_j | \mathbf{x}) = 0$ for $i \neq j$. Also, $p(\mathbf{x}_i; \boldsymbol{\phi}) > c$ for all $i$ for some fixed constant $c > 0$.

[C2] The solution to (2.6) exists and is unique almost everywhere. The function $\mathbf{h}_i(\boldsymbol{\phi}) = \mathbf{h}(\mathbf{x}_i; \boldsymbol{\phi})$ in (2.6) has a bounded fourth moment. Furthermore, the partial derivative $\partial\{\hat{\mathbf{U}}_h(\boldsymbol{\phi})\}/\partial\boldsymbol{\phi}$ is nonsingular for all $n$.

[C3] The estimating function $\hat{\mathbf{U}}_h(\boldsymbol{\phi})$ in (2.6) converges in probability to $\mathbf{U}_h(\boldsymbol{\phi}) = \sum_{i=1}^{N} \{r_i - p_i(\boldsymbol{\phi})\} \mathbf{h}_i(\boldsymbol{\phi})$ uniformly in $\boldsymbol{\phi}$. Furthermore, the partial derivative $\partial\{\hat{\mathbf{U}}_h(\boldsymbol{\phi})\}/\partial\boldsymbol{\phi}$ converges in probability to $\partial\{\mathbf{U}_h(\boldsymbol{\phi})\}/\partial\boldsymbol{\phi}$ uniformly in $\boldsymbol{\phi}$. The solution $\boldsymbol{\phi}_N$ to $\mathbf{U}_h(\boldsymbol{\phi}) = \mathbf{0}$ satisfies $N^{1/2}(\boldsymbol{\phi}_N - \boldsymbol{\phi}_0) = O_p(1)$ under the response mechanism.

Condition [C1] states the regularity conditions for the response mechanism. Condition [C2] is the regularity condition for the solution $\hat{\boldsymbol{\phi}}_h$ to (2.6). In Condition [C3], some regularity conditions are imposed on the estimating function $\hat{\mathbf{U}}_h(\boldsymbol{\phi})$ itself. By [C2] and [C3], we can establish the $\sqrt{n}$-consistency (2.2) of $\hat{\boldsymbol{\phi}}_h$.

Now, the following theorem deals with some asymptotic properties of the PSA estimator $\hat{\theta}_{PSA,h}$.

**Theorem 2.1.** *If conditions [C1]-[C4] hold, then under the joint distribution of the sampling mechanism and the response mechanism, the PSA estimator* $\hat{\theta}_{PSA,h}$ *satisfies*

$$\sqrt{n}\left(\hat{\theta}_{PSA,h} - \tilde{\theta}_{PSA,h}\right) = o_p(1), \tag{2.7}$$

*where*

$$\tilde{\theta}_{PSA,h} = \frac{1}{N}\sum_{i\in A} w_i \left\{p_i \mathbf{h}_i'\boldsymbol{\gamma}_h^* + \frac{r_i}{p_i}\left(y_i - p_i\mathbf{h}_i'\boldsymbol{\gamma}_h^*\right)\right\}, \tag{2.8}$$

$\boldsymbol{\gamma}_h^* = (\sum_{i=1}^{N} r_i\mathbf{z}_i p_i \mathbf{h}_i')^{-1}(\sum_{i=1}^{N} r_i \mathbf{z}_i y_i)$, $p_i = p(\mathbf{x}_i;\boldsymbol{\phi}_0)$, $\mathbf{z}_i = \partial\{p^{-1}(\mathbf{x}_i;\boldsymbol{\phi}_0)\}/\partial\boldsymbol{\phi}$, *and* $\mathbf{h}_i = \mathbf{h}(\mathbf{x}_i;\boldsymbol{\phi}_0)$. *Moreover, if the finite population is a random sample from a superpopulation model, then*

$$V(\tilde{\theta}_{PSA,h}) \geq V_l \equiv V(\hat{\theta}_{HT}) + \frac{1}{N^2}E\left\{\sum_{i\in A} w_i^2\left(\frac{1}{p_i} - 1\right)V(Y|\mathbf{x}_i)\right\}. \tag{2.9}$$

*The equality in (2.9) holds when* $\hat{\boldsymbol{\phi}}_h$ *satisfies*

$$\sum_{i\in A} w_i \left\{\frac{r_i}{p(\mathbf{x}_i;\hat{\boldsymbol{\phi}}_h)} - 1\right\} E(Y|\mathbf{x}_i) = 0, \tag{2.10}$$

*where $E(Y|\mathbf{x}_i)$ is the conditional expectation under the superpopulation model.*

*Proof.* Given $p_i(\boldsymbol{\phi}) = p(\mathbf{x}_i; \boldsymbol{\phi})$ and $\mathbf{h}_i(\boldsymbol{\phi}) = \mathbf{h}(\mathbf{x}_i; \boldsymbol{\phi})$, define

$$\hat{\theta}(\boldsymbol{\phi}, \boldsymbol{\gamma}) = N^{-1} \sum_{i \in A} w_i \left[ p_i(\boldsymbol{\phi}) \mathbf{h}_i'(\boldsymbol{\phi}) \boldsymbol{\gamma} + \frac{r_i}{p_i(\boldsymbol{\phi})} \left\{ y_i - p_i(\boldsymbol{\phi}) \mathbf{h}_i'(\boldsymbol{\phi}) \boldsymbol{\gamma} \right\} \right].$$

Since $\hat{\boldsymbol{\phi}}_h$ satisfies (2.6), we have $\hat{\theta}_{PSA} = \hat{\theta}(\hat{\boldsymbol{\phi}}_h, \boldsymbol{\gamma})$ for any choice of $\boldsymbol{\gamma}$. We now want to find a particular choice of $\boldsymbol{\gamma}$, say $\boldsymbol{\gamma}^*$, such that

$$\hat{\theta}(\hat{\boldsymbol{\phi}}_h, \boldsymbol{\gamma}^*) = \hat{\theta}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}^*) + o_p\left(n^{-1/2}\right). \tag{2.11}$$

As $\hat{\boldsymbol{\phi}}_h$ converges in probability to $\boldsymbol{\phi}_0$, the asymptotic equivalence (2.11) holds if

$$E\left\{ \frac{\partial}{\partial \boldsymbol{\phi}} \hat{\theta}(\boldsymbol{\phi}, \boldsymbol{\gamma}^*) | \boldsymbol{\phi} = \boldsymbol{\phi}_0 \right\} = \mathbf{0}, \tag{2.12}$$

using the theory of Randles (1982). Condition (2.12) holds if $\boldsymbol{\gamma}^* = \boldsymbol{\gamma}_h^*$, where $\boldsymbol{\gamma}_h^*$ is defined in (2.8). Thus, (2.11) reduces to

$$\hat{\theta}_{PSA,h} = \frac{1}{N} \sum_{i \in A} w_i \left\{ p_i \mathbf{h}_i' \boldsymbol{\gamma}_h^* + \frac{r_i}{p_i} \left( y_i - p_i \mathbf{h}_i' \boldsymbol{\gamma}_h^* \right) \right\} + o_p(n^{-1/2}), \tag{2.13}$$

which proves (2.7). The variance of $\tilde{\theta}_{PSA,h}$ can be derived as

$$\begin{aligned} V(\tilde{\theta}_{PSA,h}) &= V(\hat{\theta}_{HT}) + \frac{1}{N^2} E\left\{ \sum_{i \in A} w_i^2 \left( \frac{1}{p_i} - 1 \right) (y_i - p_i \mathbf{h}_i' \boldsymbol{\gamma}_h^*)^2 \right\} \\ &= V(\hat{\theta}_{HT}) + \frac{1}{N^2} E\left[ \sum_{i \in A} w_i^2 \left( \frac{1}{p_i} - 1 \right) \left\{ y_i - E(Y|\mathbf{x}_i) + E(Y|\mathbf{x}_i) - p_i \mathbf{h}_i' \boldsymbol{\gamma}_h^* \right\}^2 \right] \\ &= V(\hat{\theta}_{HT}) + \frac{1}{N^2} E\left\{ \sum_{i \in A} w_i^2 \left( \frac{1}{p_i} - 1 \right) V(Y|\mathbf{x}_i) \right\} \\ &\quad + \frac{1}{N^2} E\left[ \sum_{i \in A} w_i^2 \left( \frac{1}{p_i} - 1 \right) \left\{ E(Y|\mathbf{x}_i) - p_i \mathbf{h}_i' \boldsymbol{\gamma}_h^* \right\}^2 \right] \end{aligned} \tag{2.14}$$

where the last equality follows because $y_i$ is conditionally independent of $E(Y|\mathbf{x}_i) - p_i \mathbf{h}_i' \boldsymbol{\gamma}_h^*$, conditioning on $\mathbf{x}_i$. Since the last term in (2.14) is non-negative, the inequality in (2.9) is established. Furthermore, if $E(Y|\mathbf{x}_i) = p_i \mathbf{h}_i' \boldsymbol{\alpha}$ for some $\boldsymbol{\alpha}$, then (2.10) holds and $E(\boldsymbol{\gamma}_h^*|\mathbf{x}_i) = \boldsymbol{\alpha}$, by the definition of $\boldsymbol{\gamma}_h^*$. Thus, $E(Y|\mathbf{x}_i) - p_i \mathbf{h}_i' \boldsymbol{\gamma}_h^* = -p_i \mathbf{h}_i' \left\{ \boldsymbol{\gamma}_h^* - E(\boldsymbol{\gamma}_h^*|\mathbf{x}_i) \right\} = o_p(1)$, implying that the last term in (2.14) is negligible. $\square$

In (2.9), $V_l$ is the lower bound of the asymptotic variance of PSA estimators of the form (2.3) satisfying (2.6). Any PSA estimator that has the asymptotic variance $V_l$ in (2.9) is optimal in the sense that it achieves the lower bound of the asymptotic variance among the class of PSA estimators with $\hat{\boldsymbol{\phi}}$ satisfying (2.2). The asymptotic variance of optimal PSA estimators of $\theta$ is equal to $V_l$ in (2.9). The PSA estimator using the maximum likelihood estimator of $\boldsymbol{\phi}_0$ does not necessarily achieve the lower bound of the asymptotic variance.

Condition (2.10) provides a way of constructing an optimal PSA estimator. First, we need an assumption for $E(Y|\mathbf{x})$, which is often called the outcome regression model. If the outcome regression model is a linear regression model of the form $E(Y|\mathbf{x}) = \beta_0 + \boldsymbol{\beta}_1'\mathbf{x}$, an optimal PSA estimator of $\theta$ can be obtained by solving

$$\sum_{i \in A} w_i \frac{r_i}{p_i(\boldsymbol{\phi})} (1, \mathbf{x}_i) = \sum_{i \in A} w_i (1, \mathbf{x}_i) . \tag{2.15}$$

Condition (2.15) is appealing because it says that the PSA estimator applied to $y = a + \mathbf{b}'\mathbf{x}$ leads to the original HT estimator. Condition (2.15) is called the calibration condition in survey sampling. The calibration condition applied to $\mathbf{x}$ makes full use of the information contained in it if the study variable is well approximated by a linear function of $\mathbf{x}$. Condition (2.15) was also used in Nevo (2003) and Kott (2006) under the linear regression model.

If we explicitly use a regression model for $E(Y|\mathbf{x})$, it is possible to construct an estimator that has asymptotic variance (2.9) and is not necessarily a PSA estimator. For example, if we assume that

$$E(Y|\mathbf{x}) = m(\mathbf{x}; \boldsymbol{\beta}_0) \tag{2.16}$$

for some function $m(\mathbf{x}; \cdot)$ known up to $\boldsymbol{\beta}_0$, we can use the model (2.16) directly to construct an optimal estimator of the form

$$\hat{\theta}_{opt} = \frac{1}{N} \sum_{i \in A} w_i \left[ m(\mathbf{x}_i; \hat{\boldsymbol{\beta}}) + \frac{r_i}{p_i(\hat{\boldsymbol{\phi}})} \left\{ y_i - m(\mathbf{x}_i; \hat{\boldsymbol{\beta}}) \right\} \right] , \tag{2.17}$$

where $\hat{\boldsymbol{\beta}}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}_0$ in the superpopulation model (2.16) and $\hat{\boldsymbol{\phi}}$ is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\phi}_0$ computed by (2.6). The following theorem shows that the optimal estimator (2.17) achieves the lower bound in (2.9).

**Theorem 2.2.** *Let the conditions of Theorem 2.1 hold. Assume that $\hat{\boldsymbol{\beta}}$ satisfies $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + O_p(n^{-1/2})$. Assume that, in the superpopulation model (2.16), $m(\mathbf{x};\boldsymbol{\beta})$ has continuous first-order partial derivatives in an open set containing $\boldsymbol{\beta}_0$. Under the joint distribution of the sampling mechanism, the response mechanism, and the superpopulation model (2.16), the estimator $\hat{\theta}_{opt}$ in (2.17) satisfies*

$$\sqrt{n}\left(\hat{\theta}_{opt} - \tilde{\theta}^*_{opt}\right) = o_p(1),$$

*where*

$$\tilde{\theta}^*_{opt} = N^{-1}\sum_{i\in A} w_i\left[m(\mathbf{x}_i;\boldsymbol{\beta}_0) + \frac{r_i}{p_i}\{y_i - m(\mathbf{x}_i;\boldsymbol{\beta}_0)\}\right],$$

*$p_i = p_i(\boldsymbol{\phi}_0)$, and $V(\tilde{\theta}^*_{opt})$ is equal to $V_l$ in (2.9).*

*Proof.* Define $\hat{\theta}_{opt}(\boldsymbol{\beta},\boldsymbol{\phi}) = N^{-1}\sum_{i\in A} w_i\left[m(\mathbf{x}_i;\boldsymbol{\beta}) + r_i p_i^{-1}(\boldsymbol{\phi})\{y_i - m(\mathbf{x}_i;\boldsymbol{\beta})\}\right]$. Note that $\hat{\theta}_{opt}$ in (2.17) can be written as $\hat{\theta}_{opt} = \hat{\theta}_{opt}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\phi}})$. Since

$$\frac{\partial}{\partial\boldsymbol{\beta}}\hat{\theta}_{opt}(\boldsymbol{\beta},\boldsymbol{\phi}) = \frac{1}{N}\sum_{i\in A} w_i\left\{\breve{m}(\mathbf{x}_i;\boldsymbol{\beta}) - \frac{r_i}{p_i(\boldsymbol{\phi})}\breve{m}(\mathbf{x}_i;\boldsymbol{\beta})\right\},$$

where $\breve{m}(\mathbf{x}_i;\boldsymbol{\beta}) = \partial m(\mathbf{x}_i;\boldsymbol{\beta})/\partial\boldsymbol{\beta}$, and

$$\frac{\partial}{\partial\boldsymbol{\phi}}\hat{\theta}_{opt}(\boldsymbol{\beta},\boldsymbol{\phi}) = \frac{1}{N}\sum_{i\in A} w_i r_i \mathbf{z}_i(\boldsymbol{\phi})\{y_i - m(\mathbf{x}_i;\boldsymbol{\beta})\},$$

where $\mathbf{z}_i(\boldsymbol{\phi}) = \partial\{p_i^{-1}(\boldsymbol{\phi})\}/\partial\boldsymbol{\phi}$, we have $E[\partial\{\hat{\theta}_{opt}(\boldsymbol{\beta},\boldsymbol{\phi})\}/\partial(\boldsymbol{\beta},\boldsymbol{\phi})|\boldsymbol{\beta}=\boldsymbol{\beta}_0,\boldsymbol{\phi}=\boldsymbol{\phi}_0] = \mathbf{0}$ and the condition of Randles (1982) is satisfied. Thus,

$$\hat{\theta}_{opt}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\phi}}) = \hat{\theta}_{opt}(\boldsymbol{\beta}_0,\boldsymbol{\phi}_0) + o_p\left(n^{-1/2}\right) = \tilde{\theta}^*_{opt} + o_p\left(n^{-1/2}\right)$$

and the variance of $\tilde{\theta}^*_{opt}$ is equal to $V_l$, the lower bound of the asymptotic variance. □

The (asymptotic) optimality of the estimator in (2.17) is justified under the joint distribution of the response model (2.4) and the superpopulation model (2.16). When both models are correct, $\hat{\theta}_{opt}$ is optimal and the choice of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ does not affect the efficiency of the $\hat{\theta}_{opt}$ as long as $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ is $\sqrt{n}$-consistent. Robins et al. (1994) also advocated using $\hat{\theta}_{opt}$ in (2.17) under simple random sampling.

**Remark 2.1.** *When the response model is correct and the superpopulation model (2.16) is not necessarily correct, the choice of $\hat{\boldsymbol{\beta}}$ does affect the efficiency of the optimal estimator. Cao et al. (2009) considered optimal estimation when only the response model is correct. Using Taylor linearization, the optimal estimator in (2.17) with $\hat{\boldsymbol{\phi}}$ satisfying (2.6) is asymptotically equivalent to*

$$\tilde{\theta}(\boldsymbol{\beta}) = \sum_{i \in A} w_i \left[ m(\mathbf{x}_i; \boldsymbol{\beta}) + \frac{r_i}{p_i} \{y_i - m(\mathbf{x}_i; \boldsymbol{\beta})\} - \left(\frac{r_i}{p_i} - 1\right) \mathbf{c}_\beta{}' p_i \mathbf{h}_i \right],$$

*where $\mathbf{c}_\beta$ is the probability limit of $\hat{\mathbf{c}}_\beta = \{\sum_{i \in A} w_i r_i \mathbf{z}_i(\hat{\boldsymbol{\phi}}) \hat{p}_i \mathbf{h}_i'(\hat{\boldsymbol{\phi}})\}^{-1} \sum_{i \in A} w_i r_i \mathbf{z}_i(\hat{\boldsymbol{\phi}}) \{y_i - m(\mathbf{x}_i; \boldsymbol{\beta})\}$ and $\mathbf{z}_i(\boldsymbol{\phi}) = \partial\{p_i^{-1}(\boldsymbol{\phi})\}/\partial \boldsymbol{\phi}$. The asymptotic variance is then equal to*

$$V\left\{\tilde{\theta}(\boldsymbol{\beta})\right\} = V\left(\hat{\theta}_{HT}\right) + E\left[\sum_{i \in A} w_i^2 \frac{1 - p_i}{p_i} \{y_i - m(\mathbf{x}_i; \boldsymbol{\beta}) - \mathbf{c}_\beta{}' p_i \mathbf{h}_i\}^2\right].$$

*Thus, an optimal estimator of $\boldsymbol{\beta}$ can be computed by finding $\hat{\boldsymbol{\beta}}$ that minimizes*

$$Q(\boldsymbol{\beta}) = \sum_{i \in A} w_i^2 r_i \frac{1 - \hat{p}_i}{\hat{p}_i^2} \left\{y_i - m(\mathbf{x}_i; \boldsymbol{\beta}) - \hat{\mathbf{c}}_\beta' \hat{p}_i \mathbf{h}_i(\hat{\boldsymbol{\phi}})\right\}^2.$$

*The resulting estimator is design-optimal in the sense that it minimizes the asymptotic variance under the response model.*

## 2.3 Augmented propensity score model

In this section, we consider optimal PSA estimation. Note that the optimal estimator $\hat{\theta}_{opt}$ in (2.17) is not necessarily written as a PSA estimator form in (2.3). It is in the PSA estimator form if it satisfies $\sum_{i \in A} w_i r_i \hat{p}_i^{-1} m(\mathbf{x}_i; \hat{\boldsymbol{\beta}}) = \sum_{i \in A} w_i m(\mathbf{x}_i; \hat{\boldsymbol{\beta}})$. Thus, we can construct an optimal PSA estimator by including $m(\mathbf{x}_i; \hat{\boldsymbol{\beta}})$ in the model for the propensity score. Specifically, given $\hat{m}_i = m(\mathbf{x}_i; \hat{\boldsymbol{\beta}})$, $\hat{p}_i = p_i(\hat{\boldsymbol{\phi}})$ and $\hat{\mathbf{h}}_i = \mathbf{h}_i(\hat{\boldsymbol{\phi}})$, where $\hat{\boldsymbol{\phi}}$ is obtained from (2.6), we augment the response model by

$$p_i^*(\hat{\boldsymbol{\phi}}, \boldsymbol{\lambda}) \equiv \frac{\hat{p}_i}{\hat{p}_i + (1-\hat{p}_i)\exp\left(\lambda_0 + \lambda_1 \hat{m}_i\right)}, \tag{2.18}$$

where $\boldsymbol{\lambda} = (\lambda_0, \lambda_1)'$ is the Lagrange multiplier which is used to incorporate the additional constraint. If $(\lambda_0, \lambda_1)' = \mathbf{0}$, then $p_i^*(\hat{\boldsymbol{\phi}}, \boldsymbol{\lambda}) = \hat{p}_i$. The augmented response probability $p_i^*(\hat{\boldsymbol{\phi}}, \boldsymbol{\lambda})$ always takes values between 0 and 1. The augmented response probability model (2.18) can be derived by minimizing the Kullback-Leibler distance $\sum_{i \in A} w_i r_i q_i^* \log(q_i^*/q_i)$, where $q_i^* = (1-p_i^*)/p_i^*$ and $q_i = (1-\hat{p}_i)/\hat{p}_i$, subject to the constraint $\sum_{i \in A} w_i (r_i/p_i^*)(1, \hat{m}_i) = \sum_{i \in A} w_i (1, \hat{m}_i)$.

Using (2.18), the optimal PSA estimator is computed by

$$\hat{\theta}_{PSA}^* = \frac{1}{N} \sum_{i \in A} w_i \frac{r_i}{p_i^*(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\lambda}})} y_i, \tag{2.19}$$

where $\hat{\boldsymbol{\lambda}}$ satisfies

$$\sum_{i \in A} w_i \frac{r_i}{p_i^*(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\lambda}})} \left(1, \hat{m}_i\right) = \sum_{i \in A} w_i \left(1, \hat{m}_i\right). \tag{2.20}$$

Under the response model (2.4), it can be shown that

$$\hat{\theta}_{PSA}^* = \frac{1}{N} \sum_{i \in A} w_i \left\{ \hat{b}_0 + \hat{b}_1 \hat{m}_i + \frac{r_i}{\hat{p}_i} \left( y_i - \hat{b}_0 - \hat{b}_1 \hat{m}_i \right) \right\} + o_p\left(n^{-1/2}\right),$$

where

$$\begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \end{pmatrix} = \left\{ \sum_{i \in A} w_i r_i \left( \frac{1}{\hat{p}_i} - 1 \right) \begin{pmatrix} 1 \\ \hat{m}_i \end{pmatrix} \begin{pmatrix} 1 \\ \hat{m}_i \end{pmatrix}' \right\}^{-1} \sum_{i \in A} w_i r_i \left( \frac{1}{\hat{p}_i} - 1 \right) \begin{pmatrix} 1 \\ \hat{m}_i \end{pmatrix} y_i. \tag{2.21}$$

Furthermore, by the argument for Theorem 2.1, we can establish that

$$\hat{\theta}_{PSA}^* = \frac{1}{N} \sum_{i \in A} w_i \left\{ b_0 + b_1 \hat{m}_i + \boldsymbol{\gamma}_{h2}' p_i \mathbf{h}_i + \frac{r_i}{p_i} \left( y_i - b_0 - b_1 \hat{m}_i - \boldsymbol{\gamma}_{h2}' p_i \mathbf{h}_i \right) \right\} + o_p\left(n^{-1/2}\right),$$

where $(b_0, b_1, \boldsymbol{\gamma}'_{h2})$ is the probability limit of $(\hat{b}_0, \hat{b}_1, \hat{\boldsymbol{\gamma}}'_{h2})$ with

$$\hat{\boldsymbol{\gamma}}_{h2} = \left\{ \sum_{i \in A} w_i r_i \mathbf{z}_i(\hat{\boldsymbol{\phi}}) \hat{p}_i \mathbf{h}'_i(\hat{\boldsymbol{\phi}}) \right\}^{-1} \sum_{i \in A} w_i r_i \mathbf{z}_i(\hat{\boldsymbol{\phi}})(y_i - \hat{b}_0 - \hat{b}_1 \hat{m}_i) \tag{2.22}$$

and the effect of estimating $\boldsymbol{\phi}_0$ in $\hat{p}_i = p(\mathbf{x}_i; \hat{\boldsymbol{\phi}})$ can be safely ignored.

Note that, under the response model (2.4), $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\lambda}})$ in (2.19) converges in probability to $(\boldsymbol{\phi}_0, \mathbf{0})$, where $\boldsymbol{\phi}_0$ is the true parameter in (2.4). Thus, the propensity score from the augmented model converges to the true response probability. Because $\hat{\boldsymbol{\lambda}}$ converges to zero in probability, the choice of $\hat{\boldsymbol{\beta}}$ in $\hat{m}_i = m(x_i; \hat{\boldsymbol{\beta}})$ does not play a role for the asymptotic unbiasedness of the PSA estimator. The asymptotic variances are changed for different choices of $\hat{\boldsymbol{\beta}}$.

Under the superpopulation model (2.16), $\hat{b}_0 + \hat{b}_1 \hat{m}_i \to E(Y|\mathbf{x}_i)$ in probability. Thus, the optimal PSA estimator in (2.19) is asymptotically equivalent to the optimal estimator in (2.17). Incorporating $\hat{m}_i$ into the calibration equation to achieve optimality is close in spirit to the model-calibration method proposed by Wu and Sitter (2001).

## 2.4 Variance estimation

We now discuss variance estimation of PSA estimators under the assumed response model. Singh and Folsom (2000) and Kott (2006) discussed variance estimation for certain types of PSA estimators. Kim and Kim (2007) discussed variance estimation when the PSA estimator is computed with the maximum likelihood method.

We consider variance estimation for the PSA estimator of the form (2.3) where $\hat{p}_i = p_i(\hat{\boldsymbol{\phi}})$ is constructed to satisfy (2.6) for some $\mathbf{h}_i(\boldsymbol{\phi}) = \mathbf{h}(\mathbf{x}_i; \boldsymbol{\phi}, \hat{\boldsymbol{\beta}})$, where $\hat{\boldsymbol{\beta}}$ is obtained using the postulated superpopulation model. Let $\boldsymbol{\beta}^*$ be the probability limit of $\hat{\boldsymbol{\beta}}$ under the response model. Note that $\boldsymbol{\beta}^*$ is not necessarily equal to $\boldsymbol{\beta}_0$ in (2.16) since we are not assuming that the postulated superpopulation model is correctly specified in this section.

Using the argument for the Taylor linearization (2.13) used in the proof of Theorem 2.1, the PSA estimator satisfies

$$\hat{\theta}_{PSA} = \frac{1}{N}\sum_{i\in A} w_i\eta_i(\boldsymbol{\phi}_0, \boldsymbol{\beta}^*) + o_p\left(n^{-1/2}\right), \tag{2.23}$$

where

$$\eta_i(\boldsymbol{\phi}, \boldsymbol{\beta}) = p_i(\boldsymbol{\phi})\mathbf{h}_i'(\boldsymbol{\phi}, \boldsymbol{\beta})\boldsymbol{\gamma}_h^* + \frac{r_i}{p_i(\boldsymbol{\phi})}\left\{y_i - p_i(\boldsymbol{\phi})\mathbf{h}_i'(\boldsymbol{\phi}, \boldsymbol{\beta})\boldsymbol{\gamma}_h^*\right\}, \tag{2.24}$$

$\mathbf{h}_i(\boldsymbol{\phi}, \boldsymbol{\beta}) = \mathbf{h}(\mathbf{x}_i; \boldsymbol{\phi}, \boldsymbol{\beta})$ and $\boldsymbol{\gamma}_h^*$ is defined as in (2.8) with $\mathbf{h}_i$ replaced by $\mathbf{h}_i(\boldsymbol{\phi}_0, \boldsymbol{\beta}^*)$. Since $p_i(\hat{\boldsymbol{\phi}})$ satisfies (2.6) with $\mathbf{h}_i(\boldsymbol{\phi}) = \mathbf{h}(\mathbf{x}_i; \boldsymbol{\phi}, \hat{\boldsymbol{\beta}})$, $\hat{\theta}_{PSA} = N^{-1}\sum_{i\in A} w_i\eta_i(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\beta}})$ holds and the linearization in (2.23) can be expressed as $N^{-1}\sum_{i\in A} w_i\eta_i(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\beta}}) = N^{-1}\sum_{i\in A} w_i\eta_i(\boldsymbol{\phi}_0, \boldsymbol{\beta}^*) + o_p(n^{-1/2})$. Thus, if $(\mathbf{x}_i, y_i, r_i)$ are independent and identically distributed (IID), then $\eta_i(\boldsymbol{\phi}_0, \boldsymbol{\beta}^*)$ are IID even though $\eta_i(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\beta}})$ are not necessarily IID. Because $\eta_i\left(\boldsymbol{\phi}_0, \boldsymbol{\beta}^*\right)$ are IID, we can apply the standard complete sample method to estimate the variance of $\hat{\eta}_{HT} = N^{-1}\sum_{i\in A} w_i\eta_i\left(\boldsymbol{\phi}_0, \boldsymbol{\beta}^*\right)$, which is asymptotically equivalent to the variance of $\hat{\theta}_{PSA} = N^{-1}\sum_{i\in A} w_i\eta_i(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\beta}})$. See Kim and Rao (2009).

To derive the variance estimator, we assume that the variance estimator

$$\hat{V} = N^{-2}\sum_{i\in A}\sum_{j\in A}\Omega_{ij}g_ig_j$$

satisfies $\hat{V}/V(\hat{g}_{HT}|\mathcal{F}_N) = 1 + o_p(1)$ for some $\Omega_{ij}$ related to the joint inclusion probability, where $\hat{g}_{HT} = N^{-1}\sum_{i\in A} w_ig_i$ for any $g$ with a finite second moment and $V(g_{HT}|\mathcal{F}_N) = N^{-2}\sum_{i=1}^N\sum_{j=1}^N \Omega_{N\cdot ij}g_ig_j$,, for some $\Omega_{N\cdot ij}$. We also assume that

$$\sum_{i=1}^N |\Omega_{N\cdot ij}| = O(n^{-1}N). \tag{2.25}$$

To obtain the total variance, the *reverse framework* of Fay (1992), Shao and Steel (1999), and Kim and Rao (2009) is considered. In this framework, the finite population is first divided into two groups, a population of respondents and a population of non-respondents. Given the population, the sample $A$ is selected according to a probability sampling design. Thus, selection of the population respondents from the whole finite population is treated as the first-phase sampling and the selection of the sample respondents from the population respondents is treated as the second-phase sampling in the reverse framework. The total variance of $\hat{\eta}_{HT}$ can be written as

$$V(\hat{\eta}_{HT}|\mathcal{F}_N) = V_1 + V_2 = E\{V(\hat{\eta}_{HT}|\mathcal{F}_N, \mathcal{R}_N)|\mathcal{F}_N\} + V\{E(\hat{\eta}_{HT}|\mathcal{F}_N, \mathcal{R}_N)|\mathcal{F}_N\}. \quad (2.26)$$

The conditional variance term $V(\hat{\eta}_{HT}|\mathcal{F}_N, \mathcal{R}_N)$ in (2.26) can be estimated by

$$\hat{V}_1 = N^{-2} \sum_{i \in A} \sum_{j \in A} \Omega_{ij} \hat{\eta}_i \hat{\eta}_j, \quad (2.27)$$

where $\hat{\eta}_i = \eta_i(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\beta}})$ is defined in (2.24) with $\boldsymbol{\gamma}_h^*$ replaced by a consistent estimator such as $\hat{\boldsymbol{\gamma}}_h^* = \{\sum_{i \in A} w_i r_i \mathbf{z}_i(\hat{\boldsymbol{\phi}}) \hat{p}_i \hat{\mathbf{h}}_i'\}^{-1} \sum_{i \in A} w_i r_i \mathbf{z}_i(\hat{\boldsymbol{\phi}}) y_i$, and $\hat{\mathbf{h}}_i = \mathbf{h}(\mathbf{x}_i; \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\beta}})$. To show that $\hat{V}_1$ is also consistent for $V_1$ in (2.26), it suffices to show that $V\{n \cdot V(\hat{\eta}_{HT}|\mathcal{F}_N, \mathcal{R}_N)|\mathcal{F}_N\} = o(1)$, which follows by (2.25) and the existence of the fourth moment. See Kim et al. (2006). The second term $V_2$ in (2.26) is

$$\begin{aligned} V\{E(\hat{\eta}_{HT}|\mathcal{F}_N, \mathcal{R}_N)|\mathcal{F}_N\} &= V\left(N^{-1} \sum_{i=1}^{N} \eta_i | \mathcal{F}_N\right) \\ &= \frac{1}{N^2} \sum_{i=1}^{N} \frac{1-p_i}{p_i} \left(y_i - p_i \mathbf{h}_i^{*'} \boldsymbol{\gamma}_h^*\right)^2, \end{aligned}$$

where $\mathbf{h}_i^* = \mathbf{h}(\mathbf{x}_i; \boldsymbol{\phi}_0, \boldsymbol{\beta}^*)$. A consistent estimator of $V_2$ can be derived as

$$\hat{V}_2 = \frac{1}{N^2} \sum_{i \in A} w_i r_i \frac{1-\hat{p}_i}{\hat{p}_i^2} \left(y_i - \hat{p}_i \hat{\mathbf{h}}_i' \hat{\boldsymbol{\gamma}}_h^*\right)^2, \quad (2.28)$$

where $\boldsymbol{\gamma}_h^*$ is defined after (2.27). Therefore,

$$\hat{V}\left(\hat{\theta}_{PSA}\right) = \hat{V}_1 + \hat{V}_2, \quad (2.29)$$

is consistent for the variance of the PSA estimator defined in (2.3) with $\hat{p}_i = p_i(\hat{\boldsymbol{\phi}})$ satisfying (2.6), where $\hat{V}_1$ is in (2.27) and $\hat{V}_2$ is in (2.28).

Note that the first term of the total variance is $V_1 = O_p(n^{-1})$, but the second term is $V_2 = O_p(N^{-1})$. Thus, when the sampling fraction $nN^{-1}$ is negligible, that is, $nN^{-1} = o(1)$, the second term $V_2$ can be ignored and $\hat{V}_1$ is a consistent estimator of the total variance. Otherwise, the second term $V_2$ should be taken into consideration, so that a consistent variance estimator can be constructed as in (2.29).

**Remark 2.2.** *The variance estimation of the optimal PSA estimator with augmented propensity model (2.18) with $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\lambda}})$ satisfying (2.20) can be derived by (2.29) using $\hat{\eta}_i = \hat{b}_0 + \hat{b}_1 \hat{m}_i + \hat{\boldsymbol{\gamma}}'_{h2} \hat{p}_i \hat{\mathbf{h}}_i + r_i \hat{p}_i^{-1}(y_i - \hat{b}_0 - \hat{b}_1 \hat{m}_i - \hat{\boldsymbol{\gamma}}'_{h2} \hat{p}_i \hat{\mathbf{h}}_i)$ where $(\hat{b}_0, \hat{b}_1)$ and $\hat{\boldsymbol{\gamma}}_{h2}$ are defined in (2.21) and (2.22), respectively.*

## 2.5 Simulation study

### 2.5.1 Study One

Two simulation studies were performed to investigate the properties of the proposed method. In the first simulation, we generated a finite population of size $N = 10,000$ from the following multivariate normal distribution:

$$\begin{pmatrix} x_1 \\ x_2 \\ e \end{pmatrix} \sim N \left[ \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right].$$

The variable of interest $y$ was constructed as $y = 1 + x_1 + e$. We also generated response indicator variables $r_i$ independently from a Bernoulli distribution with probability

$$p_i = \frac{\exp(2 + x_{2i})}{1 + \exp(2 + x_{2i})}.$$

From the finite population, we used simple random sampling to select two samples of size, $n = 100$ and $n = 400$, respectively. We used $B = 5,000$ Monte Carlo samples in the simulation. The average response rate was about 69.6%.

To compute the propensity score, a response model of the form

$$p(\mathbf{x}; \boldsymbol{\phi}) = \frac{\exp(\phi_0 + \phi_1 x_2)}{1 + \exp(\phi_0 + \phi_1 x_2)} \tag{2.30}$$

was postulated and an outcome regression model of the form

$$m(\mathbf{x}; \boldsymbol{\beta}) = \beta_0 + \beta_1 x_1 \tag{2.31}$$

was postulated to obtain the optimal PSA estimators. Thus, both models are correctly specified. From each sample, we computed four estimators of $\theta = N^{-1}\sum_{i=1}^{N} y_i$:

1. (PSA-MLE) : PSA estimator in (2.3) with $\hat{p}_i = p_i(\hat{\boldsymbol{\phi}})$ and $\hat{\boldsymbol{\phi}}$ being the maximum likelihood estimator of $\boldsymbol{\phi}$.

2. (PSA-CAL) : PSA estimator in (2.3) with $\hat{p}_i$ satisfying the calibration constraint (2.15) on $(1, x_{2i})$.

3. (AUG) : Augmented PSA estimator in (2.19).

4. (OPT) : Optimal estimator in (2.17).

In the augmented PSA estimators, $\hat{\boldsymbol{\phi}}$ was computed by the maximum likelihood method. Under model (2.30), the maximum likelihood estimator of $\boldsymbol{\phi} = (\phi_0, \phi_1)'$ was computed by solving (2.6) with $\mathbf{h}_i(\boldsymbol{\phi}) = (1, x_{2i})'$. Parameter $(\beta_0, \beta_1)$ for the outcome regression model was computed using ordinary least squares, regressing $y$ on $x_1$. In addition to the point estimators, we also computed the variance estimators of the point estimators. The variance estimators of the PSA estimators were computed using the pseudo-values in (2.24) and the $\mathbf{h}_i(\boldsymbol{\phi})$ corresponding to each estimator. For the augmented PSA estimators, the pseudo-values were computed by the method in Remark 2.

Table 2.1 presents the Monte Carlo biases, variances, and mean square errors of the four point estimators and the Monte Carlo percent relative biases and t-statistics of the variance estimators of the estimators. The percent relative bias of a variance estimator $\hat{V}(\hat{\theta})$ is calculated as $100 \times \{V_{MC}(\hat{\theta})\}^{-1}[E_{MC}\{\hat{V}(\hat{\theta})\} - V_{MC}(\hat{\theta})]$, where $E_{MC}(\cdot)$ and $V_{MC}(\cdot)$ denote the Monte Carlo expectation and the Monte Carlo variance, respectively. The t-statistic in Table 1 is the test statistic for testing the zero bias of the variance estimator. See Kim (2004).

Based on the simulation results in Table 1, we have the following conclusions.

1. All of the PSA estimators are asymptotically unbiased because the response model (2.30) is correctly specified. The PSA estimator using the calibration method is slightly more efficient than the PSA estimator using the maximum likelihood estimator, because the last term of (2.14) is smaller for the calibration method as the predictor for $E(Y|\mathbf{x}_i) = \beta_0 + \beta_1 x_{1i}$ is better approximated by a linear function of $(1, x_{2i})$ than by a linear function of $(\hat{p}_i, \hat{p}_i x_{2i})$.

2. The augmented PSA estimator is more efficient than the direct PSA estimator (2.3). The augmented PSA estimator is constructed by using the correctly specified regression model (2.31) and so it is asymptotically equivalent to the optimal PSA estimator in (2.17).

3. Variance estimators are all approximately unbiased. There are some modest biases in the variance estimators of the PSA estimators when the sample size is small ($n = 100$).

### 2.5.2 Study Two

In the second simulation study, we further investigated the PSA estimators with a non-linear outcome regression model under an unequal probability sampling design. We generated two stratified finite populations of $(x, y)$ with four strata ($h = 1, 2, 3, 4$), where $x_{hi}$ were independently generated from a normal distribution $N(1, 1)$ and $y_{hi}$ were dichotomous variables that take values of 1 or 0 from a Bernoulli distribution with probability $p_{1yhi}$ or $p_{1yhi}$. Two different probabilities were used for two populations, respectively :

1. Population 1 (Pop1): $p_{1yhi} = 1/\left\{1 + \exp(0.5 - 2x)\right\}$

2. Population 2 (Pop2): $p_{2yhi} = 1/\left[1 + \exp\{0.25(x - 1.5)^2 - 1.5\}\right]$

In addition to $x_{hi}$ and $y_{hi}$, the response indicator variables $r_{hi}$ were generated from a Bernoulli distribution with probability $p_{hi} = 1/\left\{1 + \exp(-1.5 + 0.7x_{hi})\right\}$. The sizes of the four strata were $N_1 = 1,000$, $N_2 = 2,000$, $N_3 = 3,000$, and $N_4 = 4,000$, respectively. In each of the two sets of finite population, a stratified sample of size $n = 400$ was independently generated without replacement, where a simple random sample of size $n_h = 100$ was selected from each stratum. We used $B = 5,000$ Monte Carlo samples in this simulation. The average response rate was about 67%.

To compute the propensity score, a response model of the form

$$p(x; \boldsymbol{\phi}) = \frac{\exp(\phi_0 + \phi_1 x)}{1 + \exp(\phi_0 + \phi_1 x)}$$

was postulated for parameter estimation. To obtain the augmented PSA estimator, a model for the variable of interest of the form

$$m(x; \boldsymbol{\beta}) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \tag{2.32}$$

was postulated. Thus, model (2.32) is a true model under (Pop1), but it is not a true model under (Pop2).

We computed four estimators:

1. (PSA-MLE): PSA estimator in (2.3) using the maximum likelihood estimator of $\boldsymbol{\phi}$.

2. (PSA-CAL): PSA estimator in (2.3) with $\hat{p}_i$ satisfying the calibration constraint (2.15) on $(1, x)$.

3. (AUG-1) : Augmented PSA estimator $\hat{\theta}^*_{PSA}$ in (2.19) with $\hat{\boldsymbol{\beta}}$ computed by the maximum likelihood method.

4. (AUG-2) : Augmented PSA estimator $\hat{\theta}^*_{PSA}$ in (2.19) with $\hat{\boldsymbol{\beta}}$ computed by the method of Cao et al. (2009) discussed in Remark 1.

We considered the the augmented PSA estimator in (2.19) with $\hat{p}_i = p_i(\hat{\boldsymbol{\phi}})$, where $\hat{\boldsymbol{\phi}}$ is the maximum likelihood estimator of $\boldsymbol{\phi}$. The first augmented PSA estimator (AUG-1) used $\hat{m}_i = m(x_i; \hat{\boldsymbol{\beta}})$ with $\hat{\boldsymbol{\beta}}$ found by solving $\sum_{h=1}^{4} \sum_{i \in A_h} w_{hi} r_{hi} \{y_{hi} - m(x_{hi}; \boldsymbol{\beta})\}(1, x_{hi}) =$

$\mathbf{0}$, where $A_h$ is the set of indices appearing in the sample for stratum $h$ and $w_{hi}$ is the sampling weight of unit $i$ for stratum $h$.

Table 2.2 presents the simulation results for each method. In each population, the augmented PSA estimator shows some improvement comparing to the PSA estimator using the maximum likelihood estimator of $\boldsymbol{\phi}$ or the calibration estimator of $\boldsymbol{\phi}$ in terms of variance. Under (Pop1), since model (2.32) is true, there is essentially no difference between the augmented PSA estimators using different methods of estimating $\boldsymbol{\beta}$. However, under (Pop2), where the assumed outcome regression model (2.32) is incorrect, the augmented PSA estimator with $\hat{\boldsymbol{\beta}}$ computed by the method of Cao et al. (2009) results in slightly better efficiency, which is consistent with the theory in Remark 1. Variance estimates are approximately unbiased in all cases in the simulation study.

## 2.6 Conclusion

We have considered the problem of estimating the finite population mean of $y$ under nonresponse using the propensity score method. The propensity score is computed from a parametric model for the response probability, and some asymptotic properties of PSA estimators are discussed. In particular, the optimal PSA estimator is derived with an additional assumption for the distribution of $y$. The propensity score for the optimal PSA estimator can be implemented by the augmented propensity model presented in Section 2.3. The resulting estimator is still consistent even when the assumed outcome regression model fails to hold.

We have restricted our attention to missing-at-random mechanisms in which the response probability depends only on the always-observed $\mathbf{x}$. If the response mechanism also depends on $y$, PSA estimation becomes more challenging. PSA estimation when missingness is not at random is beyond the scope of this article and will be a topic of future research.

## Acknowledgement

The research was partially supported by a Cooperative Agreement between the US Department of Agriculture Natural Resources Conservation Service and Iowa State University. The authors wish to thank F. Jay Breidt, three anonymous referees, and the associate editor for their helpful comments.

**Table 2.1 Monte Carlo bias, variance and mean square error(MSE) of the four point estimators and percent relative biases (R.B.) and t-statistics(t-stat) of the variance estimators based on 5,000 Monte Carlo samples**

| n | Method | $\hat{\theta}$ | | | $V(\hat{\theta})$ | |
|---|---|---|---|---|---|---|
| | | Bias | Variance | MSE | R.B. (%) | t-stat |
| 100 | (PSA-MLE) | -0.01 | 0.0315 | 0.0317 | -2.34 | -1.12 |
| | (PSA-CAL) | -0.01 | 0.0308 | 0.0309 | -3.56 | -1.70 |
| | (AUG) | 0.00 | 0.0252 | 0.0252 | -0.61 | -0.30 |
| | (OPT) | 0.00 | 0.0252 | 0.0252 | -0.21 | -0.10 |
| 400 | (PSA-MLE) | -0.01 | 0.00737 | 0.00746 | 0.35 | 0.17 |
| | (PSA-CAL) | -0.01 | 0.00724 | 0.00728 | 0.29 | 0.14 |
| | (AUG) | 0.00 | 0.00612 | 0.00612 | 0.07 | 0.03 |
| | (OPT) | 0.00 | 0.00612 | 0.00612 | -0.14 | -0.07 |

**Table 2.2. Monte Carlo bias, variance and mean square error of the four point estimators and percent relative biases (R.B.) and t-statistics of the variance estimators, based on 5,000 Monte Carlo samples**

| Population | Method | $\hat{\theta}_{PSA}$ | | | $V(\hat{\theta}_{PSA})$ | |
|---|---|---|---|---|---|---|
| | | Bias | Variance | MSE | R.B. (%) | t-stat |
| Pop1 | (PSA-MLE) | 0.00 | 0.000750 | 0.000762 | -1.13 | -0.57 |
| | (PSA-CAL) | 0.00 | 0.000762 | 0.000769 | -1.45 | -0.72 |
| | (AUG-1) | 0.00 | 0.000745 | 0.000757 | -1.73 | -0.86 |
| | (AUG-2) | 0.00 | 0.000745 | 0.000757 | -1.83 | -0.91 |
| Pop2 | (PSA-MLE) | 0.00 | 0.000824 | 0.000826 | 0.29 | 0.14 |
| | (PSA-CAL) | 0.00 | 0.000829 | 0.000835 | -0.94 | -0.46 |
| | (AUG-1) | 0.00 | 0.000822 | 0.000823 | -0.71 | -0.35 |
| | (AUG-2) | 0.00 | 0.000820 | 0.000821 | -0.61 | -0.30 |

# CHAPTER 3. PROPENSITY SCORE ADJUSTMENT METHOD FOR NONIGNORABLE NONRESPONSE

To be submitted to Journal of the American Statistical Association

Minsun Kim Riddles and Jae-kwang Kim

## Abstract

Propensity score adjustment is a popular technique for handling unit nonresponse in sample surveys. If the response probability depends on the study variable that is subject to missingness, estimating the response probability often relies on additional distributional assumptions about the study variable. Instead of making fully parametric assumptions about the population distribution of the study variable and the response mechanism, we propose a new approach of maximum likelihood estimation that is based on the distributional assumptions of the observed part of the sample. Since the model for the observed part of the sample can be verified from the data, the proposed method is less sensitive to failure of the assumed model of the outcomes. Generalized method of moments can be used to improve the efficiency of the proposed estimator. Variance estimation is discussed and results from two limited simulation studies are presented to compare the performance of the proposed method with the existing methods.

*Key words:* Exponential tilting model, Nonresponse weighting adjustment, Nonresponse error, Not missing at random

## 3.1 Introduction

Analysis of survey data involves the assumption that data are collected from a randomly selected sample that is representative of the target population. When the representativeness of the sample at hand is in question, the relationship of the variables in the sample does not necessarily hold in the population. In practice, survey estimates are subject to various forms of selection bias that stems from a systematic difference between the sample and the target population. Sources of selection bias at the unit-level include discrepancies between the sampling frame and the target population, called coverage error, and failure to obtain responses from the full sample, called nonresponse error. Reducing the selection bias is a crucial part of improving the scientific foundation for generalizing survey results to the target population.

Nonresponse error has become a major problem in sample surveys as participation rates have declined in many surveys. Weighting adjustments are commonly used to adjust for unit nonresponse. Classical approaches include poststratification [Holt and Smith (1979)], regression weighting [Bethlehem (1988)], and raking ratio estimation [Deville et al. (1993)]. Propensity score (PS) weighting, which increases the sampling weights of the respondents using their inverse response probabilities, is a popular approach for handling unit nonresponse. Most of the existing work on PS modeling for unit nonresponse assumes an ignorable mechanism for missing data, and the auxiliary variables for the PS model are observed throughout the sample. For examples, see Ekholm and Laaksonen (1991), Fuller et al. (1994), Lindström and Särndal (1999), and Kott (2006). If the missing mechanism is not ignorable, that is, the response mechanism is related to the variable of interest directly or indirectly, then the realized sample may over-represent individuals that are interested in the topic of the survey and survey estimates may be biased [Groves et al. (2004)]. PS modeling with nonignorable nonresponse is challenging because the covariates for the PS model are not always observed.

In this paper, we consider parameter estimation for a PS model under nonignorable nonresponse. In order to estimate the parameters in the PS model consistently, additional assumptions are imposed on the distribution of the study variable that is subject to missingness. A fully parametric approach, which makes parametric assumptions about the population distribution of the study variable can be used to estimate the parameters in the response model, but the estimates can be very sensitive against the failure of the assumed response model. Instead of the fully parametric approach, we consider an alternative modeling approach that uses parametric model assumptions about the study variable in the responding part of the sample. The resulting PS weighed estimator is shown to be consistent under the correct specification of the response probability model.

In addition to PS modeling, efficient or optimal estimation of the parameters in the PS model and inference with the estimated PS model are also addressed. The maximum likelihood estimation of the PS model parameters does not necessarily lead to optimal PS estimation [Kim and Kim (2007)]. Recent works such as Tan (2006), Cao et al. (2009), Kim and Riddles (2012) have partially addressed these issues under ignorable response mechanisms, but extension to nonignorable nonresponse model is not well developed. Recent works such as Chang and Kott (2008), Kott and Chang (2010), and Wang et al. (2013) have addressed parameter estimation under nonignorable nonresponse, but the optimality was not discussed. Efficient estimation and valid inferential tools for nonignorable nonresponse are discussed in Section 3.5. The proposed estimators are directly applicable to the survey sampling setup, as illustrated in Section 3.6.2.

In Section 3.2, the basic setup is introduced. In Section 3.3, an approach to estimating response model parameters is proposed. In Section 3.4, variance estimation is discussed. In Section 3.5, the generalized method of moments is used to incorporate the auxiliary variables that are observed in the sample. Results from two simulation studies are presented in Section 3.6, and concluding remarks are made in Section 3.7.

## 3.2 Basic setup

Consider an infinite population with three random variables $(\mathbf{X}, Y, \delta)$. Let $(\mathbf{x}_i, y_i)$, $i = 1, 2, \cdots, n$ be $n$ independent realizations of $(\mathbf{X}, Y)$ in the population. In addition, we assume that $\delta$ is dichotomous taking values of 1 or 0, and $y_i$ is observed if and only if $\delta_i = 1$. The auxiliary variable $\mathbf{x}_i$ is always observed for $i = 1, 2, \cdots, n$. We are interested in estimating $\theta$, which is uniquely determined by solving $E\{U(\theta; \mathbf{X}, Y)\} = 0$.

In the presence of missing data, if the true response probability $\pi_i$ were known, an unbiased consistent estimator of $\theta$, $\hat{\theta}_{PS1}$, could be obtained by solving $U_{PS1}(\theta) = \sum_{i=1}^{n} \delta_i \pi_i^{-1} U_i(\theta) = 0$ for $\theta$, where $U_i(\theta) = U(\theta; \mathbf{x}_i, y_i)$. However, the true response probabilities are unknown in general and need to be estimated consistently from the sample.

On the other hand, if $\pi_{i2} = Pr\,(\delta_i = 1 \mid \mathbf{x}_i, y_i)$ are known, then the resulting estimator $\hat{\theta}_{PS2}$ obtained by solving

$$U_{PS2}(\theta) = \sum_{i=1}^{n} \frac{\delta_i}{\pi_{i2}} U_i(\theta) = 0 \tag{3.1}$$

is also unbiased, because

$$\begin{aligned} E\left\{U_{PS2}(\theta)\right\} &= E\left\{\sum_{i=1}^{n} E\left(\delta_i \pi_{i2}^{-1} U_i(\theta) \mid \mathbf{x}_i, y_i\right)\right\} \\ &= E\left\{\sum_{i=1}^{n} \pi_{i2}^{-1} U_i(\theta) E\left(\delta_i \mid \mathbf{x}_i, y_i\right)\right\} = E\left\{\sum_{i=1}^{n} U_i(\theta)\right\}. \end{aligned}$$

Thus, we only have to postulate a model for the conditional response probability $\pi_{i2}$, conditional on $\mathbf{x}_i$ and $y_i$. Furthermore, $\hat{\theta}_{PS2}$ from (3.1) is more efficient than $\hat{\theta}_{PS1}$ using the true response probability, because $\hat{\theta}_{PS2}$ is essentially the conditional expectation of $\hat{\theta}_{PS1}$ given the observation. See Lemma 1 of Kim and Skinner (2013). Thus, we consider estimating $\theta$ by $\hat{\theta}_{PS2}$ in (3.1).

To compute the PS estimator, $\hat{\theta}_{PS2}$, we assume that $\mathbf{x}$ can be decomposed as $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ and the dimension of $\mathbf{x}_2$ is greater than or equal to one. We consider a parametric model for the response indicator as follows:

$$P(\delta = 1 \mid \mathbf{x}, y) = P(\delta = 1 \mid \mathbf{x}_1, y) = \pi(\mathbf{x}_1, y; \boldsymbol{\phi}_0) \tag{3.2}$$

for some function $\pi(\cdot)$ known up to the response model parameter $\boldsymbol{\phi}_0$. The assumed response model (3.2) implies that the response mechanism is nonignorable in the sense that the response mechanism is not independent of the study variable $y$ even after adjusting for the auxiliary variable $\mathbf{x}$. The model (3.2) also implies that response indicator does not depend on $\mathbf{x}_2$ given $\mathbf{x}_1$ and $y$. Variable $\mathbf{x}_2$ is sometimes the called nonresponse instrumental variable, which means that it does not directly related with the response mechanism but it helps to identify the parameters in the response mechanism. The assumption of conditional independence of the response indicator $\delta$ and the instrumental variable $\mathbf{x}_2$ in (3.2) makes the parameter $\boldsymbol{\phi}_0$ in (3.2) identifiable. See Wang et al. (2013) for details.

Under the parametric assumption (3.2), we now assume that $\delta_i$ are generated from a Bernoulli distribution with probability $\pi_i(\boldsymbol{\phi}_0) \equiv \pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi}_0)$ for some $\boldsymbol{\phi}_0$. If $y_i$ were observed throughout the sample, the likelihood function of $\boldsymbol{\phi}$ would be

$$L(\boldsymbol{\phi}) = \prod_{i=1}^{n} \{\pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})\}^{\delta_i} \{1 - \pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})\}^{(1-\delta_i)},$$

and the maximum likelihood estimator (MLE) of $\boldsymbol{\phi}$ could be obtained by solving the score equation $\mathbf{S}(\boldsymbol{\phi}) = \partial \log L(\boldsymbol{\phi})/\partial \boldsymbol{\phi} = \mathbf{0}$. The score equation $\mathbf{S}(\boldsymbol{\phi}) = \mathbf{0}$ can be expressed as

$$\mathbf{S}(\boldsymbol{\phi}) = \sum_{i=1}^{n} \mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_i) = \sum_{i=1}^{n} \{\delta_i - \pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})\} \, \mathbf{z}(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi}) = \mathbf{0}, \tag{3.3}$$

where $\mathbf{z}(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi}) = \partial \text{logit}\{\pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})\}/\partial \boldsymbol{\phi}$, and $\text{logit}(p) = \log\{p/(1-p)\}$. However, as some of $y_i$ are missing, the score equation (3.3) is not applicable. Instead, we can consider maximizing the observed likelihood function

$$L_{obs}(\boldsymbol{\phi}) = \prod_{i=1}^{n} \{\pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})\}^{\delta_i} \left[ \int \{1 - \pi(\mathbf{x}_{1i}, y; \boldsymbol{\phi})\} f(y \mid \mathbf{x}_i) dy \right]^{1-\delta_i}, \tag{3.4}$$

where $f(y \mid \mathbf{x})$ is the true conditional distribution of $y$ given $\mathbf{x}$.

The MLE of $\boldsymbol{\phi}$ can be obtained by solving the observed score equation, $\mathbf{S}_{obs}(\boldsymbol{\phi}) = \partial \log L_{obs}(\boldsymbol{\phi})/\partial \boldsymbol{\phi} = \mathbf{0}$. Finding the solution to the observed score equation is computationally challenging, because the observed likelihood involves integration with unknown parameters. Instead of solving the observed score equation, another way to find the MLE of $\boldsymbol{\phi}$ is to solve the mean score equation $\bar{\mathbf{S}}(\boldsymbol{\phi}) = \mathbf{0}$, where

$$\begin{aligned}
\bar{\mathbf{S}}(\boldsymbol{\phi}) = \sum_{i=1}^{n} \bar{\mathbf{s}}_i(\boldsymbol{\phi}) &= \sum_{i=1}^{n} E\{\mathbf{s}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, y) \mid \delta_i, \mathbf{x}_i, y_{obs,i}\} \\
&= \sum_{i=1}^{n} \left[\delta_i \mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_i) + (1-\delta_i) E_0\{\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, Y) \mid \mathbf{x}_i\}\right], \quad (3.5)
\end{aligned}$$

where $\mathbf{s}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, y)$ is defined in (3.3), $y_{obs,i}$ is $y_i$ if $\delta_i = 1$ and is null if $\delta_i = 0$, and $E_0(\cdot \mid \mathbf{x}_i) = E(\cdot \mid \mathbf{x}_i, \delta_i = 0)$. The mean score function (3.5) is the conditional expectation of the score function given all the observed data $(\delta_i, \mathbf{x}_i, y_{obs,i})$, for $i = 1, 2, \cdots, n$. The equivalence of $\mathbf{S}_{obs}(\boldsymbol{\phi})$ and $\bar{\mathbf{S}}(\boldsymbol{\phi})$ is given in Lemma 3.1, which was originally discussed in Louis (1982).

**Lemma 3.1.** *Under some regularity conditions,*

$$\mathbf{S}_{obs}(\boldsymbol{\phi}) = \bar{\mathbf{S}}(\boldsymbol{\phi})$$

*holds, where* $\mathbf{S}_{obs}(\boldsymbol{\phi}) = \partial \log L_{obs}(\boldsymbol{\phi}) / \partial \boldsymbol{\phi}$, $L_{obs}(\boldsymbol{\phi})$ *is defined in (3.4), and* $\bar{\mathbf{S}}(\boldsymbol{\phi})$ *is defined in (3.5).*

The proof of Lemma 3.1 is provided in Appendix A.1.

In order to solve the mean score equation $\bar{\mathbf{S}}(\boldsymbol{\phi}) = \mathbf{0}$, the conditional distribution for nonrespondents, $f_0(y \mid \mathbf{x}) = f(y \mid \mathbf{x}, \delta = 0)$, is needed to compute the conditional expectation of the score function of nonrespondents. If a parametric model $f(y \mid \mathbf{x}) = f(y \mid \mathbf{x}; \boldsymbol{\beta})$ is assumed in addition to (3.2), the conditional expectation for nonrespondents can be derived using two models, $f(y \mid \mathbf{x}; \boldsymbol{\beta})$ and $P(\delta = 1 \mid \mathbf{x}, y; \boldsymbol{\phi})$, and the maximum likelihood estimate for $(\boldsymbol{\beta}, \boldsymbol{\phi})$ can be computed jointly, as considered by Greenlees et al. (1982), Baker and Laird (1988), and Ibrahim et al. (1999). The fully parametric approach finds the maximum likelihood estimator that maximizes the following likelihood function

$$L(\boldsymbol{\beta}, \boldsymbol{\phi}) = \prod_{i=1}^{n} \left\{\pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi}) f(y \mid \mathbf{x}; \boldsymbol{\beta})\right\}^{\delta_i} \left[\int \left\{1 - \pi(\mathbf{x}_{1i}, y; \boldsymbol{\phi})\right\} f(y \mid \mathbf{x}_i; \boldsymbol{\beta}) dy\right]^{1-\delta_i}.$$

However, the fully parametric model approach, which assumes parametric models for both the response mechanism and the conditional distribution $f(y \mid \mathbf{x})$, is known to be sensitive to the failure of model assumptions [Kenward and Molenberghs (1988)]. Also, it can be challenging to check both models under nonignorable nonresponse. We will consider an alternative approach of computing the MLE of $\boldsymbol{\phi}$ in Section 3.3.

Instead of maximum likelihood estimation for the response model parameter $\boldsymbol{\phi}$, one can find a consistent estimator of $\boldsymbol{\phi}$ by forcing the PS estimator of auxiliary variables to match the complete sample mean of auxiliary variables as follows:

$$\sum_{i=1}^{n} \frac{\delta_i}{\pi_i(\boldsymbol{\phi})} \mathbf{x}_i = \sum_{i=1}^{n} \mathbf{x}_i. \tag{3.6}$$

Condition (3.6) is often called the calibration condition in survey sampling. Chang and Kott (2008) showed the consistency of the PS estimator satisfying the calibration condition (3.6) under some regularity conditions when the parametric response model (3.2) is correctly specified and there exists a linear relationship between the auxiliary variables $\mathbf{X}$ and the study variable $Y$. Wang et al. (2013) also proved the asymptotic normality of the PS estimator satisfying (3.6) without assuming the linear models. The proposed method in Chang and Kott (2008) and Wang et al. (2013), which is based on the generalized method of moments (GMM), find estimates by minimizing $\{\mathbf{A}(\theta; \boldsymbol{\phi})\}^T \mathbf{W}^{-1} \mathbf{A}(\theta; \boldsymbol{\phi})$, where

$$\mathbf{A}(\theta; \boldsymbol{\phi}) = \begin{bmatrix} \sum_{i=1}^{n} \{\delta_i \pi_i^{-1}(\boldsymbol{\phi}) \mathbf{x}_i - \mathbf{x}_i\} \\ \sum_{i=1}^{n} \{\delta_i \pi_i^{-1}(\boldsymbol{\phi}) y_i - \theta\} \end{bmatrix},$$

$\mathbf{x}_i$ is $i$-th observation of benchmark covariates and $\mathbf{W}^{-1}$ is some weight matrix. By this method, calibration can correct the nonresponse bias. Here, use of $\mathbf{x}_i$ in (3.6) makes the parameter $\boldsymbol{\phi}$ estimable under some regularity condition discussed in Wang et al. (2013).

## 3.3 Proposed method

In this section, we consider an alternative approach of obtaining the maximum likelihood estimator of response model parameter $\boldsymbol{\phi}$ without specifying the conditional distribution of $f(y \mid \mathbf{x})$. Note that with the fully parametric approach, the conditional expectation in (3.5) is taken with respect to the conditional distribution,

$$f(y \mid \mathbf{x}, \delta = 0) \quad = \quad f(y \mid \mathbf{x}) \frac{P(\delta = 0 \mid \mathbf{x}, y)}{E\{P(\delta = 0 \mid \mathbf{x}, Y) \mid \mathbf{x}\}}.$$

If a parametric model for $f(y \mid \mathbf{x})$ were correctly specified, the MLE of the response model parameter $\boldsymbol{\phi}$ could be obtained by solving $\bar{\mathbf{S}}(\boldsymbol{\phi}) = \mathbf{0}$, where $\bar{\mathbf{S}}(\boldsymbol{\phi})$ is defined in (3.5), and the resulting MLE will be consistent and the most efficient in the sense that it achieves the Cramer-Rao lower bound. On the other hand, these attractive features are not guaranteed when the parametric model for $f(y \mid \mathbf{x})$ is not correctly specified. Besides, finding a correct model is quite challenging when only a part of $y$ is observed.

We consider an alternative approach that uses a model for the conditional distribution of the study variable $y$ given the auxiliary variables $\mathbf{x}$ for respondents, denoted by $f_1(y \mid \mathbf{x}) = f(y \mid \mathbf{x}, \delta = 1)$, instead of using a model for the conditional distribution of $y$ given $\mathbf{x}$, $f(y \mid \mathbf{x})$. To obtain the conditional distribution of $y$ given $\mathbf{x}$ for nonrespondents, denoted by $f_0(y \mid \mathbf{x}) = f(y \mid \mathbf{x}, \delta = 0)$, from the conditional distribution for respondents, $f_1(y \mid \mathbf{x})$, the following Bayes formula can be used

$$f_0(y \mid \mathbf{x}) \quad = \quad f_1(y \mid \mathbf{x}) \frac{O(\mathbf{x}, y)}{E\{O(\mathbf{x}, Y) \mid \mathbf{x}, \delta = 1\}}, \tag{3.7}$$

where $O(\mathbf{x}, y) = P(\delta = 0 \mid \mathbf{x}, y)/P(\delta = 1 \mid \mathbf{x}, y)$ is the conditional odds of nonresponse. In (3.7), we only need the response model (3.2) and the conditional distribution of study variable given the auxiliary variables for respondents $f_1(y \mid \mathbf{x})$, which is relatively easy to verify from the observed part of the sample. Using (3.7), the conditional expectation in the mean score function (3.5) can be computed by

$$\sum_{i=1}^{n} \left[ \delta_i \mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_i) + (1 - \delta_i) \frac{\int \mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y) O(\mathbf{x}_i, y) f_1(y \mid \mathbf{x}_i) dy}{\int O(\mathbf{x}_i, y) f_1(y \mid \mathbf{x}_i) dy} \right].$$

If the response model follows from the following logistic model

$$P(\delta = 1|\mathbf{X} = \mathbf{x}, Y = y) = \frac{\exp(\phi_0 + \boldsymbol{\phi}_1\mathbf{x}_1 + \phi_2 y)}{1 + \exp(\phi_0 + \boldsymbol{\phi}_1\mathbf{x}_1 + \phi_2 y)},$$

then $O(\mathbf{x}, y) = \exp(-\phi_0 - \boldsymbol{\phi}_1\mathbf{x}_1 - \phi_2 y)$, and (3.7) becomes

$$f_0(y|\mathbf{x}) = f_1(y|\mathbf{x})\frac{\exp(-\phi_2 y)}{E\{\exp(-\phi_2 y)|\mathbf{x}, \delta = 1\}},$$

which is often called the exponential tilting model [Kim and Yu (2011)].

We assume a parametric model for the conditional distribution for respondents $f_1(y \mid \mathbf{x})$. The parametric model can be expressed as

$$f_1(y \mid \mathbf{x}) = f_1(y \mid \mathbf{x}; \boldsymbol{\gamma}_0), \tag{3.8}$$

for some $\boldsymbol{\gamma}_0$. Thus, the two parametric model assumptions, (3.2) and (3.8), are used to compute the conditional expectation in (3.5). The approach using the respondents' model in (3.8) is weaker than the fully parametric model approach that assumes parametric models for $f(y \mid \mathbf{x})$. In addition, model diagnostics for $f_1(y \mid \mathbf{x})$ are more feasible than those for $f(y \mid \mathbf{x})$, since $(\mathbf{x}_i, y_i)$ are only observed for respondents. The two assumed parametric models are combined to obtain the nonrespondents' conditional distribution $f_0(y \mid \mathbf{x})$ by (3.7).

To estimate response model parameter $\boldsymbol{\phi}$ in (3.2) using the mean score equation (3.5), we first need to obtain a consistent estimator of $\boldsymbol{\gamma}_0$ in (3.8), which can be obtained by solving the following score equation for $\boldsymbol{\gamma}$ :

$$\mathbf{S}_1(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \delta_i \mathbf{s}_{1i}(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \delta_i \frac{\partial \log f_1(y_i \mid \mathbf{x}_i; \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \mathbf{0}. \tag{3.9}$$

Since the conditional distribution for respondents, $f_1(y \mid \mathbf{x})$, only involves respondents, we can obtain the MLE of $\boldsymbol{\gamma}$ from the respondents. Using the MLE $\hat{\boldsymbol{\gamma}}$ from (3.9), the mean score function in (3.5) can be obtained by substituting $E_0\{\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, Y) \mid \mathbf{x}_i\}$ with

$$E_0\{\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, Y) \mid \mathbf{x}_i, \hat{\boldsymbol{\gamma}}\} = \frac{\int \mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y) O(\mathbf{x}_{1i}, y; \boldsymbol{\phi}) f_1(y \mid \mathbf{x}_i, \hat{\boldsymbol{\gamma}}) dy}{\int O(\mathbf{x}_{1i}, y; \boldsymbol{\phi}) f_1(y \mid \mathbf{x}_i, \hat{\boldsymbol{\gamma}}) dy}, \tag{3.10}$$

where $O(\mathbf{x}_{1i}, y; \boldsymbol{\phi}) = \pi^{-1}(\boldsymbol{\phi}; \mathbf{x}_i, y) - 1$.

However, computing the conditional expectation in (3.10) involves integration, which can be computationally challenging. To avoid this difficulty, we propose using

$$\tilde{E}_0\{\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, Y) \mid \mathbf{x}_i; \boldsymbol{\phi}, \hat{\boldsymbol{\gamma}}\} = \frac{\sum_{j;\delta_j=1} \mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_j) O(\mathbf{x}_{1i}, y_j; \boldsymbol{\phi}) f_1(y_j \mid \mathbf{x}_i; \hat{\boldsymbol{\gamma}}) / \hat{f}_1(y_j)}{\sum_{k;\delta_k=1} O(\mathbf{x}_{1i}, y_k; \boldsymbol{\phi}) f_1(y_k \mid \mathbf{x}_i; \hat{\boldsymbol{\gamma}}) / \hat{f}_1(y_k)}, \tag{3.11}$$

where $\hat{f}_1(y_j) = n_r^{-1} \sum_{k:\delta_k=1} f_1(y_j|\mathbf{x}_k; \hat{\boldsymbol{\gamma}})$ is a consistent estimator of the marginal density $f_1(y) = f(y \mid \delta = 1)$ among the respondents, evaluated at $y = y_j$ and $n_r$ is the number of respondents in the sample.

To justify (3.11), we use

$$\begin{aligned} E_1\{Q(\mathbf{x}_i, Y)|\mathbf{x}_i\} &= \int Q(\mathbf{x}_i, y) f_1(y|\mathbf{x}_i) dy \\ &= \int Q(\mathbf{x}_i, y) \frac{f_1(y|\mathbf{x}_i)}{f_1(y)} f_1(y) dy, \end{aligned}$$

which, using the idea of importance sampling, can be estimated by

$$\tilde{E}_1\{Q(\mathbf{x}_i, Y)|\mathbf{x}_i\} = \frac{\sum_{j;\delta_j=1} \mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_j) O(\mathbf{x}_{1i}, y_j) f_1(y_j \mid \mathbf{x}_i) / \hat{f}_1(y_j)}{\sum_{k;\delta_k=1} O(\mathbf{x}_{1i}, y_k) f_1(y_k \mid \mathbf{x}_i) / \hat{f}_1(y_k)}.$$

Since

$$f\,(y \mid \delta = 1) = \int f_1(y \mid \mathbf{x}) f(\mathbf{x} \mid \delta = 1) d\mathbf{x},$$

we can use the empirical distribution of $f(\mathbf{x}|\delta = 1)$ to obtain

$$\hat{f}_1(y_j) \propto \sum_{k;\delta_k=1} f_1(y_j \mid \mathbf{x}_k; \hat{\boldsymbol{\gamma}}). \tag{3.12}$$

Thus, the mean score equation (3.5) for $\boldsymbol{\phi}$ can be approximated by

$$\mathbf{S}_2(\boldsymbol{\phi}; \hat{\boldsymbol{\gamma}}) = \sum_{i=1}^{n} \left[\delta_i \mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_i) + (1 - \delta_i) \tilde{E}_0\{\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, Y) \mid \mathbf{x}_i; \boldsymbol{\phi}, \hat{\boldsymbol{\gamma}}\}\right] = \mathbf{0}, \tag{3.13}$$

where $\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_i, y_i)$ is defined in (3.3), $\tilde{E}_0\{\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, Y) \mid \mathbf{x}_i; \boldsymbol{\phi}, \hat{\boldsymbol{\gamma}}\}$ is defined in (3.11) with $\hat{f}(y_j \mid \delta_j = 1)$ in (3.12).

Once the mean score equation is computed by (3.13), the solution to the mean score equation can be obtained by the following EM algorithm:

$$\hat{\boldsymbol{\phi}}^{(t+1)} \leftarrow \text{ solve } \bar{\mathbf{S}}(\boldsymbol{\phi} \mid \hat{\boldsymbol{\phi}}^{(t)}, \hat{\boldsymbol{\gamma}}) = \mathbf{0}$$

where

$$\bar{\mathbf{S}}(\boldsymbol{\phi} \mid \hat{\boldsymbol{\phi}}^{(t)}, \hat{\boldsymbol{\gamma}}) = \sum_{i=1}^{n} \left[ \delta_i \mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_i) + (1-\delta_i) \tilde{E}_0\{\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, Y) \mid \mathbf{x}_i; \hat{\boldsymbol{\phi}}^{(t)}, \hat{\boldsymbol{\gamma}}\} \right],$$

$\tilde{E}_0\{\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, Y) \mid \mathbf{x}_i; \hat{\boldsymbol{\phi}}^{(t)}, \hat{\boldsymbol{\gamma}}\} = \sum_{j;\delta_j=1} w_{ij}^*(\hat{\boldsymbol{\phi}}^{(t)}, \hat{\boldsymbol{\gamma}}) \mathbf{s}(\boldsymbol{\phi}, \delta_i, \mathbf{x}_{1i}, y_j)$,

$$w_{ij}^*(\boldsymbol{\phi}, \boldsymbol{\gamma}) = \frac{O(\mathbf{x}_{1i}, y_j; \boldsymbol{\phi}) f_1(y_j \mid \mathbf{x}_i; \boldsymbol{\gamma}) / C(y_j; \boldsymbol{\gamma})}{\sum_{k;\delta_k=1} O(\mathbf{x}_{1i}, y_k; \boldsymbol{\phi}) f_1(y_k \mid \mathbf{x}_i; \boldsymbol{\gamma}) / C(y_k; \boldsymbol{\gamma})}, \quad (3.14)$$

$O(\mathbf{x}_1, y; \boldsymbol{\phi})$ is defined in (3.11), and $C(y; \boldsymbol{\gamma}) = \sum_{l;\delta_l=1} f_1(y \mid \mathbf{x}_l; \boldsymbol{\gamma})$. The weight $w_{ij}^*$ in (3.14) can be viewed as a fractional weight assigned to the imputed values, where $y_i^* = y_j$ is the imputed value [Kim (2011)].

Once the solution $\hat{\boldsymbol{\phi}}_p$ to (3.13) is obtained, the parameter of interest $\theta$ is estimated by solving

$$U_{PS}(\theta; \hat{\boldsymbol{\phi}}_p) = \sum_{i=1}^{n} \frac{\delta_i}{\pi(\mathbf{x}_i, y_i; \hat{\boldsymbol{\phi}}_p)} u(\theta; \mathbf{x}_i, y_i) = 0. \quad (3.15)$$

The following theorem presents some asymptotic properties of the response model parameter estimator $\hat{\boldsymbol{\phi}}_p$ and PS estimator, $\hat{\theta}_{PS,p}$, satisfying (3.15), using the fact that the response model parameter estimator $\hat{\boldsymbol{\phi}}_p$ is the solution to (3.13), and the PS estimator is the solution to (3.15).

**Theorem 3.1.** *Assume the regularity conditions stated in Appendix A hold. The response model estimator $\hat{\boldsymbol{\phi}}_p$ satisfies*

$$\sqrt{n}\left(\hat{\boldsymbol{\phi}}_p - \boldsymbol{\phi}_0\right) \xrightarrow{\mathcal{L}} N(0, \boldsymbol{\Sigma}_\phi) \quad (3.16)$$

*where*

$$\boldsymbol{\Sigma}_\phi = \mathcal{I}_{22}^{-1} V\left[\mathbf{s}(\boldsymbol{\phi}_0; \boldsymbol{\gamma}_0) - \boldsymbol{\kappa}\mathbf{s}_1(\boldsymbol{\gamma}_0)\right] (\mathcal{I}_{22}^{-1})^T,$$

$\boldsymbol{\kappa} = \mathcal{I}_{21}\mathcal{I}_{11}^{-1}$,

$$\begin{aligned}
\mathcal{I}_{11} &= V\left\{\mathbf{s}_1(\boldsymbol{\gamma}_0)\right\}, \\
\mathcal{I}_{21} &= E\left[(1-\delta)\{\mathbf{s}(\boldsymbol{\phi}_0) - \bar{\mathbf{s}}_0(\boldsymbol{\phi}_0;\boldsymbol{\gamma}_0)\}\mathbf{s}_1^T(\boldsymbol{\gamma}_0)\right], \\
\mathcal{I}_{22} &= E\{(1-\delta)\bar{\mathbf{s}}_0(\boldsymbol{\phi}_0;\boldsymbol{\gamma}_0)\mathbf{s}^T(\boldsymbol{\phi}_0)/\pi(\boldsymbol{\phi})\},
\end{aligned}$$

$\bar{\mathbf{s}}_0(\boldsymbol{\phi};\boldsymbol{\gamma}) = E_0\{\mathbf{s}(\boldsymbol{\phi};\delta,\mathbf{x},Y)|\mathbf{x};\boldsymbol{\gamma}\}$, $\mathbf{s}_2(\boldsymbol{\phi};\boldsymbol{\gamma}) = \delta\mathbf{s}(\boldsymbol{\phi};\delta,\mathbf{x},y) + (1-\delta)\bar{\mathbf{s}}_0(\boldsymbol{\phi};\boldsymbol{\gamma})$, $\mathbf{s}_1(\boldsymbol{\gamma}) = (\partial/\partial\boldsymbol{\gamma})\log f_1(y|\mathbf{x};\boldsymbol{\gamma})$, *and* $\mathbf{s}(\boldsymbol{\phi}) = \mathbf{s}(\boldsymbol{\phi};\delta,\mathbf{x},y)$ *is defined in (3.3). Also, the PS estimator* $\hat{\theta}_{PS,p}$ *satisfies*

$$\sqrt{n}\left(\hat{\theta}_{PS,p} - \theta\right) \rightarrow N(0,\sigma_\theta^2) \tag{3.17}$$

*where*

$$\sigma_\theta^2 = \boldsymbol{\tau}^{-1}V\left[u_{PS}(\theta_0;\boldsymbol{\phi}_0) - \mathbf{B}\left\{\mathbf{s}(\boldsymbol{\phi}_0;\boldsymbol{\gamma}_0) - \boldsymbol{\kappa}\mathbf{s}_1(\boldsymbol{\gamma}_0)\right\}\right](\boldsymbol{\tau}^{-1})^T,$$

$u_{PS}(\theta;\boldsymbol{\phi}) = \delta\pi^{-1}(\boldsymbol{\phi})u(\theta)$, $\boldsymbol{\tau} = E\{\partial u(\theta;\boldsymbol{\phi}_0)/\partial\theta^T\}$, $\mathbf{B} = Cov\{u_{PS}(\theta;\boldsymbol{\phi}_0),\mathbf{s}(\boldsymbol{\phi}_0)\}\mathcal{I}_{22}^{-1}$, *and* $\boldsymbol{\kappa}$ *is defined in (3.16).*

Note that $\mathcal{I}_{21} = \mathbf{0}$ and $\mathcal{I}_{22} = V\{\mathbf{S}(\boldsymbol{\phi})\}$ under missing at random assumptions, since the study variable $y$ is not involved in the response model given $\mathbf{x}$.

The proof of Theorem 3.1 is provided in Appendix A.2. Theorem 3.1 shows that the response model parameter estimator $\hat{\boldsymbol{\phi}}_p$ and the PS estimator $\hat{\theta}_{PS,p}$ are consistent estimators for the corresponding parameters and have asymptotic normal distributions.

## 3.4 Variance estimation

We consider two ways of estimating the variance of the proposed estimator: one is by linearization using Theorem 3.1 in Section 3.3, and the other is by the jackknife method.

Firstly, the linearization variance estimator will be established using Theorem 3.1. By Theorem 3.1, the variance of the PS estimator, $\hat{\theta}_{PS,p}$, can be estimated by

$$\hat{V}_{lin}(\hat{\theta}_{PS,p}) = \frac{1}{n}\hat{\boldsymbol{\tau}}^{-1}\hat{V}_{Ul}(\hat{\boldsymbol{\tau}}^{-1})^T \tag{3.18}$$

where $\hat{\boldsymbol{\tau}} = n^{-1}\sum_{i=1}^{n}\delta_i\pi^{-1}(\mathbf{x}_{1i}, y_i; \hat{\boldsymbol{\phi}})\dot{u}(\hat{\theta}_{PS,p}; \mathbf{x}_i, y_i)$, $\dot{u}(\theta; \mathbf{x}, y) = \partial u(\theta; \mathbf{x}, y)/\partial\theta)^T$,

$$\hat{V}_{Ul} \;=\; (n-1)^{-1}\sum_{i=1}^{n}\left(\hat{u}_{li} - \bar{u}_n\right)^2,$$

$\hat{u}_{li} = u_{li}(\hat{\theta}_{PS,p}, \hat{\boldsymbol{\phi}}_p, \hat{\boldsymbol{\gamma}})$, $\bar{u}_n = n^{-1}\sum_{i=1}^{n}\hat{u}_{li}$,

$$\begin{aligned}
u_{li}(\theta, \boldsymbol{\phi}, \boldsymbol{\gamma}) &= -\hat{\mathbf{B}}\bar{\mathbf{s}}_0^*(\boldsymbol{\phi}; \mathbf{x}_i, \boldsymbol{\phi}, \boldsymbol{\gamma}) \\
&\quad + \delta_i\left[\frac{u(\theta; \mathbf{x}_i, y_i)}{\pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})} - \hat{\mathbf{B}}\{\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_i) - \bar{\mathbf{s}}_0^*(\boldsymbol{\phi}; \mathbf{x}_i, \boldsymbol{\gamma}, \boldsymbol{\phi}) - \hat{\boldsymbol{\kappa}}\mathbf{s}_1(\boldsymbol{\gamma}; \mathbf{x}_i, y_i)\}\right], \\
\hat{\mathbf{B}} &= \left\{\sum_{i=1}^{n}\delta_i\pi^{-1}(\mathbf{x}_{1i}, y_i; \hat{\boldsymbol{\phi}}_p)u(\hat{\theta}_{PS,p}; \mathbf{x}_i, y_i)\mathbf{s}^T(\hat{\boldsymbol{\phi}}_p; \delta_i, \mathbf{x}_{1i}, y_i)\right\}\hat{\mathbf{I}}_{22}^{-1}(\hat{\boldsymbol{\phi}}_p, \hat{\boldsymbol{\gamma}}), \\
\hat{\boldsymbol{\kappa}} &= \hat{\mathbf{I}}_{21}(\hat{\boldsymbol{\phi}}_p, \hat{\boldsymbol{\gamma}})\hat{\mathbf{I}}_{11}^{-1}(\hat{\boldsymbol{\phi}}_p, \hat{\boldsymbol{\gamma}}), \\
\hat{\mathbf{I}}_{22}(\boldsymbol{\phi}, \boldsymbol{\gamma}) &= \sum_{i=1}^{n}(1-\delta_i)\bar{\mathbf{s}}_0^*(\boldsymbol{\phi}; \mathbf{x}_i, \boldsymbol{\phi}, \boldsymbol{\gamma})\sum_{j;\delta_j=1}w_{ij}^*(\boldsymbol{\phi}; \boldsymbol{\gamma})\mathbf{s}^T(\boldsymbol{\phi}; \mathbf{x}_{1i}, y_j)/\pi(\boldsymbol{\phi}; \mathbf{x}_{1i}, y_j), \\
\hat{\mathbf{I}}_{21}(\boldsymbol{\phi}, \boldsymbol{\gamma}) &= \sum_{i=1}^{n}(1-\delta_i)\sum_{j;\delta_j=1}w_{ij}^*(\boldsymbol{\phi}; \boldsymbol{\gamma})\{\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_j) - \bar{\mathbf{s}}_0^*(\boldsymbol{\phi}; \mathbf{x}_i, \boldsymbol{\phi}, \boldsymbol{\gamma})\}\mathbf{s}_1^T(\boldsymbol{\gamma}; \mathbf{x}_i, y_j), \\
\hat{\mathbf{I}}_{11}(\boldsymbol{\phi}, \boldsymbol{\gamma}) &= -\sum_{i=1}^{n}\delta_i\dot{\mathbf{s}}_1(\boldsymbol{\gamma}; \mathbf{x}_i, y_i),
\end{aligned}$$

$w_{ij}^*(\boldsymbol{\phi}, \boldsymbol{\gamma})$ is defined in (3.14), $\bar{\mathbf{s}}_0^*(\boldsymbol{\phi}; \mathbf{x}_i, \boldsymbol{\phi}, \boldsymbol{\gamma}) = \sum_{j;\delta_j=1}w_{ij}^*(\boldsymbol{\phi}, \boldsymbol{\gamma})\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_j)$, $\dot{\mathbf{s}}(\boldsymbol{\phi}; \delta, \mathbf{x}, y) = \partial\mathbf{s}(\boldsymbol{\phi}; \delta, \mathbf{x}, y)/\partial\boldsymbol{\phi}^T$, $\dot{\mathbf{s}}_1(\boldsymbol{\gamma}; \mathbf{x}, y) = \partial\mathbf{s}_1(\boldsymbol{\gamma}; \mathbf{x}, y)/\partial\boldsymbol{\gamma}^T$, and $C(y; \boldsymbol{\gamma}) = \sum_{l;\delta_l=1}f_1(y \mid \mathbf{x}_l; \boldsymbol{\gamma})$.

Another way to estimate the variance of the PS estimator is to use the jackknife method. Let $w_i^{(k)}$ be the $k$-th replicate weight under simple random sampling, which is defined by

$$w_i^{(k)} = \begin{cases} (n-1)^{-1} & \text{if} \quad i \neq k \\ 0 & \text{if} \quad i = k. \end{cases}$$

First, the $k$-th jackknife replicate of $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\gamma}}^{(k)}$ is obtained by solving $\mathbf{S}_1^{(k)}(\boldsymbol{\gamma}) = \mathbf{0}$, where $\mathbf{S}_1^{(k)}(\boldsymbol{\gamma}) = \sum_{i=1}^{n}w_i^{(k)}\delta_i\mathbf{S}_1(\boldsymbol{\gamma}; \mathbf{x}_i, y_i)$. Next, the $k$-th jackknife replicate of $\hat{\boldsymbol{\phi}}_p$ can be computed by solving $\mathbf{S}_2^{(k)}(\boldsymbol{\phi}; \hat{\boldsymbol{\gamma}}^{(k)}) = \mathbf{0}$, where

$$\mathbf{S}_2^{(k)}(\boldsymbol{\phi}; \hat{\boldsymbol{\gamma}}^{(k)}) = \sum_{i=1}^{n}w_i^{(k)}\left\{\delta_i\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_i) + (1-\delta_i)\sum_{j;\delta_j=1}w_{ij}^{*(k)}(\boldsymbol{\phi}, \hat{\boldsymbol{\gamma}}^{(k)})\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_j)\right\}, \tag{3.19}$$

$w_i^{(k)}$ is $k$-th jackknife replicate weight,

$$w_{ij}^{*(k)}(\boldsymbol{\phi}, \boldsymbol{\gamma}) = \frac{w_i^{(k)} O(\mathbf{x}_{1i}, y_j; \boldsymbol{\phi}) f_1(y_j \mid \mathbf{x}_i; \boldsymbol{\gamma}) / C^{(k)}(y_j; \boldsymbol{\gamma})}{\sum_{l;\delta_l=1} w_i^{(k)} O(\mathbf{x}_{1i}, y_l; \boldsymbol{\phi}) f_1(y_l \mid \mathbf{x}_i; \boldsymbol{\gamma}) / C^{(k)}(y_l; \boldsymbol{\gamma})}, \tag{3.20}$$

and $C^{(k)}(y; \boldsymbol{\gamma}) = \sum_{l:\delta_l=1} w_l^{(k)} f_1(y \mid \mathbf{x}_l; \boldsymbol{\gamma})$.

**Remark 3.1.** *Solving $\mathbf{S}_2^{(k)}(\boldsymbol{\phi}) = \mathbf{0}$ using the EM algorithm may involve heavy computation. We can avoid this issue by approximating the solution to (3.19) by a one-step method as follows*

$$\hat{\boldsymbol{\phi}}_p^{(k)} = \hat{\boldsymbol{\phi}}_p - \left\{ \frac{\partial \mathbf{S}_2^{(k)}(\hat{\boldsymbol{\phi}}_p)}{\partial \boldsymbol{\phi}} \right\}^{-1} \mathbf{S}_2^{(k)}(\hat{\boldsymbol{\phi}}_p; \hat{\boldsymbol{\gamma}}^{(k)}),$$

*where*

$$\frac{\partial \mathbf{S}_2^{(k)}(\boldsymbol{\phi}; \hat{\boldsymbol{\gamma}}^{(k)})}{\partial \boldsymbol{\phi}} = \sum_{i=1}^{n} w_i^{(k)} (1-\delta_i) \sum_{j;\delta_j=1} w_{ij}^{*(k)}(\boldsymbol{\phi}; \boldsymbol{\gamma}^{(k)}) \bar{\mathbf{s}}_0^{*(k)}(\boldsymbol{\phi}; \mathbf{x}_i) \mathbf{s}^T(\boldsymbol{\phi}; \mathbf{x}_{1i}, y_j) / \pi(\boldsymbol{\phi}; \mathbf{x}_{1i}, y_j),$$

$\bar{\mathbf{s}}_0^{*(k)}(\boldsymbol{\phi}; \mathbf{x}_i) = \sum_{j;\delta_j=1} w_{ij}^{*(k)}(\boldsymbol{\phi}; \hat{\boldsymbol{\gamma}}^{(k)}) \mathbf{s}(\boldsymbol{\phi}; \mathbf{x}_{1i}, y_j)$, *and $w_{ij}^{*(k)}(\boldsymbol{\phi}; \boldsymbol{\gamma})$ is defined in (3.20).*

Once we obtain $\hat{\boldsymbol{\gamma}}^{(k)}$ and $\hat{\boldsymbol{\phi}}^{(k)}$, the $k$-th jackknife replicate of the PS estimator of $\theta = E(Y)$ can be computed as

$$\hat{\theta}_{PS,p}^{(k)} = \frac{\sum_{i=1}^{n} w_i^{(k)} \delta_i \pi^{-1}(\mathbf{x}_{1i}, y_i; \hat{\boldsymbol{\phi}}^{(k)}) y_i}{\sum_{i=1}^{n} w_i^{(k)} \delta_i \pi^{-1}(\mathbf{x}_{1i}, y_i; \hat{\boldsymbol{\phi}}^{(k)})}, \tag{3.21}$$

and the variance of the PS estimator can be estimated by

$$\hat{V}_{JK}(\hat{\theta}_{PS,p}) = \frac{(n-1)}{n} \sum_{k=1}^{n} \left\{ \hat{\theta}_{PS,p}^{(k)} - \bar{\theta}_{PS,p}^{JK} \right\}^2, \tag{3.22}$$

where $\hat{\theta}_{PS,p}^{(k)}$ is defined in (3.21) and $\bar{\theta}_{PS,p}^{JK} = n^{-1} \sum_{k=1}^{n} \hat{\theta}_{PS,p}^{(k)}$.

## 3.5 Generalized Method of Moment Estimation

We now discuss incorporating the auxiliary variables into estimation. Note that the PS estimator discussed in Section 3.3 does not satisfy the calibration constraints in (3.6). Thus, the PS estimator can be adjusted to satisfy the calibration condition and improve the efficiency. The proposed method is based on the theory of Generalized Method of Moment (GMM) applied to PS estimation. Zhou and Kim (2012) provide the GMM theory for PS estimation under ignorable missing.

The PS estimator discussed in Section 3.3 can also be derived by minimizing the following objective function with respect to $(\boldsymbol{\eta}, \theta)$ :

$$Q(\boldsymbol{\eta}, \theta) = \begin{bmatrix} \mathbf{S}_*(\boldsymbol{\eta}) \\ U_{PS}(\theta; \boldsymbol{\phi}) \end{bmatrix}^T \left[ V \begin{Bmatrix} \mathbf{S}_*(\boldsymbol{\eta}) \\ U_{PS}(\theta; \boldsymbol{\phi}) \end{Bmatrix} \right]^{-1} \begin{bmatrix} \mathbf{S}_*(\boldsymbol{\eta}) \\ U_{PS}(\theta; \boldsymbol{\phi}) \end{bmatrix}, \tag{3.23}$$

where $\boldsymbol{\eta}^T = (\boldsymbol{\gamma}^T, \boldsymbol{\phi}^T)$,

$$\mathbf{S}_*(\boldsymbol{\eta}) = \begin{bmatrix} \mathbf{S}_1(\boldsymbol{\gamma}) \\ \mathbf{S}_2(\boldsymbol{\phi}; \boldsymbol{\gamma}) \end{bmatrix}, \tag{3.24}$$

$\mathbf{S}_1(\boldsymbol{\gamma})$ is defined in (3.9), $\mathbf{S}_2(\boldsymbol{\phi}; \boldsymbol{\gamma})$ is defined in (3.13), and $U_{PS}(\theta; \boldsymbol{\phi})$ is defined in (3.15). To justify (3.23) by partitioning the variance matrix as follows

$$\begin{aligned} V \begin{bmatrix} \mathbf{S}_*(\boldsymbol{\eta}) \\ U_{PS}(\theta; \boldsymbol{\phi}) \end{bmatrix} &= \begin{bmatrix} V\{\mathbf{S}_*(\boldsymbol{\eta})\} & C\{\mathbf{S}_*(\boldsymbol{\eta}), U_{PS}(\theta; \boldsymbol{\phi})\} \\ C\{U_{PS}(\theta; \boldsymbol{\phi}), \mathbf{S}_*(\boldsymbol{\eta})\} & V\{U_{PS}(\theta; \boldsymbol{\phi})\} \end{bmatrix} \\ &\triangleq \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}, \end{aligned}$$

we can write (3.23) as

$$\begin{bmatrix} \mathbf{S}_*(\boldsymbol{\eta}) \\ U_{PS}(\theta; \boldsymbol{\phi}) \end{bmatrix}^T \begin{pmatrix} \mathbf{V}_{11}^{-1} + \mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{V}_{2|1}^{-1}\mathbf{V}_{21}\mathbf{V}_{11}^{-1} & -\mathbf{V}_{2|1}^{-1}\mathbf{V}_{21}\mathbf{V}_{11}^{-1} \\ -\mathbf{V}_{11}^{-1}\mathbf{V}_{12}\mathbf{V}_{2|1}^{-1} & \mathbf{V}_{2|1}^{-1} \end{pmatrix} \begin{bmatrix} \mathbf{S}_*(\boldsymbol{\eta}) \\ U_{PS}(\theta; \boldsymbol{\phi}) \end{bmatrix},$$

where $\mathbf{V}_{2|1} = \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}$. Thus, the objective function in (3.23) can be simplified as

$$Q(\boldsymbol{\eta}, \theta) = Q_1(\boldsymbol{\eta}) + Q_2(\theta | \boldsymbol{\eta}), \tag{3.25}$$

where $Q_1(\boldsymbol{\eta}) = \mathbf{S}_*^T(\boldsymbol{\eta}) V_{11}^{-1} \mathbf{S}_*(\boldsymbol{\eta})$ and

$$Q_2(\theta | \boldsymbol{\eta}) = \left\{ U_{PS}(\theta; \boldsymbol{\phi}) - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{S}_*(\boldsymbol{\eta}) \right\}^T \mathbf{V}_{2|1}^{-1} \left\{ U_{PS}(\theta; \boldsymbol{\phi}) - \mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{S}_*(\boldsymbol{\eta}) \right\}.$$

If $\boldsymbol{\eta} = \hat{\boldsymbol{\eta}}$, where $\hat{\boldsymbol{\eta}}^T = (\hat{\boldsymbol{\gamma}}^T, \hat{\boldsymbol{\phi}}^T)$ and $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\phi}}$ are the maximum likelihood estimator of $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$, $Q_1(\boldsymbol{\eta})$ is zero and $Q(\boldsymbol{\eta}, \theta) = Q_2(\theta|\boldsymbol{\eta}) = U_{PS}^T(\theta; \boldsymbol{\phi})\mathbf{V}_{2|1}^{-1}U_{PS}(\theta; \boldsymbol{\phi})$. Also, $Q_2(\theta|\hat{\boldsymbol{\eta}})$ is zero at $\theta = \hat{\theta}_{PS}$, where $\hat{\theta}_{PS}$ is the solution to (3.15). Thus, the estimator $(\hat{\boldsymbol{\eta}}^T, \hat{\theta}_{PS})^T$ that was found in Section 3.3 also minimizes the objective function $Q(\boldsymbol{\eta}, \theta)$ in (3.25).

Here, we can expand $U_{PS}(\theta; \boldsymbol{\phi})$ to incorporate auxiliary information from $\mathbf{x}$. As mentioned in Section 3.2, if the study variable $y$ can be well approximated by a linear combination of $\mathbf{x}$, the calibration condition (3.6) can lead to an efficient PS estimator. However, in general, the PS estimator using the MLE of $\boldsymbol{\phi}$ applied to the auxiliary variable $\mathbf{x}$ does not satisfy the calibration condition (3.6). That is, the PS estimator of auxiliary variable $\mathbf{x}$ using the MLE of $\boldsymbol{\phi}$ is not equal to the complete sample mean $\bar{\mathbf{x}}_n = n^{-1}\sum_{i=1}^n \mathbf{x}_i$. Thus, incorporating the complete sample mean $\bar{\mathbf{x}}_n$ information into estimating the parameter of interest $\theta$ can be considered.

For simplicity, we consider $\boldsymbol{\theta} = (\boldsymbol{\mu}_{\mathbf{x}}^T, \mu_y)^T$, where $\boldsymbol{\mu}_{\mathbf{x}} = E(\mathbf{X})$ and $\mu_y = E(Y)$. Define

$$
\begin{aligned}
\mathbf{U}(\boldsymbol{\theta}) &= \begin{bmatrix} \mathbf{U}_1(\boldsymbol{\theta}) \\ \mathbf{U}_2(\theta) \end{bmatrix} = \sum_{i=1}^n \begin{bmatrix} \mathbf{U}_1(\boldsymbol{\mu}_{\mathbf{x}}; \mathbf{x}_i) \\ \mathbf{U}_2(\mu_y; y_i) \end{bmatrix} \\
\mathbf{U}_{PS}(\boldsymbol{\theta}; \boldsymbol{\phi}) &= \begin{bmatrix} \mathbf{U}_{PS,1}(\theta; \boldsymbol{\phi}) \\ \mathbf{U}_{PS,2}(\theta; \boldsymbol{\phi}) \end{bmatrix} = \sum_{i=1}^n \frac{\delta_i}{\pi_i(\boldsymbol{\phi})} \begin{bmatrix} \mathbf{U}_1(\boldsymbol{\mu}_{\mathbf{x}}; \mathbf{x}_i) \\ \mathbf{U}_2(\mu_y; y_i) \end{bmatrix},
\end{aligned}
$$

where $\mathbf{U}_1(\boldsymbol{\mu}_{\mathbf{x}}; \mathbf{x}_i) = \mathbf{x}_i - \boldsymbol{\mu}_{\mathbf{x}}$ and $\mathbf{U}_2(\mu_y; y_i) = y_i - \mu_y$. To combine all available information, one may consider an estimator minimizing the following GMM-type objective function

$$
Q^*(\boldsymbol{\eta}, \theta) = \begin{bmatrix} \mathbf{S}_*(\boldsymbol{\eta}) \\ \mathbf{U}_1(\boldsymbol{\theta}) \\ \mathbf{U}_{PS}(\boldsymbol{\theta}; \boldsymbol{\phi}) \end{bmatrix}^T \left[ V \left\{ \begin{array}{c} \mathbf{S}_*(\boldsymbol{\eta}) \\ \mathbf{U}_1(\boldsymbol{\theta}) \\ \mathbf{U}_{PS}(\boldsymbol{\theta}; \boldsymbol{\phi}) \end{array} \right\} \right]^{-1} \begin{bmatrix} \mathbf{S}_*(\boldsymbol{\eta}) \\ \mathbf{U}_1(\boldsymbol{\theta}) \\ \mathbf{U}_{PS}(\boldsymbol{\theta}; \boldsymbol{\phi}) \end{bmatrix}, \quad (3.26)
$$

where $\mathbf{S}_*(\boldsymbol{\eta})$ is defined in (3.24). First, we partition the variance matrix in (3.26) as follows

$$
\begin{aligned}
& V\left\{\begin{array}{c}\mathbf{S}_*(\boldsymbol{\eta}) \\ \mathbf{U}_1(\boldsymbol{\theta}) \\ \mathbf{U}_{PS}(\boldsymbol{\theta};\boldsymbol{\phi})\end{array}\right\} \\
= & \left[\begin{array}{l|cr}
V\{\mathbf{S}_*(\boldsymbol{\eta})\} & C\{\mathbf{S}_*(\boldsymbol{\eta}),\mathbf{U}_1(\boldsymbol{\theta})\} & C\{\mathbf{S}_*(\boldsymbol{\eta}),\mathbf{U}_{PS}(\boldsymbol{\theta};\boldsymbol{\phi})\} \\ \hline
C\{\mathbf{U}_1(\boldsymbol{\theta}),\mathbf{S}_*(\boldsymbol{\eta})\} & V\{\mathbf{U}_1(\boldsymbol{\theta})\} & C\{\mathbf{U}_1(\boldsymbol{\theta}),\mathbf{U}_{PS}(\boldsymbol{\theta};\boldsymbol{\phi})\} \\
C\{\mathbf{U}_{PS}(\boldsymbol{\theta};\boldsymbol{\phi}),\mathbf{S}_*(\boldsymbol{\eta})\} & C\{\mathbf{U}_{PS}(\boldsymbol{\theta};\boldsymbol{\phi}),\mathbf{U}_1(\boldsymbol{\theta})\} & V\{\mathbf{U}_{PS}(\boldsymbol{\theta};\boldsymbol{\phi})\}
\end{array}\right] \\
\triangleq & \left(\begin{array}{c|c}\mathbf{V}_{ss} & \mathbf{V}_{su} \\ \hline \mathbf{V}_{us} & \mathbf{V}_{uu}\end{array}\right),
\end{aligned}
$$

$\mathbf{V}_{su} = \mathbf{V}_{us}^T$ and

$$
\begin{aligned}
\mathbf{V}_{su} &= \left[\ \mathbf{V}_{su,0} \quad \mathbf{V}_{su,1} \quad \mathbf{V}_{su,2}\ \right] \\
&= \left[\ C\{\mathbf{S}_*(\boldsymbol{\eta}),\mathbf{U}_1(\boldsymbol{\mu}_\mathbf{x})\} \quad C\{\mathbf{S}_*(\boldsymbol{\eta}),\mathbf{U}_{PS,1}(\theta;\boldsymbol{\phi})\} \quad C\{\mathbf{S}_*(\boldsymbol{\eta}),\mathbf{U}_{PS,2}(\theta;\boldsymbol{\phi})\}\ \right] \\
\mathbf{V}_{uu} &= \begin{bmatrix} \mathbf{V}_{uu,00} & \mathbf{V}_{uu,01} & \mathbf{V}_{uu,02} \\ \mathbf{V}_{uu,10} & \mathbf{V}_{uu,11} & \mathbf{V}_{uu,12} \\ \mathbf{V}_{uu,20} & \mathbf{V}_{uu,21} & \mathbf{V}_{uu,22} \end{bmatrix} \\
&= \begin{bmatrix}
V\{\mathbf{U}(\theta)\} & C\{\mathbf{U}(\theta),\mathbf{U}_{PS,1}(\theta;\boldsymbol{\phi})\} & C\{\mathbf{U}(\theta),\mathbf{U}_{PS,2}(\theta;\boldsymbol{\phi})\} \\
C\{\mathbf{U}_{PS,1}(\theta;\boldsymbol{\phi}),\mathbf{U}(\theta)\} & V\{\mathbf{U}_{PS,1}(\theta;\boldsymbol{\phi})\} & C\{\mathbf{U}_{PS,1}(\theta;\boldsymbol{\phi}),\mathbf{U}_{PS,2}(\theta;\boldsymbol{\phi})\} \\
C\{\mathbf{U}_{PS,2}(\theta;\boldsymbol{\phi}),\mathbf{U}(\theta)\} & C\{\mathbf{U}_{PS,2}(\theta;\boldsymbol{\phi}),\mathbf{U}_{PS,1}(\theta;\boldsymbol{\phi})\} & V\{\mathbf{U}_{PS,2}(\theta;\boldsymbol{\phi})\}
\end{bmatrix}.
\end{aligned}
$$

Similar to the decomposition in (3.25), the objective function $Q^*(\boldsymbol{\eta},\theta)$ in (3.26) can be decomposed as follows:

$$
Q^*(\boldsymbol{\eta},\theta) = Q_1^*(\boldsymbol{\eta}) + Q_2^*(\boldsymbol{\mu}_\mathbf{x}|\boldsymbol{\eta}) + Q_3^*(\mu_y|\boldsymbol{\eta},\boldsymbol{\mu}_\mathbf{x}), \tag{3.27}
$$

where $Q_1^*(\boldsymbol{\eta}) = \mathbf{S}_*^T(\boldsymbol{\eta})\mathbf{V}_{ss}^{-1}\mathbf{S}_*(\boldsymbol{\eta})$,

$$\begin{aligned}
Q_2^*(\boldsymbol{\mu}_{\mathbf{x}}|\boldsymbol{\eta}) &= \mathbf{G}_2^T(\boldsymbol{\mu}_{\mathbf{x}}|\boldsymbol{\eta})\mathbf{V}_{u|s,\mathbf{xx}}^{-1}\mathbf{G}_2(\boldsymbol{\mu}_{\mathbf{x}}|\boldsymbol{\eta}), \\
Q_3^*(\mu_y|\boldsymbol{\eta},\boldsymbol{\mu}_{\mathbf{x}}) &= \mathbf{G}_3^T(\mu_y|\boldsymbol{\eta},\boldsymbol{\mu}_{\mathbf{x}})\mathbf{V}_{u|s,y|\mathbf{x}}^{-1}\mathbf{G}_3(\mu_y|\boldsymbol{\eta},\boldsymbol{\mu}_{\mathbf{x}}), \\
\mathbf{G}_2(\boldsymbol{\mu}_{\mathbf{x}}|\boldsymbol{\eta}) &= \begin{bmatrix} \mathbf{U}_1(\boldsymbol{\theta}) - \mathbf{V}_{us,0}\mathbf{V}_{ss}^{-1}\mathbf{S}_*(\boldsymbol{\eta}) \\ \mathbf{U}_{PS,1}(\boldsymbol{\theta};\boldsymbol{\phi}) - \mathbf{V}_{us,1}\mathbf{V}_{ss}^{-1}\mathbf{S}_*(\boldsymbol{\eta}) \end{bmatrix}, \\
\mathbf{G}_3(\mu_y|\boldsymbol{\eta},\boldsymbol{\mu}_{\mathbf{x}}) &= \mathbf{U}_{PS,2}(\boldsymbol{\theta};\boldsymbol{\phi}) - \mathbf{V}_{u|s,y\mathbf{x}}\mathbf{V}_{u|s,\mathbf{xx}}^{-1}\mathbf{G}_2(\boldsymbol{\mu}_{\mathbf{x}}|\boldsymbol{\eta}) - \mathbf{V}_{us,2}\mathbf{V}_{ss}^{-1}\mathbf{S}_*(\boldsymbol{\eta}),
\end{aligned}$$

$\mathbf{V}_{u|s} = \mathbf{V}_{uu} - \mathbf{V}_{us}\mathbf{V}_{ss}^{-1}\mathbf{V}_{su}$,

$$\mathbf{V}_{u|s} = \left[\begin{array}{c|c} \mathbf{V}_{u|s,\mathbf{xx}} & \mathbf{V}_{u|s\mathbf{x},y} \\ \hline \mathbf{V}_{u|s,y\mathbf{x}} & \mathbf{V}_{u|s,yy} \end{array}\right] = \left[\begin{array}{cc|c} \mathbf{V}_{u|s,00} & \mathbf{V}_{u|s,01} & \mathbf{V}_{u|s,02} \\ \mathbf{V}_{u|s,10} & \mathbf{V}_{u|s,11} & \mathbf{V}_{u|s,12} \\ \hline \mathbf{V}_{u|s,20} & \mathbf{V}_{u|s,21} & \mathbf{V}_{u|s,22} \end{array}\right],$$

$\mathbf{V}_{u|s,ij} = \mathbf{V}_{uu,ij} - \mathbf{V}_{us,i}\mathbf{V}_{ss}^{-1}\mathbf{V}_{su,j}$, for $i = 0,1,2$ and $j = 0,1,2$

$$\mathbf{V}_{u|s,y|\mathbf{x}} = \mathbf{V}_{u|s,yy} - \mathbf{V}_{u|s,y\mathbf{x}}\mathbf{V}_{u|s,\mathbf{xx}}^{-1}\mathbf{V}_{u|s,\mathbf{x}y}.$$

Since $\mathbf{V}_{us,0} = C\{\mathbf{S}_*(\boldsymbol{\eta}), \mathbf{U}_1(\boldsymbol{\theta})\} = \mathbf{0}$ and

$$C\{\mathbf{U}_{PS}(\boldsymbol{\theta};\boldsymbol{\phi}), \mathbf{U}_1(\boldsymbol{\theta})\} = \begin{bmatrix} V\{\mathbf{U}_1(\boldsymbol{\theta})\} \\ C\{\mathbf{U}_2(\boldsymbol{\theta}), \mathbf{U}_1(\boldsymbol{\theta})\} \end{bmatrix},$$

$Q_1^*(\boldsymbol{\eta})$ is minimized at $\boldsymbol{\eta} = \hat{\boldsymbol{\eta}}$, $Q_2^*(\boldsymbol{\mu}_{\mathbf{x}}|\hat{\boldsymbol{\eta}})$ is minimized at $\boldsymbol{\mu}_{\mathbf{x}} = \bar{\mathbf{x}}_n$, and $Q_3(\mu_y|\hat{\boldsymbol{\eta}}, \bar{\mathbf{x}}_n)$ is minimized at $\mu_y = \hat{\mu}_y^*$, where

$$\hat{\mu}_y^* = \hat{\mu}_{y,PS} - \mathbf{B}_x^*(\hat{\boldsymbol{\mu}}_{\mathbf{x},PS} - \bar{\mathbf{x}}_n), \tag{3.28}$$

where $\hat{\mu}_{y,PS} = \sum_{i=1}^n \delta_i\pi_i^{-1}(\hat{\boldsymbol{\phi}})y_i / \sum_{i=1}^n \delta_i\pi_i^{-1}(\hat{\boldsymbol{\phi}})$, $\hat{\boldsymbol{\mu}}_{\mathbf{x},PS} = \sum_{i=1}^n \delta_i\pi_i^{-1}(\hat{\boldsymbol{\phi}})\mathbf{x}_i / \sum_{i=1}^n \delta_i\pi_i^{-1}(\hat{\boldsymbol{\phi}})$, and

$$\mathbf{B}_x^* = (\mathbf{V}_{uu,21} - \mathbf{V}_{us,2}\mathbf{V}_{ss}^{-1}\mathbf{V}_{su,1})(\mathbf{V}_{uu,11} - \mathbf{V}_{uu,00} - \mathbf{V}_{us,1}\mathbf{V}_{ss}^{-1}\mathbf{V}_{su,1})^{-1}.$$

Note that $\hat{\mu}_y^*$ in (3.28) is an optimal regression estimator since $\mathbf{B}_x^* = C\{\hat{\mu}_{y,PS}, \hat{\boldsymbol{\mu}}_{\mathbf{x},PS} - \bar{\mathbf{x}}_n\}[V\{\hat{\boldsymbol{\mu}}_{\mathbf{x},PS} - \bar{\mathbf{x}}_n\}]^{-1}$ minimizing the variance among the class of linear estimators of

$(\hat{\boldsymbol{\mu}}_{\mathbf{x},PS}, \bar{\mathbf{x}}_n, \hat{\mu}_{y,PS})$. Yet, $\mathbf{B}_x^*$ is still a parameter which needs to be estimated to obtain the estimator minimizing (3.27). Thus, finding a consistent estimator for

$$V \begin{bmatrix} \mathbf{S}_*(\boldsymbol{\eta}) \\ \mathbf{U}_1(\boldsymbol{\theta}) \\ \mathbf{U}_{PS}(\boldsymbol{\theta};\boldsymbol{\phi}) \end{bmatrix}$$

is critical. Using $E\{\mathbf{S}_*(\boldsymbol{\eta})\} = \mathbf{0}$, $E\{\mathbf{U}_1(\boldsymbol{\theta})\} = \mathbf{0}$, and $E\{\mathbf{U}_{PS}(\boldsymbol{\theta};\boldsymbol{\phi})\} = \mathbf{0}$ at $\boldsymbol{\eta} = (\boldsymbol{\gamma}_0^T, \boldsymbol{\phi}_0^T)^T$ and $\theta = \theta_0$, the variance matrix can be estimated by

$$\begin{bmatrix} \hat{\mathbf{V}}_{\mathbf{ss}} & \hat{\mathbf{V}}_{\mathbf{su}} \\ \hat{\mathbf{V}}_{\mathbf{us}} & \hat{\mathbf{V}}_{\mathbf{uu}} \end{bmatrix}, \tag{3.29}$$

where

$$\begin{aligned}
\hat{\mathbf{V}}_{\mathbf{ss}} &= \sum_{i=1}^{n} \hat{\mathbf{S}}_i \hat{\mathbf{S}}_i^T \\
\hat{\mathbf{V}}_{\mathbf{su}} &= \begin{bmatrix} \hat{\mathbf{V}}_{su,0} & \hat{\mathbf{V}}_{su,1} & \hat{\mathbf{V}}_{su,2} \end{bmatrix} = \sum_{i=1}^{n} \begin{bmatrix} \mathbf{0} & \delta_i \hat{\pi}_i^{-1} \hat{\mathbf{S}}_i \hat{\mathbf{U}}_{1i}^T & \delta_i \hat{\pi}_i^{-1} \hat{\mathbf{S}}_i \hat{\mathbf{U}}_{2i}^T \end{bmatrix} \\
\hat{\mathbf{V}}_{\mathbf{us}}^T &= \begin{bmatrix} \hat{\mathbf{V}}_{us,0}^T & \hat{\mathbf{V}}_{us,1}^T & \hat{\mathbf{V}}_{us,2}^T \end{bmatrix} = \hat{\mathbf{V}}_{\mathbf{su}} \\
\hat{\mathbf{V}}_{\mathbf{uu}} &= \begin{bmatrix} \hat{\mathbf{V}}_{uu,11} & \hat{\mathbf{V}}_{uu,12} \\ \hat{\mathbf{V}}_{uu,21} & \hat{\mathbf{V}}_{uu,22} \end{bmatrix}, \\
&= \begin{bmatrix} \delta_i \hat{\pi}_i^{-2}(1-\hat{\pi}_i) \hat{\mathbf{U}}_{1i} \hat{\mathbf{U}}_{1i}^T & \delta_i \hat{\pi}_i^{-2}(1-\hat{\pi}_i) \hat{\mathbf{U}}_{1i} \hat{\mathbf{U}}_{2i}^T \\ \delta_i \hat{\pi}_i^{-2}(1-\hat{\pi}_i) \hat{\mathbf{U}}_{2i} \hat{\mathbf{U}}_{1i}^T & \delta_i \hat{\pi}_i^{-2}(1-\hat{\pi}_i) \hat{\mathbf{U}}_{2i} \hat{\mathbf{U}}_{2i}^T \end{bmatrix},
\end{aligned}$$

where $\hat{\pi}_i = \pi_i(\hat{\boldsymbol{\phi}})$, $\hat{\mathbf{S}}_i = \mathbf{S}_{*i}(\hat{\boldsymbol{\eta}})$, $\hat{\mathbf{U}}_{1i} = \mathbf{U}_1(\hat{\boldsymbol{\mu}}_{\mathbf{x}}; \mathbf{x}_i)$, $\hat{\mathbf{U}}_i = [\mathbf{U}_1^T(\hat{\boldsymbol{\mu}}_{\mathbf{x}}; \mathbf{x}_i), \mathbf{U}_2^T(\hat{\mu}_y; y_i)]^T$,

$$\mathbf{S}_{*i}(\boldsymbol{\eta}) = \begin{bmatrix} \delta_i \mathbf{s}_{1i}(\boldsymbol{\gamma}) \\ \mathbf{s}_{2i}(\boldsymbol{\phi};\boldsymbol{\gamma}) \end{bmatrix}.$$

Note that each variance estimator term in (3.29) is computable, since the $y_i$'s are observed for $\delta_i = 1$. Thus, the resulting estimator minimizing (3.27) can be written as

$$\hat{\mu}_y^* = \hat{\mu}_{y,PS} - \hat{\mathbf{B}}_x^*(\hat{\boldsymbol{\mu}}_{\mathbf{x},PS} - \bar{\mathbf{x}}_n),$$

where

$$\hat{\mathbf{B}}_x^* = [\hat{\mathbf{V}}_{uu,21} - \hat{\mathbf{V}}_{us,2}\hat{\mathbf{V}}_{ss}^{-1}\hat{\mathbf{V}}_{su,2}][\hat{\mathbf{V}}_{uu,11} - \hat{\mathbf{V}}_{uu,00} - \hat{\mathbf{V}}_{us,1}\hat{\mathbf{V}}_{ss}^{-1}\hat{\mathbf{V}}_{su,1}]^{-1}.$$

Using that $\hat{\mu}_y^*$ is the solution to $\partial Q_*(\boldsymbol{\theta}, \boldsymbol{\phi})/\partial \mu_y$, the variance of $\hat{\mu}_y^*$ can be estimated by

$$\hat{V}_\mu = \left\{\sum_{i=1}^n \frac{\delta_i}{\hat{\pi}_i}\right\}^{-1} \hat{V}_U \left\{\sum_{i=1}^n \frac{\delta_i}{\hat{\pi}_i}\right\}^{-1},$$

where

$$\begin{aligned}\hat{V}_U &= \hat{\mathbf{V}}_{uu,22} - \hat{\mathbf{V}}_{us,2}\hat{\mathbf{V}}_{ss}^{-1}\hat{\mathbf{V}}_{su,2} - \hat{\mathbf{V}}_{uu,20}\hat{\mathbf{V}}_{uu,00}^{-1}\hat{\mathbf{V}}_{uu,02} \\ &\quad -\hat{\boldsymbol{\alpha}}^T[\hat{\mathbf{V}}_{uu,11} - \hat{\mathbf{V}}_{uu,00} - \hat{\mathbf{V}}_{su,1}^T\hat{\mathbf{V}}_{ss}^{-1}\hat{\mathbf{V}}_{su,1}]^{-1}\hat{\boldsymbol{\alpha}}.\end{aligned}$$

and $\hat{\boldsymbol{\alpha}} = \hat{\mathbf{V}}_{uu,12} - \hat{\mathbf{V}}_{uu,02} - \hat{\mathbf{V}}_{us,1}\hat{\mathbf{V}}_{ss}^{-1}\hat{\mathbf{V}}_{su,2}$. Note that $\hat{V}_U$ is the linearized variance estimator of the estimating equation for $\theta$, $U_{PS,2}(\theta; \boldsymbol{\phi})$, taking into account the effect of using estimated $\boldsymbol{\phi}$

## 3.6 Simulation Studies

### 3.6.1 Simulation Study I

To investigate the finite sample properties of the proposed method, limited simulation studies are performed. In the simulation, two auxiliary variables are generated from the following bivariate normal distribution:

$$\mathbf{x}_i = \begin{pmatrix} x_{1i} \\ x_{2i} \end{pmatrix} \overset{iid}{\sim} N\left[\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right].$$

We generated the two sets of study variable $y$ as follows:

$$\begin{array}{lll} \text{Population A :} & y = -1 + x_{1i} + 0.5x_{2i} + e_i & e_i \sim N(0,1) \\ \text{Population B :} & y = (x_2 - 2)^2 + e_i & e_i \sim N(0,1) \end{array}$$

The value of the parameter of interest $\theta = E(y)$ is 1 in both cases.

For the response mechanism, we generated $\delta_i \overset{indep}{\sim} Bernoulli(p_i)$, where $p_i(\boldsymbol{\phi}) = [1+\exp(-0.2-0.5x_{1i}-0.3y_i)]^{-1}$. Thus, variable $x_{2i}$ plays the role of the nonresponse instrumental variable. The average response rate is 70% for both Population A and Population B. For each population, Monte Carlo samples of size $n = 500$ are independently generated $B = 2,000$ times.

Two response models of the form

$$\text{Model 1}: \quad P(\delta=1|x_1,x_2,y) = \pi(x_1,y;\boldsymbol{\phi}) = \frac{\exp(\phi_0+\phi_1x_1+\phi_2y)}{1+\exp(\phi_0+\phi_1x_1+\phi_2y)}, \tag{3.30}$$

$$\text{Model 2}: \quad P(\delta=1|x_1,x_2,y) = \pi(x_1,y;\boldsymbol{\phi}) = \Phi(\phi_0+\phi_1x_1+\phi_2y), \tag{3.31}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, were postulated for parameter estimation. The postulated model (3.30) is correctly specified for this simulation study, while the second postulated model (3.31) is misspecified. Given $\pi_i(\boldsymbol{\phi}) = \pi(x_{1i},y_i;\boldsymbol{\phi})$, in each sample, we considered estimating $\theta$ by the following 3 estimators :

1. PS estimator : the solution to $U_{PS}(\theta;\hat{\boldsymbol{\phi}}_p) = \sum_{i=1}^n \delta_i\pi_i^{-1}(\hat{\boldsymbol{\phi}}_p)(y_i-\theta) = 0$,

$$\hat{\theta}_1 = \frac{\sum_{i=1}\delta_i\pi_i^{-1}(\hat{\boldsymbol{\phi}}_p)y_i}{\sum_{i=1}\delta_i\pi_i^{-1}(\hat{\boldsymbol{\phi}}_p)},$$

where $\hat{\boldsymbol{\phi}}_p$ is the solution to (3.13) assuming the response model (3.30) and a model,

$$f_1(y|\mathbf{x};\boldsymbol{\gamma}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(y-\mathbf{x}^T\boldsymbol{\beta})^2}{2\sigma^2}\right\}, \tag{3.32}$$

for population A and

$$f_1(y|\mathbf{x};\boldsymbol{\gamma}) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(y-\beta_0-\beta_1x_{2i}-\beta_2x_{2i}^2)^2}{2\sigma^2}\right\}, \tag{3.33}$$

for population B.

2. Chang and Kott (2008) estimator (CK estimator)

$$\hat{\theta}_2 = \frac{\sum_{i=1}\delta_iy_i\pi_i^{-1}(\hat{\boldsymbol{\phi}}_{ck})}{\sum_{i=1}\delta_i\pi_i^{-1}(\hat{\boldsymbol{\phi}}_{ck})},$$

where $\hat{\boldsymbol{\phi}}_{ck}$ is the solution to calibration condition (3.6) for population A and

$$\sum_{i=1}^{n} \frac{\delta_i}{\pi_i(\boldsymbol{\phi})} \begin{pmatrix} 1 \\ x_{2i} \\ x_{2i}^2 \end{pmatrix} = \sum_{i=1}^{n} \begin{pmatrix} 1 \\ x_{2i} \\ x_{2i}^2 \end{pmatrix}$$

for population B.

3. GMM estimator : the estimator minimizing (3.26) in Section 3.5, denoted by $\hat{\theta}_3$.

4. MAR estimator : the solution to $U_{PS}(\theta; \hat{\boldsymbol{\phi}}_p) = \sum_{i=1}^{n} \delta_i \pi_i^{-1}(\hat{\boldsymbol{\phi}}_p)(y_i - \theta) = 0$,

$$\hat{\theta}_4 = \frac{\sum_{i=1} \delta_i \pi_i^{-1}(\hat{\boldsymbol{\phi}}) y_i}{\sum_{i=1} \delta_i \pi_i^{-1}(\hat{\boldsymbol{\phi}})},$$

where $\hat{\boldsymbol{\phi}}$ is the maximum likelihood estimator assuming ignorable nonresponse $P(\delta = 1|x_1, x_2, y) = \{1 + \exp(-\phi_0 - \phi_1 x_1 - \phi_2 x_2)\}^{-1}$ for model 1 and $P(\delta = 1|x_1, x_2, y) = \Phi(\phi_0 + \phi_1 x_1 + \phi_2 x_2)$ for model 2.

Table 3.1 presents the Monte Carlo biases, variances, and mean squared errors (MSE) of the four point estimators and percent relative biases and t-statistics of the variance estimators of the proposed methods that are computed from the Monte Carlo samples of size $B = 2,000$. The linearized variance estimator proposed in Section 3.4 is computed for $\hat{\theta}_1$, and the linearized variance estimates for $\hat{\theta}_3$ derived in Section 3.5 is computed. The percent relative bias of the variance estimator $\hat{V}(\hat{\theta})$ is calculated as $100 \times [E_{MC}\{\hat{V}(\hat{\theta})\} - V_{MC}(\hat{\theta})]/V_{MC}(\hat{\theta})$, where $E_{MC}(\cdot)$ and $V_{MC}(\cdot)$ denote the Monte Carlo expectation and the Monte Carlo variance, respectively. The t-statistic is computed as

$$t = \frac{E_{MC}\{\hat{V}(\hat{\theta})\} - V_{MC}(\hat{\theta})}{\sqrt{E_{MC}\left[[\hat{V}(\hat{\theta}) - \{\hat{\theta} - E_{MC}(\hat{\theta})\}^2 - \{E_{MC}(\hat{V}(\hat{\theta})) - V_{MC}(\hat{\theta})\}]^2\right]/B}}. \tag{3.34}$$

The test statistic (3.34) is for testing if the variance estimator has zero bias. See Kim (2004).

While the PS estimator under ignorable nonresponse assumption still shows significant nonresponse bias, the first three estimators (PS estimator, CK estimator, GMM estimator) are nearly unbiased in both populations with model 1 even though the postulated model (3.32) is not correctly specified for the PS estimator. The two proposed estimators are less biased than the CK estimator under model 1. With model 2, the incorrectly specified model for response probabilities, the bias reduction of PS and CK estimators was not large enough for all three estimators, but the GMM estimator presents considerable bias reduction in population A. In terms of variances, the PS estimator shows slightly smaller values than the CK estimator in all scenarios, because the response model parameter $\boldsymbol{\phi}$ is estimated more efficiently using the maximum likelihood method than the CK estimator using the method of moments estimation. Table 3.2 shows that the maximum likelihood method is much more efficient than the CK method when estimating the response model parameters. In Table 3.1, the GMM estimator shows the best efficiency because it incorporates the auxiliary information of $x_1$ and $x_2$ efficiently.

For variance estimation, results in Table 3.1 show that the linearization variance estimator works well in all scenarios, with the absolute values of relative biases all less than 5%.

We also present the Monte Carlo biases and variances of the two estimators for the response model parameter $\boldsymbol{\phi}$ under the correctly specified response probability model (model 1) in Table 3.2. One is obtained by solving the mean score equation (3.13) and is used for $\hat{\theta}_1$, and the other is obtained by solving the calibration equation (3.6) and is used for $\hat{\theta}_2$. The two estimators are nearly unbiased, and the MLE solving the mean score equation is much more efficient than the CK estimators.

### 3.6.2 Simulation Study II

In the second simulation study, we investigate the PS estimators with the discrete study variable $y$ and the discrete auxiliary variables $x_1$ and $x_2$. We generate a stratified finite population of $(x_1, x_2, y)$ with four strata $(h = 1, 2, 3, 4)$ of size $N_1 = 1,000$, $N_2 =$

**Table 3.1 Monte Carlo Bias, Variance, and Mean Squared Error(MSE) of $\hat{\theta}$ and Monte Carlo Relative Bias(R.Bias) and Test Statistics(t.stat) of $\hat{V}(\hat{\theta})$ from Simulation I**

| Population | Model | Estimator | Point Estimation | | | Variance Estimation | |
|---|---|---|---|---|---|---|---|
| | | | Bias | Variance | MSE | R.Bias | t-stat |
| A | Model 1 | PS | 0.000 | 0.0125 | 0.0125 | -1.60 | -0.91 |
| | | CK | 0.006 | 0.0135 | 0.0135 | | |
| | | GMM | 0.000 | 0.0101 | 0.0101 | -1.39 | -0.63 |
| | | MAR | 0.086 | 0.0071 | 0.0145 | | |
| | Model 2 | PS | 0.115 | 0.0075 | 0.0207 | -3.57 | -1.88 |
| | | CK | 0.105 | 0.0086 | 0.0196 | | |
| | | GMM | 0.010 | 0.0062 | 0.0063 | -3.11 | -1.54 |
| | | MAR | 0.090 | 0.0072 | 0.0152 | | |
| B | Model 1 | PS | -0.001 | 0.0162 | 0.0162 | -1.90 | -0.81 |
| | | CK | 0.003 | 0.0189 | 0.0189 | | |
| | | GMM | 0.000 | 0.0150 | 0.0150 | -1.50 | -0.76 |
| | | MAR | 0.130 | 0.0123 | 0.0292 | | |
| | Model 2 | PS | 0.125 | 0.0134 | 0.0291 | -3.49 | -1.81 |
| | | CK | 0.112 | 0.0154 | 0.0279 | | |
| | | GMM | 0.113 | 0.0126 | 0.0255 | -3.21 | -1.61 |
| | | MAR | 0.128 | 0.0123 | 0.0287 | | |

**Table 3.2 Monte Carlo Bias and Variance of $\hat{\phi}$ from Simulation I**

| Population | Model | Parameter | Point Estimation | |
|---|---|---|---|---|
| | | | Bias | Variance |
| A | Model 1 | MLE $\hat{\phi}_0$ | 0.024 | 0.0252 |
| | | $\hat{\phi}_1$ | 0.014 | 0.0679 |
| | | $\hat{\phi}_2$ | 0.003 | 0.0437 |
| | | CK $\hat{\phi}_0$ | 0.040 | 0.0730 |
| | | $\hat{\phi}_1$ | -0.002 | 0.1438 |
| | | $\hat{\phi}_2$ | 0.027 | 0.1000 |
| B | Model 1 | MLE $\hat{\phi}_0$ | 0.021 | 0.0273 |
| | | $\hat{\phi}_1$ | 0.020 | 0.0444 |
| | | $\hat{\phi}_2$ | 0.010 | 0.0914 |
| | | CK $\hat{\phi}_0$ | 0.033 | 0.1360 |
| | | $\hat{\phi}_1$ | 0.006 | 0.1232 |
| | | $\hat{\phi}_2$ | 0.021 | 0.2616 |

**Table 3.3 Conditional probability of $y$ given auxiliary variables $x_1$ and $x_2$ for Simulation II**

| $(P(y=1\|\mathbf{x}), P(y=2\|\mathbf{x}), P(y=3\|\mathbf{x}))$ | $x_2 = 1$ | $x_2 = 2$ | $x_2 = 3$ |
|---|---|---|---|
| $x_1 = 1$ | $(0.6, 0.2, 0.2)$ | $(0.5, 0.25, 0.25)$ | $(0.3, 0.35, 0.35)$ |
| $x_1 = 0$ | $(0.5, 0.25, 0.25)$ | $(0.4, 0.3, 0.3)$ | $(0.2, 0.4, 0.4)$ |

$2,000$, $N_3 = 3,000$ and $N_4 = 4,000$. The first auxiliary variable, $x_1$, is generated from a Bernoulli distribution with $P(x_1 = 1) = 0.51$ and the other auxiliary variable, $x_2$, takes a value from $\{1, 2, 3\}$ with probability $(0.3, 0.6, 0.1)$. The study variable also takes one value out of $\{1, 2, 3\}$ with different probability depending on the auxiliary variables:

The parameter of interest is $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ where $\theta_l = N^{-1} \sum_{i=1}^{N} I_{i(++l)}$ and $I_{i(++l)}$ takes the value of one if $y_i = l$ and takes the value of zero otherwise. The parameter value is about $(\theta_1, \theta_2, \theta_3) = (0.4641, 0.2688, 0.2671)$. For the response mechanism, we generate $\delta_i$ from a Bernoulli distribution where the response probability is different depending on $\mathbf{x}_2$ and $y$:

$$P(\delta_i = 1|x_{1i}, x_{2i}, y_i) = 1/\{1+\exp\left(-1 - 0.5I_{i(+1+)} + I_{i(+3+)} - I_{i(++1)} + 0.5I_{i(++2)} + I_{i(++3)}\right)\},$$

where $I_{i(+k+)}$ takes the value of 1 if $x_{2i} = k$ and takes the value of 0 otherwise. The average response rate is about 71%.

We use $B = 2,000$ Monte Carlo samples of size $n = 400$ in this simulation. A stratified sample is independently generated without replacement, where a simple random sample of size $n_1 = n_2 = n_3 = n_4 = 100$ is selected from each stratum. Thus, the sampling weights are $N_h/n_h = 10, 20, 30, 40$ for $h = 1, 2, 3, 4$, respectively. We assume a response model

$$P(\delta_i = 1|x_{1i} = j, x_{2i} = k, y_i = l) = \pi_i(\boldsymbol{\phi}) = \{1 + \exp\left(-\phi_0 - \phi_{1k} I_{i(+k+)} - \phi_{2l} I_{i(++l)}\right)\}^{-1}$$

for $k = 1, 3$ and $l = 2, 3$. Thus, $x_1$ plays the role of a nonresponse instrument in this setup. Since the conditional distribution of $y$ given $x_1$ and $x_2$ for respondents is assumed to be $P(y_i = l | x_{1i} = j, x_{2i} = k, \delta_i = 1) = p^*_{i,jkl}$, the respondent's model $f_1(y|\mathbf{x})$ is fully nonparametric. Under complete response, the (pseudo) MLE of $\boldsymbol{\phi}$ can be obtained by solving the pseudo score equation :

$$\sum_{i \in A} d_i \{\delta_i - \pi_i(\boldsymbol{\phi})\} \mathbf{h}_i(\boldsymbol{\phi}) = \mathbf{0},$$

where $d_i$ is the sampling weight and $\mathbf{h}_i(\boldsymbol{\phi}) = (1, I_{i(+1+)}, I_{i(+3+)}, I_{i(++2)}, I_{i(++3)})^T$. To obtain the pseudo maximum likelihood estimator of $\boldsymbol{\phi} = (\phi_0, \phi_{11}, \phi_{13}, \phi_{22}, \phi_{23})$ in the presence of missing data, we solve

$$\sum_{i \in A} \sum_{\delta_i = 1} d_i \{\delta_i - \pi_i(\boldsymbol{\phi})\} \mathbf{h}_i(\boldsymbol{\phi}) + \sum_{i \in A} \sum_{\delta_i = 0} \sum_{l=1}^{3} d_i w_{i,l}^{(t)} \{\delta_i - \pi_{i,l}(\boldsymbol{\phi})\} \mathbf{h}_{i,l}(\boldsymbol{\phi}) = 0,$$

where

$$w_{i,l}^{(t)} = \frac{p_{i,jkl}\{\pi_{i,l}^{-1}(\boldsymbol{\phi}^{(t)}) - 1\}}{\sum_{m=1}^{3} p_{i,jkm}\{\pi_{i,m}^{-1}(\boldsymbol{\phi}^{(t)}) - 1\}},$$

$\pi_{i,l}(\boldsymbol{\phi}) = \{1 + \exp(-\phi_0 - \phi_1 I_{i(+x_{2i}+)} - \phi_{2l} I_{i(++l)})\}^{-1}$, $\mathbf{h}_{i,1}(\boldsymbol{\phi}) = (1, I_{i(+1+)}, I_{i(+3+)}, 0, 0)^T$, $\mathbf{h}_{i,2}(\boldsymbol{\phi}) = (1, I_{i(+1+)}, I_{i(+3+)}, 1, 0)^T$, and $\mathbf{h}_{i,3}(\boldsymbol{\phi}) = (1, I_{i(+1+)}, I_{i(+3+)}, 0, 1)^T$, given $\boldsymbol{\phi}^{(t)}$ and update $\boldsymbol{\phi}^{(t+1)}$ until convergence. Once the MLE of $\boldsymbol{\phi}$ is found, the resulting estimator is the solution to

$$\mathbf{U}_{PS}(\boldsymbol{\theta}, \hat{\boldsymbol{\phi}}) \equiv \sum_{i=1}^{n} d_i \frac{\delta_i}{\pi_i(\hat{\boldsymbol{\phi}})} \left\{ \begin{pmatrix} I_{i(++1)} \\ I_{i(++2)} \\ I_{i(++3)} \end{pmatrix} - \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} \right\} = \mathbf{0},$$

which can be expressed as

$$\hat{\boldsymbol{\theta}}_{PS} = (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)^T = \frac{\sum_{i \in A} d_i \delta_i \hat{\pi}_i^{-1}}{\sum_{j \in A} d_j \delta_j \hat{\pi}_j^{-1}} \left( I_{i(++1)}, I_{i(++2)}, I_{i(++3)} \right)^T = \mathbf{0}, \tag{3.35}$$

where $\hat{\pi}_i = \pi_i(\hat{\boldsymbol{\phi}})$. The Monte Carlo mean, variance, and MSE of $\hat{\boldsymbol{\theta}}_{PS}$ in (3.35) are presented in Table 3.3. with the Monte Carlo relative bias(R.Bias) and t-statistic(t-stat) for the linearization variance estimator of $\hat{\boldsymbol{\theta}}_{PS}$. The variance estimator of $\hat{\theta}_l$ is derived from Theorem 3.1 to incorporate stratified sampling design. The resulting variance estimator is given by

$$\hat{V}_l = \left\{ \sum_{i \in A} d_i \delta_i \pi^{-1}(\hat{\boldsymbol{\phi}}) \right\}^2 \hat{V}_{u,l},$$

for $l = 1, 2, 3$, where

$$\begin{aligned}
\hat{V}_{u,l} &= \hat{V}_{u,l,1} + \hat{V}_{u,l,2}, \\
\hat{V}_{u,l,1} &= \sum_{h=1}^{4} \frac{N_h^2}{N^2} \left( \frac{1}{n_h} - \frac{1}{N_h} \right) s_{u,h}^2, \\
\hat{V}_{u,l,2} &= \frac{1}{N^2} \sum_{i \in A} d_i \delta_i \{1 - \pi_i(\boldsymbol{\phi})\} \{\hat{\pi}_i^{-1}(I_{i(++l)} - \hat{\theta}_l) - \hat{\mathbf{B}}(\hat{\mathbf{s}}_i - \bar{\mathbf{s}}_{0i})\}^2, \\
s_{u,h}^2 &= (n_h - 1)^{-1} \sum_{i \in A_h} (\hat{u}_{lin,i} - \bar{u}_{lin,h})^2, \\
\bar{u}_{lin,h} &= n_h^{-1} \sum_{i \in A_h} \hat{u}_{lin,i}, \\
\hat{u}_{lin,i} &= -\hat{\mathbf{B}}\bar{\mathbf{s}}_{0i} - \delta_i \{\hat{\pi}_i^{-1}(I_{i(++l)} - \hat{\theta}_l) - \hat{\mathbf{B}}(\hat{\mathbf{s}}_i - \bar{\mathbf{s}}_{0i})\}, \\
\hat{\mathbf{B}} &= \left\{ \sum_{i \in A} d_i \delta_i \hat{\pi}_i^{-1}(I_{i(++l)} - \hat{\theta}_l)\hat{\mathbf{s}}_i \right\}^{-1} \left\{ \sum_{i \in A} d_i (1 - \delta_i) \sum_{m=1}^{3} w_{i,m} \hat{\pi}_{i,m}^{-1} \bar{\mathbf{s}}_{0i} \hat{\mathbf{s}}_{i,m}^T \right\}, \\
\bar{\mathbf{s}}_{0i} &= \sum_{m=1}^{3} w_{i,m} \hat{\mathbf{s}}_{i,m}, \\
\hat{\mathbf{s}}_{i,m} &= \{\delta_i - \pi_{i,m}(\hat{\boldsymbol{\phi}})\} \mathbf{h}_{i,m}(\hat{\boldsymbol{\phi}}), \\
w_{i,m} &= \frac{p_{i,jkl}\{\pi_{i,m}^{-1}(\hat{\boldsymbol{\phi}}) - 1\}}{\sum_{p=1}^{3} p_{i,jkp}\{\pi_{i,p}^{-1}(\hat{\boldsymbol{\phi}}) - 1\}}
\end{aligned}$$

and $A_h$ is the set of indices selected in the sample for stratum $h$.

For comparison, the Monte Carlo bias, variance, and MSE of a naive estimator,

$$\hat{\boldsymbol{\theta}}_{l,naive} = \frac{\sum_{i \in A} d_i \delta_i I_{i(++l)}}{\sum_{j \in A} d_j \delta_j},$$

are also presented

Table 3.4 shows that the PS estimator adjusts nonresponse bias while the naive estimator is biased. The linearization variance estimator for the PS estimator is approximately unbiased.

**Table 3.4 MC Bias, Variance, and MSE of $\hat{\theta}$ and R.Bias and t-stat of $\hat{V}(\hat{\theta})$ from Simulation II**

| Parameter | Method | Point Estimation | | | Variance Estimation | |
|---|---|---|---|---|---|---|
| | | Bias | Variance | MSE | R.Bias | t-stat |
| $\theta_1 = 0.4641$ | Naive | 0.1186 | 0.00106 | 0.01509 | | |
| | PS | 0.0038 | 0.00245 | 0.00247 | -1.99 | -1.42 |
| $\theta_2 = 0.2688$ | Naive | -0.0376 | 0.00079 | 0.00220 | | |
| | PS | 0.0044 | 0.00366 | 0.00368 | -1.37 | -0.91 |
| $\theta_3 = 0.2671$ | Naive | -0.0808 | 0.00061 | 0.00714 | | |
| | PS | -0.0082 | 0.00365 | 0.00372 | -2.64 | -1.62 |

## 3.7 Concluding remarks

We have proposed a maximum likelihood method of parameter estimation for the response probability when the response mechanism is nonignorable. In Section 3.3, maximum likelihood estimation for response model parameter $\boldsymbol{\phi}$ assuming the conditional distribution for respondents, $f_1(y|\mathbf{x})$, is discussed. By the nonresponse odds ratio model in (3.7), the information from the respondents' conditional distribution $f_1(y|\mathbf{x})$ and the information from the response mechanism $P(\delta = 1|\mathbf{x}, y)$ are combined to obtain the nonrespondents' conditional distribution $f_0(y|\mathbf{x})$. The EM algorithm is used to compute the MLE of $\boldsymbol{\phi}$, and the E-step is computed without requiring that $f(y \mid \mathbf{x})$ is correctly specified. Once the MLE of $\boldsymbol{\phi}$ is obtained by solving the mean score equation in (3.13), the PS estimator can be computed using the estimated response probabilities and the asymptotic properties are presented in Theorem 3.1. The proposed method is based on the conditional model $f(y \mid \mathbf{x}, \delta = 1)$, but the result does not seem to be sensitive from the departure of the correct specification of the conditional model. For example, in simulation I, we used a normal distribution for $f(y \mid \mathbf{x}, \delta = 1)$, which is not correctly specified, but the simulation results show that the resulting estimates are nearly unbiased. One could consider a nonparametric regression model for $f(y \mid \mathbf{x}, \delta = 1)$, as considered in Kim and Yu (2011), but such a nonparametric method may not be feasible when the dimension of $\mathbf{x}$ is large. Use of the GMM framework is also discussed to improve the efficiency of the PS estimator.

The proposed method also provides consistent estimates for the standard errors of the parameter estimates. Thus, we can test the null hypothesis that the response mechanism is ignorable. If the null hypothesis is not rejected, then we can simply fit a response model with ignorable response. Such a procedure will lead to a pretest procedure, and further investigation on this direction will be a topic of future research.

# CHAPTER 4. CONCLUSION

To handle unit nonresponse in sample surveys or data in the presence of missingness, a technique known as propensity score adjustment is often used. While there are many methods which utilize propensity score adjustment, and much research has been conducted in this area, this dissertation offered new contributions by examining the asymptotic properties of the PS estimators under ignorable nonresponse and nonignorable nonresponse. The variance estimator of PS estimators was also presented as well as the GMM-type methods to improve the efficiency of the PS estimators by incorporating auxiliary variables, which are observed throughout the sample.

In Chapter 2, we discussed the asymptotic properties of PS estimators when the parameter of interest is the finite population mean of the study variable, which is subject to missingness under ignorable nonresponse, and the propensity score was computed from a parametric model for the response probability. We also derived the optimal PS estimator with an additional assumption for the distribution of the study variable, where the propensity score for the optimal PS estimator can be implemented by the augmented propensity model.

In Chapter 3, we considered PS estimation under nonignorable nonresponse. Since the response probability model involves study variable that is subject to missingness, estimation is much more difficult than under ignorable nonresponse. To handle this situation, we have proposed a maximum likelihood method to estimate parameters for the response probability assuming a parametric model for the conditional distribution of the study variable for respondents. The conditional distribution of the study variable for nonrespondents is obtained by combining the response probability model with the conditional distribution of the study variable for respondents. The response probability model is incorporated in the form of a nonresponse odds ratio. An advantage of utilizing the conditional distribution for respondents is that it allows for feasible model diagnostics and model validation, whereas the overall conditional distribution of the study variable does not allow this type of process, because a part of the study variable is not observed. The combination of these two models, the conditional distribution of the study variable for respondents and the response probability model, facilitate the computability of the mean score function. We presented the asymptotic properties of the PS estimator using the maximum likelihood estimators of the PS model parameters that were obtained by solving the mean score equation. The variance estimator of the PS estimator was derived, allowing inference for the PS estimator. Additionally, we incorporated the auxiliary variable information in order to improve the efficiency of the PS estimation via generalized method of moments.

Our research will be extended in the near future. In particular, we will consider a nonparametric regression model for the conditional distribution of the study variable for respondents instead of a parametric model. While nonparametric models may not be feasible when the dimension of the predictors is large, they have an advantage of having weaker assumptions, which in turn reduces the likelihood of model failure. Since, under nonignorable nonresponse, estimation necessarily relies heavily on models for unobserved information, a sensitivity analysis can bring in crucial information into the estimation, and this potential development will be investigated in the future. In addition, Korean general election data that are believed to have a nonignorable response mechanism will be analyzed using the proposed methodology in Chapter 3.

# APPENDIX A. APPENDIX FOR CHAPTER 3

Prior to the main proof of Theorem 3.1, we present notations, regularity conditions, lemmas that will be used in the proof:

- Notations

  - PS estimator

$$\hat{\theta}_{PS}(\boldsymbol{\phi}) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})}y_i. \tag{A.1}$$

  - PS equation

$$U_{PS}(\theta; \boldsymbol{\phi}) = \sum_{i=1}^{n}\frac{\delta_i}{\pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})}u(\theta; \mathbf{x}_i, y_i) \triangleq \sum_{i=1}^{n} u_{PS,i}(\theta; \boldsymbol{\phi}) = 0. \tag{A.2}$$

  - Likelihood function for $\boldsymbol{\phi}$

$$L(\boldsymbol{\phi}) = \prod_{i=1}^{n}\{\pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})\}^{\delta_i}\{1 - \pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})\}^{(1-\delta_i)}, \tag{A.3}$$

  - Score function for $\boldsymbol{\phi}$

$$\begin{aligned}
\mathbf{S}(\boldsymbol{\phi}) &= \mathbf{S}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, y) \\
&= \partial \log L(\boldsymbol{\phi})/\partial \boldsymbol{\phi} \\
&= \sum_{i=1}^{n} \partial[\delta_i \log \pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi}) + (1-\delta_i)\log\{1 - \pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})\}]/\partial \boldsymbol{\phi} \\
&= \sum_{i=1}^{n}\{\delta_i - \pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})\}\mathbf{z}(\mathbf{x}_{1i}, y_1; \boldsymbol{\phi}) \qquad (A.4)\\
&\triangleq \sum_{i=1}^{n}\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_i, y_i), \qquad (A.5)
\end{aligned}$$

$L(\boldsymbol{\phi})$ is defined in (A.3),

$$\mathbf{z}(\mathbf{x}_1, y; \boldsymbol{\phi}) = \partial \text{logit}\{\pi(\mathbf{x}_1, y; \boldsymbol{\phi})\}/\partial \boldsymbol{\phi} \tag{A.6}$$

and $\text{logit}(p) = \log\{p/(1-p)\}$

– Mean score function for $\boldsymbol{\phi}$

$$\mathbf{S}_2(\boldsymbol{\phi}; \boldsymbol{\gamma}) = \sum_{i=1}^{n} \left[\delta_i \mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_i) + (1-\delta_i) E_0 \left\{\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y) \mid \mathbf{x}_i; \boldsymbol{\phi}, \boldsymbol{\gamma}\right\}\right], \tag{A.7}$$

where $\mathbf{s}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, y)$ is defined in (A.3),

$$E_0\{g(y) \mid \mathbf{x}; \boldsymbol{\gamma}, \boldsymbol{\phi}\} = \frac{\int g(y) f_1(y \mid \mathbf{x}; \boldsymbol{\gamma}) O(\mathbf{x}, y; \boldsymbol{\phi}) dy}{\int f_1(y \mid \mathbf{x}; \boldsymbol{\gamma}) O(\mathbf{x}, y; \boldsymbol{\phi}) dy},$$

$O(\mathbf{x}, y; \boldsymbol{\phi}) = 1/\pi(\mathbf{x}, y; \boldsymbol{\phi}) - 1.$

– Score function for $\boldsymbol{\gamma}$

$$\mathbf{S}_1(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \delta_i \mathbf{s}_1(\boldsymbol{\gamma}; \mathbf{x}_i, y_i) = \sum_{i=1}^{n} \delta_i \left\{\frac{\partial}{\partial \boldsymbol{\gamma}} \log f_1(y_i \mid \mathbf{x}_i; \boldsymbol{\gamma})\right\} \tag{A.8}$$

– Let $\mathbf{I}(\boldsymbol{\phi}, \boldsymbol{\gamma})$ be defined as follows:

$$\begin{aligned}
\mathbf{I}(\boldsymbol{\phi}, \boldsymbol{\gamma}) &= \begin{bmatrix} \mathbf{I}_{11}(\boldsymbol{\phi}, \boldsymbol{\gamma}) & \mathbf{I}_{12}(\boldsymbol{\phi}, \boldsymbol{\gamma}) \\ \mathbf{I}_{21}(\boldsymbol{\phi}, \boldsymbol{\gamma}) & \mathbf{I}_{22}(\boldsymbol{\phi}, \boldsymbol{\gamma}) \end{bmatrix} \\
&= -\begin{bmatrix} \partial \mathbf{s}_1(\boldsymbol{\gamma})/\partial \boldsymbol{\gamma}^T & \partial \mathbf{s}_1(\boldsymbol{\gamma})/\partial \boldsymbol{\phi}^T \\ \partial \mathbf{s}_2(\boldsymbol{\phi}; \boldsymbol{\gamma})/\partial \boldsymbol{\gamma}^T & \partial \mathbf{s}_2(\boldsymbol{\phi}; \boldsymbol{\gamma})/\partial \boldsymbol{\phi}^T \end{bmatrix} \\
&= -\begin{bmatrix} \partial \mathbf{s}_1(\boldsymbol{\gamma})/\partial \boldsymbol{\gamma}^T & \mathbf{0} \\ \partial \mathbf{s}_2(\boldsymbol{\phi}; \boldsymbol{\gamma})/\partial \boldsymbol{\gamma}^T & \partial \mathbf{s}_2(\boldsymbol{\phi}; \boldsymbol{\gamma})/\partial \boldsymbol{\phi}^T \end{bmatrix}
\end{aligned} \tag{A.9}$$

– Let $\mathcal{I}(\boldsymbol{\phi}, \boldsymbol{\gamma})$ be the expectation of $\mathbf{I}(\boldsymbol{\phi}; \boldsymbol{\gamma})$ defined in (A.8), i.e.,

$$E\left[\mathbf{I}(\boldsymbol{\phi}, \boldsymbol{\gamma})\right] \triangleq \mathcal{I}(\boldsymbol{\phi}, \boldsymbol{\gamma}) = \begin{bmatrix} \mathcal{I}_{11}(\boldsymbol{\phi}, \boldsymbol{\gamma}) & \mathcal{I}_{12}(\boldsymbol{\phi}, \boldsymbol{\gamma}) \\ \mathcal{I}_{21}(\boldsymbol{\phi}, \boldsymbol{\gamma}) & \mathcal{I}_{22}(\boldsymbol{\phi}, \boldsymbol{\gamma}) \end{bmatrix}. \tag{A.10}$$

- Regularity conditions

  Assume that the study variable has a common support for both respondents and nonrespondents. To discuss asymptotic properties of the PS estimator, which is the solution to (3.15), we assume that the response mechanism satisfies the following regularity conditions:

  **(A1)** The true response probability follows a parametric model in (3.2). The response probability function $\pi(\mathbf{x}_1, y; \boldsymbol{\phi})$ is continuous in $\boldsymbol{\phi}$ with continuous first and second derivatives in an open set containing $\boldsymbol{\phi}_0$ as an interior point.

  **(A2)** The responses are independent. That is, $Cov(\delta_i, \delta_j) = 0$ for $i \neq j$.

  **(A3)** The response probability is bounded below. That is, $\pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi}_0) > K_\pi$ for some $K_\pi > 0$ for all $i = 1, 2, \cdots, n$, uniformly in $n$.

  **(A4)** Under complete observation of $(\mathbf{x}_i, y_i)$ for $i = 1, 2, \cdots, n$, the unique solution $\hat{\boldsymbol{\phi}}_n$ to the score equation in (3.3) satisfies

  $$\sqrt{n}\left(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0\right) \to N(\mathbf{0}, \boldsymbol{\Sigma})$$

  for some $\boldsymbol{\Sigma}$ for sufficiently large $n$.

  Note that the condition (A4) can be broken down into the following three conditions

  **(A4.1)** There is a function $K(\mathbf{x}_1, y)$ such that $E\{K(\mathbf{x}_1, y)\} < \infty$ and $\mathbf{z}(\mathbf{x}_1, y; \boldsymbol{\phi})$ is bounded by $K(\mathbf{x}_1, y)$. The second moment of $\mathbf{z}(\boldsymbol{\phi})$ and the first moment of $\partial \mathbf{z}(\boldsymbol{\phi})/\partial \boldsymbol{\phi}$ exist.

  **(A4.2)** $E[\pi(\mathbf{x}_1, y; \boldsymbol{\phi}_0)\{1 - \pi(\mathbf{x}_1, y; \boldsymbol{\phi}_0)\}\mathbf{z}(\mathbf{x}_1, y; \boldsymbol{\phi}_0)^{\otimes 2}]$ exists and is nonsingular.

  **(A4.3)** $\pi(\mathbf{x}, y; \boldsymbol{\phi}) = \pi(\mathbf{x}, y; \boldsymbol{\phi}_0)$ if and only if $\boldsymbol{\phi} = \boldsymbol{\phi}_0$.

  Here, condition (A4) can be established by (A1)-(A3) and (A4.1)-(A4.3) However, there are different sets of conditions that guarantee the (A4) condition.

We also need a set of regularity conditions for the model for conditional distribution for respondents, $f_1(y|\mathbf{x};\boldsymbol{\gamma})$ in addition to the regularity conditions for the response mechanism. The conditions are listed as follows:

**(B1)** The independent random variables $y_i$ if $\delta_i = 1$ have common distribution $f_1(y|\mathbf{x};\boldsymbol{\gamma}_0)$. The distribution function $f_1(y|\mathbf{x};\boldsymbol{\gamma})$ is continuous in $\boldsymbol{\gamma}$ with continuous first and second derivatives in an open set containing $\boldsymbol{\gamma}_0$ as an interior point.

**(B2)** $\left|\dfrac{\partial^2 \log f_1(y|\mathbf{x};\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^T}\right|$ and $\left|\dfrac{\partial \log f_1(y|\mathbf{x};\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}}\dfrac{\partial \log f_1(y|\mathbf{x};\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}^T}\right|$ are dominated by functions that are integrable with respect to the density $f_1(y|\mathbf{x};\boldsymbol{\gamma}_0)$ for all $\mathbf{x}$ and $y$ and all $\boldsymbol{\gamma}$.

**(B3)** $E\left[\dfrac{\partial^2 \log f_1(y|\mathbf{x};\boldsymbol{\gamma}_0)}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}^T}\right]$ is nonsingular.

With the conditions from (B1) through (B3) holding, the solution $\hat{\boldsymbol{\gamma}}$ to (3.9) satisfies

$$\sqrt{n}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \to N(\mathbf{0}, \boldsymbol{\Sigma}_\gamma),$$

for some $\boldsymbol{\Sigma}_\gamma$ for sufficiently large $n$.

The last set of conditions allow to exchange the order of differentiation and integration. Let $v(\boldsymbol{\phi},\boldsymbol{\gamma},y;\mathbf{x}) = \mathbf{s}(\boldsymbol{\phi};\mathbf{x},y)f_1(y|\mathbf{x};\boldsymbol{\gamma})O(\mathbf{x},y;\boldsymbol{\phi})$

**(C1)** $v(\boldsymbol{\phi},\boldsymbol{\gamma},y;\mathbf{x})$ is integrable with respect to $y$ for all $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$.

**(C2)** The derivatives of $v(\boldsymbol{\phi},\boldsymbol{\gamma},y;\mathbf{x})$ with respect to $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$ exist for almost every $y$ for all $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$

**(C3)** There exist $K_v(y)$, where $|v(\boldsymbol{\phi},\boldsymbol{\gamma},y;\mathbf{x})| < K_v(y)$ for all $\boldsymbol{\phi}$ and $\boldsymbol{\gamma}$.

**Lemma A.1.** *The partial derivatives of $U_{PS}(\theta;\boldsymbol{\phi})$ in (A.2) with respect to $\boldsymbol{\phi}$ satisfy*

$$E\left\{\frac{\partial u_{PS}(\theta;\boldsymbol{\phi})}{\partial\boldsymbol{\phi}}\right\} = -Cov\left\{u_{PS}(\theta;\boldsymbol{\phi}), \mathbf{s}(\boldsymbol{\phi})\right\}. \tag{A.11}$$

*Proof.* Since

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\phi}}\left\{\frac{1}{\pi(\mathbf{x}_1,y;\boldsymbol{\phi})}\right\} &= -\frac{1}{\pi^2(\mathbf{x}_1,y;\boldsymbol{\phi})}\frac{\partial \pi(\mathbf{x}_1,y;\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} \\
&= -\frac{\pi(\mathbf{x}_1,y;\boldsymbol{\phi})\{1-\pi(\mathbf{x}_1,y;\boldsymbol{\phi})\}}{\pi^2(\mathbf{x}_1,y;\boldsymbol{\phi})}\mathbf{z}(\mathbf{x}_1,y;\boldsymbol{\phi}) \\
&= -\frac{1-\pi(\mathbf{x}_1,y;\boldsymbol{\phi})}{\pi(\mathbf{x}_1,y;\boldsymbol{\phi})}\mathbf{z}(\mathbf{x}_1,y;\boldsymbol{\phi}),
\end{aligned}$$

where $\mathbf{z}(\mathbf{x}_1,y;\boldsymbol{\phi})$ is defined in (A.6).

$$\begin{aligned}
& E\left\{\frac{\partial u_{PS}(\theta;\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}\right\} \\
&= -E\left[\frac{\delta}{\pi(\mathbf{x}_1,y;\boldsymbol{\phi})}\{1-\pi(\mathbf{x}_1,y;\boldsymbol{\phi})\}\mathbf{z}(\mathbf{x}_1,y;\boldsymbol{\phi})u(\theta;\mathbf{x},y)\right] \\
&= -E\left[\frac{\delta}{\pi(\mathbf{x}_1,y;\boldsymbol{\phi})}u(\theta;\mathbf{x},y)\cdot\{\delta-\pi(\mathbf{x}_1,y;\boldsymbol{\phi})\}\mathbf{z}(\mathbf{x}_1,y;\boldsymbol{\phi})\right] \\
&= -Cov\left\{u_{PS}(\theta;\boldsymbol{\phi}),\mathbf{s}(\boldsymbol{\phi})\right\}.
\end{aligned}$$

The last equation holds since $E\{\mathbf{s}(\boldsymbol{\phi})\}=\mathbf{0}$. □

**Lemma A.2.** *The partial derivatives of $\mathbf{s}_2(\boldsymbol{\phi};\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$ satisfy*

$$\mathcal{I}_{21}(\boldsymbol{\phi};\boldsymbol{\gamma}) = -E\left[(1-\delta)Cov_0\left\{\mathbf{s}(\boldsymbol{\phi};\delta,\mathbf{x}_1,Y),\mathbf{s}_1(\boldsymbol{\gamma};\mathbf{x}_1,Y)\mid\mathbf{x}\right\}\right], \tag{A.12}$$

*where $Cov_0(\cdot|\mathbf{x}) = Cov(\cdot|\delta=0,\mathbf{x})$.*

*Proof.* Since $\mathbf{s}_1(\boldsymbol{\gamma};\mathbf{x},y) = \partial\log f_1(y\mid\mathbf{x};\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}$, the partial derivative of $\mathbf{s}_2(\boldsymbol{\phi};\boldsymbol{\gamma})$ in (A.4) with respect to $\boldsymbol{\gamma}$ can be written as

$$\begin{aligned}
\frac{\partial \mathbf{s}_2(\boldsymbol{\phi};\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}^T} &= (1-\delta)E_0\left\{\mathbf{s}(\boldsymbol{\phi};\delta,\mathbf{x}_1,Y)\mathbf{s}_1^T(\boldsymbol{\gamma};\mathbf{x},Y)\mid\mathbf{x}\right\} \\
&\quad -(1-\delta)E_0\left\{\mathbf{s}(\boldsymbol{\phi};\delta,\mathbf{x}_1,Y)\mid\mathbf{x}\right\}E_0\left\{\mathbf{s}_1^T(\boldsymbol{\gamma};\mathbf{x},Y)\mid\mathbf{x}\right\} \\
&= (1-\delta)Cov_0\left\{\mathbf{s}(\boldsymbol{\phi};\delta,\mathbf{x}_1,Y),\mathbf{s}_1(\boldsymbol{\gamma};\mathbf{x},Y)\mid\mathbf{x}\right\}.
\end{aligned}$$

□

**Lemma A.3.** *The partial derivatives of $\mathbf{s}_2(\boldsymbol{\phi};\boldsymbol{\gamma})$ with respect to $\boldsymbol{\phi}$ satisfy*

$$\mathcal{I}_{22}(\boldsymbol{\phi};\boldsymbol{\gamma}) = E\left\{(1-\delta)\bar{\mathbf{s}}_0(\boldsymbol{\phi};\mathbf{x})\mathbf{s}^T(\boldsymbol{\phi};\delta,\mathbf{x}_1,Y)/\pi(\boldsymbol{\phi};\mathbf{x}_1,Y)\right\} \tag{A.13}$$

*Proof.* Since $\partial O(\mathbf{x}_1, y; \boldsymbol{\phi})/\partial \boldsymbol{\phi} = -O(\mathbf{x}_1, y; \boldsymbol{\phi})\mathbf{z}(\mathbf{x}_1, y; \boldsymbol{\phi})$, the partial derivatives of $\mathbf{s}_2(\boldsymbol{\phi}; \boldsymbol{\gamma})$ in (A.4) with respect to $\boldsymbol{\phi}$ can be written as

$$\begin{aligned}\frac{\partial \mathbf{s}_2(\boldsymbol{\phi}; \boldsymbol{\gamma})}{\partial \boldsymbol{\phi}^T} &= [\delta \dot{\mathbf{s}}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, y) + (1-\delta)E_0\left\{\dot{\mathbf{s}}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, Y) \mid \mathbf{x}\right\}] \\ &\quad -(1-\delta)E_0\left\{\mathbf{s}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, Y)\mathbf{z}^T(\mathbf{x}_1, Y; \boldsymbol{\phi}) \mid \mathbf{x}\right\} \\ &\quad +(1-\delta)E_0\left\{\mathbf{s}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, Y) \mid \mathbf{x}\right\} E_0\left\{\mathbf{z}^T(\mathbf{x}_1, Y; \boldsymbol{\phi}) \mid \mathbf{x}\right\},\end{aligned}$$

where $\dot{\mathbf{s}}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, y) = \partial \mathbf{s}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, y)/\partial \boldsymbol{\phi}^T$. Since

$$\begin{aligned}& E[\delta \dot{\mathbf{s}}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, y) + (1-\delta)E_0\left\{\dot{\mathbf{s}}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, Y) \mid \mathbf{x}\right\}] \\ =\; & E[\pi(\boldsymbol{\phi}; \mathbf{x}_1, y)\dot{\mathbf{s}}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, y) + E\left\{(1-\delta)\dot{\mathbf{s}}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, Y) \mid \mathbf{x}, \delta\right\}] \\ =\; & E[\pi(\boldsymbol{\phi}; \mathbf{x}_1, y)\dot{\mathbf{s}}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, y) + \{1-\pi(\boldsymbol{\phi}; \mathbf{x}_1, y)\}\dot{\mathbf{s}}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, y)] \\ =\; & E[\dot{\mathbf{s}}(\boldsymbol{\phi}; \mathbf{x}_1, y)],\end{aligned}$$

and

$$\begin{aligned}& E[(1-\delta)E_0\left\{\mathbf{s}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, Y)\mathbf{z}^T(\mathbf{x}_1, Y; \boldsymbol{\phi}) \mid \mathbf{x}\right\}] \\ =\; & E[E\left\{(1-\delta)\mathbf{s}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, Y)\mathbf{z}^T(\mathbf{x}_1, Y; \boldsymbol{\phi}) \mid \mathbf{x}, \delta\right\}] \\ =\; & E[E\left\{(1-\delta)\{\delta - \pi(\boldsymbol{\phi}; \mathbf{x}_1, Y)\}\mathbf{z}(\mathbf{x}_1, Y; \boldsymbol{\phi})\mathbf{z}^T(\mathbf{x}_1, Y; \boldsymbol{\phi}) \mid \mathbf{x}, \delta\right\}] \\ =\; & -E\left\{(1-\delta)\pi(\boldsymbol{\phi}; \mathbf{x}_1, y)\mathbf{z}(\mathbf{x}_1, y; \boldsymbol{\phi})\mathbf{z}^T(\mathbf{x}_1, y; \boldsymbol{\phi})\right\} \\ =\; & -E[\{1-\pi(\boldsymbol{\phi}; \mathbf{x}_1, y)\}\pi(\boldsymbol{\phi}; \mathbf{x}_1, y)\mathbf{z}(\mathbf{x}_1, y; \boldsymbol{\phi})\mathbf{z}^T(\mathbf{x}_1, y; \boldsymbol{\phi})] \\ =\; & E[\dot{\mathbf{s}}(\boldsymbol{\phi}; \mathbf{x}_1, y)]\end{aligned}$$

hold, the following equation holds

$$E\left\{\frac{\partial \mathbf{s}_2(\boldsymbol{\phi}; \boldsymbol{\gamma})}{\partial \boldsymbol{\phi}^T}\right\} = E[(1-\delta)E_0\left\{\mathbf{s}(\boldsymbol{\phi}; \delta, \mathbf{x}_1, Y) \mid \mathbf{x}\right\} E_0\left\{\mathbf{z}^T(\mathbf{x}_1, Y; \boldsymbol{\phi}) \mid \mathbf{x}\right\}].$$

Since $\mathbf{s}(\boldsymbol{\phi};\delta,\mathbf{x}_1,y) = -\pi(\boldsymbol{\phi};\mathbf{x}_1,y)\mathbf{z}(\mathbf{x},Y;\boldsymbol{\phi})$ when $\delta=0$,

$$\begin{aligned} & E[(1-\delta)E_0\left\{\mathbf{s}(\boldsymbol{\phi};\delta,\mathbf{x}_1,Y)\mid\mathbf{x}\right\}E_0\left\{\mathbf{z}^T(\mathbf{x}_1,Y;\boldsymbol{\phi})\mid\mathbf{x}\right\}] \\ = \; & E[(1-\delta)\bar{\mathbf{s}}_0(\boldsymbol{\phi};\mathbf{x})E\left\{(1-\delta)\mathbf{z}^T(\mathbf{x}_1,Y;\boldsymbol{\phi})\mid\mathbf{x},\delta\right\}] \\ = \; & -E[E\left\{(1-\delta)\bar{\mathbf{s}}_0(\boldsymbol{\phi};\mathbf{x})\mathbf{s}^T(\boldsymbol{\phi};\delta,\mathbf{x}_1,Y)/\pi(\boldsymbol{\phi};\mathbf{x}_1,Y)\mid\mathbf{x},\delta\right\}] \\ = \; & -E\left\{(1-\delta)\bar{\mathbf{s}}_0(\boldsymbol{\phi};\mathbf{x})\mathbf{s}^T(\boldsymbol{\phi};\delta,\mathbf{x}_1,Y)/\pi(\boldsymbol{\phi};\mathbf{x}_1,Y)\right\}, \end{aligned}$$

where $\bar{\mathbf{s}}_0(\boldsymbol{\phi};\mathbf{x}) = E_0\{\mathbf{s}(\boldsymbol{\phi};\delta,\mathbf{x}_1,Y)|\mathbf{x}\}$. Thus, (A.13) holds. □

## A.1 Proof of Lemma 3.1.

*Proof.* The logarithm of the observed likelihood $L_{obs}(\boldsymbol{\phi})$ in (3.4) can be written as

$$\log L_{obs}(\boldsymbol{\phi}) = \sum_{i=1}^{n}\left[\delta_i\log\pi(\mathbf{x}_{1i},y_i;\boldsymbol{\phi}) + (1-\delta_i)\log\int\{1-\pi(\mathbf{x}_{1i},y;\boldsymbol{\phi})\}f(y|\mathbf{x}_i)dy\right].$$

Since

$$\partial\pi(\mathbf{x}_1,y;\boldsymbol{\phi})/\partial\boldsymbol{\phi} = \pi(\mathbf{x}_1,y;\boldsymbol{\phi})\{1-\pi(\mathbf{x}_1,y;\boldsymbol{\phi})\}\mathbf{z}(\mathbf{x}_1,y;\boldsymbol{\phi}),$$

where $\mathbf{z}(\mathbf{x}_1,y;\boldsymbol{\phi}) = \partial\text{logit}\{\pi(\mathbf{x}_1,y;\boldsymbol{\phi})\}/\partial\boldsymbol{\phi}$, the observed score function $\mathbf{S}_{obs}(\boldsymbol{\phi}) = \partial\log L_{obs}(\boldsymbol{\phi})/\partial\boldsymbol{\phi}$, is as follows:

$$\begin{aligned} \mathbf{S}_{obs}(\boldsymbol{\phi}) &= \partial\log L_{obs}(\boldsymbol{\phi})/\partial\boldsymbol{\phi} \\ &= \sum_{i=1}^{n}\delta_i\frac{\pi(\mathbf{x}_{1i},y_i;\boldsymbol{\phi})\{1-\pi(\mathbf{x}_{1i},y_i;\boldsymbol{\phi})\}}{\pi(\mathbf{x}_{1i},y_i;\boldsymbol{\phi})}\mathbf{z}(\mathbf{x}_{1i},y_i;\boldsymbol{\phi}) \\ &\quad -\sum_{1=1}^{n}(1-\delta_i)\frac{\int\pi(\mathbf{x}_{1i},y;\boldsymbol{\phi})\{1-\pi(\mathbf{x}_{1i},y;\boldsymbol{\phi})\}\mathbf{z}(\mathbf{x}_{1i},y;\boldsymbol{\phi})f(y|\mathbf{x}_i)dy}{\int\{1-\pi(\mathbf{x}_{1i},y;\boldsymbol{\phi})\}f(y|\mathbf{x}_i)dy} \end{aligned} \tag{A.14}$$

Since

$$f(y|\mathbf{x},\delta=0) = \frac{\{1-\pi(\mathbf{x}_1,y;\boldsymbol{\phi})\}f(y|\mathbf{x})}{\int\{1-\pi(\mathbf{x}_1,y;\boldsymbol{\phi})\}f(y|\mathbf{x})dy},$$

the equation in (A.14) can be rewritten as

$$
\begin{aligned}
& \mathbf{S}_{obs}(\boldsymbol{\phi}) \\
= & \sum_{i=1}^{n} [\delta_i\{1-\pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})\}\mathbf{z}(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi}) - (1-\delta_i)E_0\{\pi(\mathbf{x}_{1i}, Y; \boldsymbol{\phi})\mathbf{z}(\mathbf{x}_{1i}, Y; \boldsymbol{\phi})|\mathbf{x}_i\}] \\
= & \sum_{i=1}^{n} [\delta_i\{\delta_i-\pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})\}\mathbf{z}(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi}) + (1-\delta_i)E_0[\{\delta_i-\pi(\mathbf{x}_{1i}, Y; \boldsymbol{\phi})\}\mathbf{z}(\mathbf{x}_{1i}, Y; \boldsymbol{\phi})|\mathbf{x}_i]] \\
= & \sum_{i=1}^{n} [\delta_i \mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_i) + (1-\delta_i)E_0\{\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, Y)|\mathbf{x}_i\}] \\
= & \bar{\mathbf{S}}(\boldsymbol{\phi}),
\end{aligned}
$$

where $E_0(\cdot|\mathbf{x}) = E(\cdot|\mathbf{x}, \delta = 0)$. □

## A.2 Proof of Theorem 3.1.

*Proof.* Since $(\hat{\boldsymbol{\gamma}}^T, \hat{\boldsymbol{\phi}}_p^T)^T$ is the solution to

$$
\begin{bmatrix} \mathbf{S}_1(\boldsymbol{\gamma}) \\ \mathbf{S}_2(\boldsymbol{\phi}; \boldsymbol{\gamma}) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix},
$$

where $\mathbf{S}_1(\boldsymbol{\gamma})$ and $\mathbf{S}_2(\boldsymbol{\phi}; \boldsymbol{\gamma})$, are defined in (A.8) and (A.7), respectively, we can obtain variance of $(\hat{\boldsymbol{\gamma}}^T, \hat{\boldsymbol{\phi}}_p^T)^T$ by using sandwich formula as follows

$$
V\left[\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 \\ \hat{\boldsymbol{\phi}}_p - \boldsymbol{\phi}_0 \end{pmatrix}\right] = \mathcal{I}^{-1} V \begin{bmatrix} \mathbf{s}_1(\boldsymbol{\gamma}_0) \\ \mathbf{s}_2(\boldsymbol{\phi}_0; \boldsymbol{\gamma}_0) \end{bmatrix} \left(\mathcal{I}^{-1}\right)^T, \tag{A.15}
$$

where $\mathcal{I} = \mathcal{I}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}_0)$ and $\mathcal{I}(\boldsymbol{\phi}, \boldsymbol{\gamma})$ is defined in (A.10). Using

$$
\mathcal{I}^{-1} = \begin{bmatrix} \mathcal{I}_{11}^{-1} & \mathbf{0} \\ -\mathcal{I}_{22}^{-1}\mathcal{I}_{21}\mathcal{I}_{11}^{-1} & \mathcal{I}_{22}^{-1} \end{bmatrix}, \tag{A.16}
$$

where $\mathcal{I}_{11} = \mathcal{I}_{11}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}_0)$, $\mathcal{I}_{12} = \mathcal{I}_{12}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}_0)$, and $\mathcal{I}_{22} = \mathcal{I}_{22}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}_0)$, the variance of $\hat{\boldsymbol{\phi}}_p$ can be derived as

$$
V\left(\hat{\boldsymbol{\phi}}_p\right) = V\left[\mathcal{I}_{22}^{-1}\left\{\mathbf{s}_2(\boldsymbol{\phi}_0; \boldsymbol{\gamma}_0) - \boldsymbol{\kappa}\mathbf{s}_1(\boldsymbol{\gamma}_0)\right\}\right],
$$

where $\boldsymbol{\kappa} = \mathcal{I}_{21}\mathcal{I}_{11}^{-1}$. The result in (3.16) holds by Lemma A.2. and Lemma A.3.

We now prove the second part of Theorem 1. Since $\hat{\boldsymbol{\eta}}^T = (\hat{\boldsymbol{\gamma}}^T, \hat{\boldsymbol{\phi}}_p^T, \hat{\theta}_{PS,p})^T$ is the solution to

$$\begin{bmatrix} \mathbf{S}_1(\boldsymbol{\gamma}) \\ \mathbf{S}_2(\boldsymbol{\phi};\boldsymbol{\gamma}) \\ U_{PS}(\theta;\boldsymbol{\phi}) \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ 0 \end{bmatrix},$$

where $\hat{\theta}_{PS,p}$, $\mathbf{S}_1(\boldsymbol{\gamma})$, $\mathbf{S}_2(\boldsymbol{\phi};\boldsymbol{\gamma})$, and $U_{PS}(\theta;\boldsymbol{\phi})$ are defined in (A.1), (A.8), (A.7), and (A.2), respectively, we can obtain variance of $\hat{\boldsymbol{\eta}}$ by using sandwich formula as follows

$$V\left[\sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 \\ \hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0 \\ \hat{\theta}_{PS,p} - \theta \end{pmatrix}\right] = \mathcal{T}^{-1} V \begin{bmatrix} \mathbf{s}_1(\boldsymbol{\gamma}_0) \\ \mathbf{s}_2(\boldsymbol{\phi}_0;\boldsymbol{\gamma}_0) \\ u_{PS}(\theta;\boldsymbol{\phi}) \end{bmatrix} \left(\mathcal{T}^{-1}\right)^T, \tag{A.17}$$

where $u_{PS}(\theta;\boldsymbol{\phi}) = \{\delta/\pi(\boldsymbol{\phi};\mathbf{x},y) - 1\}u(\theta;\mathbf{x},y)$ and

$$\mathcal{T} = E\begin{bmatrix} \partial\mathbf{s}_1(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^T & \partial\mathbf{s}_1(\boldsymbol{\gamma})/\partial\boldsymbol{\phi}^T & \partial\mathbf{s}_1(\boldsymbol{\gamma})/\partial\theta \\ \partial\mathbf{s}_2(\boldsymbol{\phi};\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}^T & \partial\mathbf{s}_2(\boldsymbol{\phi};\boldsymbol{\gamma})/\partial\boldsymbol{\phi}^T & \partial\mathbf{s}_2(\boldsymbol{\phi};\boldsymbol{\gamma})/\partial\theta \\ \partial u_{PS}(\theta;\boldsymbol{\phi})/\partial\boldsymbol{\gamma}^T & \partial u_{PS}(\theta;\boldsymbol{\phi})/\partial\boldsymbol{\phi}^T & \partial u_{PS}(\theta;\boldsymbol{\phi})/\partial\theta^T \end{bmatrix},$$

which can be simplified as

$$\mathcal{T} = \begin{bmatrix} \mathcal{I} & \mathbf{0} \\ \boldsymbol{\upsilon} & \boldsymbol{\tau}, \end{bmatrix}$$

where $\mathcal{I}$ is defined in (A.15), $\boldsymbol{\upsilon} = [\boldsymbol{\upsilon}_1 \;\; \boldsymbol{\upsilon}_2] = E[\partial u_{PS}(\theta;\boldsymbol{\phi})/\partial\boldsymbol{\gamma}^T \;\; \partial u_{PS}(\theta;\boldsymbol{\phi})/\partial\boldsymbol{\phi}^T]$, and $\boldsymbol{\tau} = E[\partial u_{PS}(\theta;\boldsymbol{\phi})/\partial\theta]$. Using

$$\mathcal{T}^{-1} = \begin{bmatrix} \mathcal{I}^{-1} & \mathbf{0} \\ -\boldsymbol{\tau}^{-1}\boldsymbol{\upsilon}\mathcal{I}^{-1} & \boldsymbol{\tau}^{-1}, \end{bmatrix}$$

$\boldsymbol{\upsilon}_1 = \mathbf{0}$, and $\mathcal{I}^{-1}$ in (A.16), the right hand side of the equation in (A.17) can be rewritten as

$$V\begin{bmatrix} \mathcal{I}_{11}^{-1}\mathbf{s}_1(\boldsymbol{\gamma}_0) \\ \mathcal{I}_{22}^{-1}\left\{\mathbf{s}_2(\boldsymbol{\phi}_0;\boldsymbol{\gamma}_0) - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathbf{s}_1(\boldsymbol{\gamma}_0)\right\} \\ \boldsymbol{\tau}^{-1}\left[U_{PS}(\theta_0;\boldsymbol{\phi}_0) - \boldsymbol{\upsilon}_2\mathcal{I}_{22}^{-1}\left\{\mathbf{s}_2(\boldsymbol{\phi}_0;\boldsymbol{\gamma}_0) - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathbf{s}_1(\boldsymbol{\gamma}_0)\right\}\right] \end{bmatrix}. \tag{A.18}$$

From (A.18), the variance of the PS estimator can be derived as

$$V\left\{\hat{\theta}_{PS,p}\right\} = \boldsymbol{\tau}^{-1} V\left[U_{PS}\left(\theta_0; \boldsymbol{\phi}_0\right) - \mathbf{B}\left\{\mathbf{s}_2(\boldsymbol{\phi}_0; \boldsymbol{\gamma}_0)\right) - \boldsymbol{\kappa}\mathbf{s}_1(\boldsymbol{\gamma}_0)\}\right]\left(\boldsymbol{\tau}^{-1}\right)^T$$

where $\boldsymbol{\tau} = E\{\partial u_{PS}/\partial\theta^T\}$, and $\mathbf{B}$ and $\boldsymbol{\kappa}$ are defined in (3.17), by Lemma A.1-Lemma A.3. $\square$

# BIBLIOGRAPHY

Baker, S. G. and Laird, N. M. (1988). Regression analysis for categorical variables with outcome subject to nonignorable nonresponse. *Journal of the American Statistical Association*, 83:62–69.

Bethlehem, J. G. (1988). Reduction in nonresponse bias through regression estimation. *Journal of Official Statistics*, 4:251–260.

Cao, W., Tsiatis, A., and Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, 96:723–734.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34.

Chang, T. and Kott, P. (2008). Using calibration weighting to adjust for nonresponse under a plausible model. *Biometrika*, 95:555–571.

Da Silva, D. N. and Opsomer, J. D. (2006). A kernel smoothing method to adjust for unit nonresponse in sample surveys. *Canadian Journal of Statistics*, 34:563–579.

Da Silva, D. N. and Opsomer, J. D. (2009). Nonparametric propensity weighting for survey nonresponse through local polynomial regression. *SURVEY METHODOLOGY*, 35(2):165–176.

Deville, J. C., Sarndal, C. E., and Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88:1013–1020.

Duncan, K. B. and Stasny, E. A. (2001). Using propensity scores to control coverages bias in telephone surveys. *Survey Methodology*, 27:121–130.

Durrant, G. B. and Skinner, C. (2006). Using missing data methods to correct for measurement error in a distribution function. *Survey Methodology*, 32:25–36.

Ekholm, A. and Laaksonen, S. (1991). Weighting via response modeling in the Finnish Household Budget Survey. *Journal of Official Statistics*, 7:325–337.

Fay, R. E. (1992). When are inferences from multiple imputation valid? In *Proceedings of the Survey Research Methods Section*, pages 227–232. American Statistical Association.

Folsom, R. E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. In *Proceedings of the Social Statistics Section*, pages 197–202. American Statistical Association.

Fuller, W. A., Loughin, M. M., and Baker, H. D. (1994). Regression weighting for the 1987-1988 nationwide food consumption survey. *Survey Methodology*, 20:75–85.

Greenlees, J., Reece, W., and Zieschang, K. (1982). Imputation of missing values when the probability of response depends on the variable being imputed. *Journal of the American Statistical Association*, 77:251–261.

Groves, R. M., Presser, S., and Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly*, 68:2–31.

Holt, D. and Smith, T. M. F. (1979). Post-stratification. *Journal of the Royal Statistical Society, Series A*, 142:33–46.

Iannacchione, V. G., Milne, J. G., and Folsom, R. (1991). Response probability weight adjustments using logistic regression. In *Proceedings of the Survey Research Methods Section*, pages 637–642. American Statistical Association.

Ibrahim, J. G., Lipsitz, S. R., and Chen, M.-H. (1999). Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society, Series B*, 61:173–190.

Isaki, C. and Fuller, W. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77:89–96.

Kenward, M. G. and Molenberghs, G. (1988). Likelihood based frequentist inference when data are missing at random. *Statistical Science*, 13:236–247.

Kim, J. K. (2004). Finite sample properties of multiple imputation estimators. *Annals of Statistics*, 32:766–783.

Kim, J. K. (2011). Parametric fractional imputation for missing data analysis. *Biometrika*, 98:119–132.

Kim, J. K. and Kim, J. J. (2007). Nonresponse weighting adjustment using estimated response probability. *Canadian Journal of Statistics*, 35:501–514.

Kim, J. K., Navarro, A., and Fuller, W. A. (2006). Replicate variance estimation after multi-phase stratified sampling. *Journal of the American Statistical Association*, 101:312–320.

Kim, J. K. and Rao, J. N. K. (2009). Unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika*, 96:917–932.

Kim, J. K. and Riddles, M. K. (2012). Some theory for propensity scoring adjustment estimator. *Survey Methodology*, 38:157–165.

Kim, J. K. and Skinner, C. (2013). Weighting in survey analysis under informative sampling. *Biometrika*, in press (ass085v1-ass085).

Kim, J. K. and Yu, C. L. (2011). A semi-parametric estimation of mean functionals with non-ignorable missing data. *Journal of the American Statistical Association*, 106:157–165.

Kott, P. S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, 32:133–142.

Kott, P. S. and Chang, T. (2010). Using calibration weighting to adjust for nonignorable unit nonresponse. *Journal of the American Statistical Association*, 105:1265–1275.

Lee, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *Journal of Official Statistics*, 22:329–349.

Lindström, S. and Särndal, C.-E. (1999). Calibration as a standard method for treatment of nonresponse. *Journal of Official Statistics*, 15:305–327.

Little, R. J. A. (1988). Missing-data adjustments in large surveys. *Journal of Business and Economic Statistics*, 6:287–296.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. John Wiley & Sons, London, 2 edition.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 44:226–233.

Nevo, A. (2003). Using weights to adjust for sample selection when auxiliary information is available. *Journal of Business and Economic Statistics*, 21:43 – 52.

Pfeffermann, D. and Sverchkov, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhyā, Series B*, 61:166–186.

Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *Annals of Statistics*, 10:462–474.

Rizzo, L., Kalton, G., and Brick, J. M. (1996). A comparison of some weighting adjustment methods for panel nonresponse. *Survey Methodology*, 22:43–53.

Robins, J., Rotnitzky, A., and Scharfstein, D. O. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89:846–866.

Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association*, 82:387–394.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63:581–590.

Shao, J. and Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association*, 94:254–265.

Singh, A. C. and Folsom, R. E. (2000). Bias corrected estimating function approach for variance estimation adjusted for poststratification. In *Proceedings of the Survey Research Methods Section*, pages 610–615. American Statistical Association.

Tan, Z. (2006). A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association*, 101:1619–1637.

Wang, S., Shao, J., and Kim, J. K. (2013). An instrument variable approach for identification and estimation with nonignorable nonresponse. Manuscript submitted for publication.

Wu, C. and Sitter, R. R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, 96:185–193.

Zhou, M. and Kim, J. K. (2012). An efficient method of estimation for longitudinal surveys with monotone missing data. *Biometrika*, 99:631–648.