M.S. THESIS

Doubly Robustness and Efficiency of Weighted Estimators incorporating Cohort Data

코호트 데이터를 활용한 가중 추정량의 이중견고성과 효율성

February 2025

DEPARTMENT OF STATISTICS
COLLEGE OF NATURAL SCIENCES
SEOUL NATIONAL UNIVERSITY

Taeyeon Kim

M.S. THESIS

Doubly Robustness and Efficiency of Weighted Estimators incorporating Cohort Data

코호트 데이터를 활용한 가중 추정량의 이중견고성과 효율성

February 2025

DEPARTMENT OF STATISTICS
COLLEGE OF NATURAL SCIENCES
SEOUL NATIONAL UNIVERSITY

Taeyeon Kim

Doubly Robustness and Efficiency of Weighted Estimators incorporating Cohort Data

코호트 데이터를 활용한 가중 추정량의 이중견고성과 효율성

지도교수 신 예 은

이 논문을 이학석사 학위논문으로 제출함

2024년 10월

서울대학교 대학원

통계학과

김 태 연

김 태 연의 이학석사 학위논문을 인준함

2025년 1월

위 원 장	이 재 용	(인)
부위원장	신 예 은	(인)
위 원	이 권 상	(인)

Abstract

We investigate the theoretical and empirical relationship between Augmented Inverse Probability Weighting (AIPW) and Weight Calibration (WC) estimators, emphasizing their double-robust properties. Both methods are commonly employed for partially or selectively observed data, each integrating auxiliary information either via an outcome model (AIPW) or through calibration constraints (WC). Despite seeming differences, we show that—provided either the inclusion probability model or the outcome model is correctly specified—AIPW and WC estimators share the same first-order asymptotic behavior, yielding consistent and efficient estimates with identical influence functions. Building on these insights, we offer a general proof for M-estimation parameters that unifies their equivalence. Through simulation studies under both fully and partially correct models, we find that AIPW and WC maintain nearly identical finite-sample performance in bias and variance, confirming their double robustness. These findings suggest that practitioners can choose either method without sacrificing optimality, as long as the same auxiliary information is incorporated. We also discuss avenues for extending these approaches to more complex designs and survival analyses, where calibration-based or augmented frameworks have shown promise.

Keywords: Augmented Inverse Probability Weighting, Weight Calibration, Double Robustness, M-Estimation, Asymptotic Equivalence

Student Number: 2023-22449

Contents

List of Tables

Chapter 1

Introduction

1.1 Introduction

It is often infeasible to obtain complete data from the target population due to cost, time, or ethical constraints. As a result, observational and survey data have only partial or selective information, leading to potential bias and inefficiency if not addressed properly [1]. One well-known strategy to handle such selective sampling or missing data is inverse probability weighting (IPW), introduced by [2]. For each observation, a weight is assigned proportional to the inverse of its inclusion probability, which would estimate population parameters without bias if those probabilities were correctly specified according to sampling design or missing mechanism. IPW have also been extensively applied in other contexts, such as causal inference [3], where the probability of treatment assignment can be similarly modeled.

However, IPW alone is inefficient because only a complete subset is used while additional information is available in the entire data [1]. To address this shortcoming, [4] proposed augmented inverse probability weighting (AIPW), which adds an extra augmentation term into the IPW estimator. Importantly, it is doubly robust in that AIPW estimators are consistent if either (1) the inclusion probability model or (2) the outcome regression model is correctly specified. Another popular approach to incorporating auxiliary information is weight calibration (WC) method, first introduced by [5]. It directly modifies the IPW weights such that weighted estimators satisfies certain calibration equations consisting of auxiliary information. Like AIPW, WC estimators are also doubly robust as long as either the base inclusion probability model or the specified outcome/auxiliary model is correct.

Despite their different schemes of using additional information, AIPW and WC share strong theoretical connections. Under appropriate conditions, particularly when the two methods use the same auxiliary variables capturing the conditional mean of the outcome, they have the same point estimates and asymptotic variances [6, 7, 8]. From a practical perspective, researchers often choose between AIPW and WC based on implementation convenience, available software, or personal familiarity with the methods. Demonstrating that they have the same optimal performance under correct model specification affirms that one can freely choose either approach without sacrificing statistical efficiency.

Nevertheless, there are relatively few studies that jointly examine and compare AIPW and WC; even among those that do, most have centered on relatively simple estimands such as population totals or means. Perhaps the most direct comparison appears in [7], which highlights how AIPW and WC share a closely related influence-function framework in estimating-equation form, yet it stops short of proving their stochastic convergence. Likewise, [6] and [8] discuss AIPW and calibration-based approaches in survey contexts but chiefly in the setting of population-total estimation, with more emphasis on how adding calibration ideas to existing estimators can improve efficiency. Beyond these works, most papers appear to focus on either AIPW or WC individually rather than systematically comparing the two. For instance, in the AIPW literature, [1] and [9] present illustrative simulations but do not fully explore WC counterparts. Meanwhile, the WC-focused works of [10, 11, 12, 13] primarily develop calibration-based methods in semiparametric or survival contexts without systematically contrasting them against AIPW. To our knowledge, no study has rigorously established, in a fully general estimating-equation framework, whether AIPW and WC are theoretically and empirically equivalent for parameters beyond simple population means.

In this study, we investigate the theoretical and empirical connections between AIPW and WC, focusing on their doubly robust properties. We first confirm that if the two estimators use the same auxiliary information about the conditional expectation of the outcome, they are asymptotically equivalent by having the same first-order behavior in point estimates and asymptotic variance. We then present extensive numerical studies, ranging from population total estimation to linear regression, to illustrate how AIPW and WC perform when the inclusion probability model is correctly specified, when the outcome model is correctly specified, and when either or both are misspecified.

Chapter 2

Theoretical Background

2.1 Notation and Basic Setup

We consider a finite population of size N, indexed by $i = 1, \ldots, N$. Let δ_i be an indicator with $\delta_i = 1$ if the ith unit is observed, and $\delta_i = 0$ otherwise. Denote $\pi_i = P(\delta_i = 1 \mid Z_i)$, where Z_i is a vector of fully observed covariates or design variables. Let $n = \sum_{i=1}^{N} \delta_i$ be the total number of observed units.

- **MAR (Missing at Random).** If δ_i depends only on Z_i, then the data are MAR.

- **MCAR (Missing Completely at Random).** If δ_i is independent of both observed and unobserved outcomes, then the data are MCAR.

- **MNAR (Missing Not at Random).** If δ_i depends on y_i (or other unobserved variables) even after conditioning on Z_i, the data are MNAR.

We observe y_i only when $\delta_i = 1$. Our main task is to construct estimators of population parameters using $\{\delta_i, y_i, Z_i\}$ that are consistent under the assumptions above and ideally have small variance. In what follows, we focus on two primary weighting approaches for handling selective samples: Augmented Inverse Probability Weighting (AIPW) and Weight Calibration (WC).

2.2 Inverse Probability Weighting (IPW)

Inverse Probability Weighting (IPW) is a general framework for handling selective observation or unequal selection probabilities by weighting each observed unit by the inverse of its sampling (or observation) probability. Historically, a well-known example of such an approach in survey sampling is the Horvitz–Thompson (HT) method [2], where each unit i in the population has a known inclusion probability π_i.

Under this design-based framework, a population quantity of interest

$$U(\theta) = \sum_{i=1}^{N} U_i(\theta)$$

is estimated by

$$\hat{U}_{IPW}(\theta) = \sum_{i=1}^{N} \frac{\delta_i}{\pi_i} U_i(\theta), \tag{2.1}$$

where $\delta_i = 1$ if unit i is sampled (or observed) and 0 otherwise. When π_i is correctly specified, IPW of this form is unbiased with respect to the sampling design.

In practice, however, small π_i values can lead to large inverse weights $1/\pi_i$ and inflate the variance significantly [1]. Furthermore, if there exist additional auxiliary variables that are not incorporated into π_i, IPW may lose efficiency by ignoring this extra information. These issues motivate more general IPW approaches beyond the original design-based setting, where π_i may be modeled based on covariates or estimated from the data, rather than being taken as a known inclusion probability.

In many observational studies, π_i represents the probability of being observed or treated, which is typically modeled rather than determined by design [3]. Under this broader IPW framework, a misspecified model for π_i can yield biased estimates, and near-zero π_i values can still cause numerical instability or high variance. These drawbacks motivate extensions such as Augmented IPW (AIPW) and Weight Calibration (WC), which exploit auxiliary variables more fully to improve both robustness and efficiency.

2.3 Augmented Inverse Probability Weighting (AIPW)

Augmented IPW (AIPW) was introduced by [4] to improve upon the basic IPW estimator by adding an augmentation term. Let $U_i(\theta)$ be the quantity of interest and $\pi_i = P\big(\delta_i = 1 \mid Z_i\big)$. Define

$$\begin{aligned}\hat{U}_{AIPW}(\theta) &= \sum_{i=1}^{N} \frac{\delta_i}{\pi_i} U_i(\theta) \;+\; \sum_{i=1}^{N} \big(1 - \tfrac{\delta_i}{\pi_i}\big)\, \phi_i(Z) \\ &= \sum_{i=1}^{N} U_i(\theta) \;+\; \sum_{i=1}^{N} \frac{\delta_i - \pi_i}{\pi_i} \big(U_i(\theta) - \phi_i(Z)\big), \qquad (2.2)\end{aligned}$$

where $\phi_i(Z)$ is a function of fully observed covariates Z.

2.3.1 Doubly Robust Consistency

Under MAR (Missing at Random) or a correctly modeled π_i, taking expectations of (2.2) yields

$$E\big[\hat{U}_{AIPW}(\theta)\big] = E\Big[\sum_{i=1}^{N} U_i(\theta)\Big] \ + \ \sum_{i=1}^{N} E\Big[\tfrac{\delta_i - \pi_i}{\pi_i}\,\big(U_i(\theta) - \phi_i(Z)\big)\Big]. \tag{2.3}$$

If either

1. the missingness probability π_i is correctly specified, or

2. $\phi_i(Z) = E\big[U_i(\theta) \mid Z_i\big]$,

then the second summation in (2.3) vanishes, implying $E[\hat{U}_{AIPW}(\theta)] = E\big[\sum_{i=1}^{N} U_i(\theta)\big]$. Hence, $\hat{U}_{AIPW}(\theta)$ is consistent under at least one correctly specified model.

2.3.2 Variance Decomposition

Applying $\text{Var}(\cdot)$ to (2.2) yields

$$\text{Var}\Big[\hat{U}_{AIPW}(\theta)\Big] = \sum_{i=1}^{N} \text{Var}\big[U_i(\theta)\big] \ + \ \sum_{i=1}^{N} E\Big[\tfrac{1-\pi_i}{\pi_i}\,\big(U_i(\theta) - \phi_i(Z)\big)^2\Big] \tag{2.4}$$

$$\geq \sum_{i=1}^{N} \text{Var}[U_i(\theta)]. \tag{2.5}$$

We interpret the first summation, $\sum_{i=1}^{N} \text{Var}[U_i(\theta)]$, as the variance we would have if the data were fully observed, and the second summation, $\sum_{i=1}^{N} E\big[\frac{1-\pi_i}{\pi_i}\,(U_i(\theta) - \phi_i(Z))^2\big]$, as the additional variance contributed by missingness or selective sampling. Equality holds if $\phi_i(Z) = E[U_i(\theta) \mid Z_i]$, in which case $(U_i(\theta) - \phi_i(Z))$ has conditional mean zero and the extra variance term is minimized. For detailed proofs of the doubly robust property and the variance decomposition, refer to [4].

2.4 Weight Calibration (WC)

The weight calibration estimation was introduced by [5] as a procedure of minimizing a distance measure between initial weights and final weights subject to calibration equations. Let

$$\hat{U}_{IPW}(\theta) \;=\; \sum_{i=1}^{N} \delta_i \, \omega_i^0 \, U_i(\theta)$$

denote the IPW estimator in (2), where $\omega_i^0 = 1/\pi_i$. Suppose we have auxiliary variables A_i, and we define a set of calibrated weights $\omega_i = \omega_i^0 \, g_i(\lambda)$ that satisfy the calibration constraint

$$\sum_{i=1}^{N} \delta_i \, \omega_i \, A_i \;=\; \sum_{i=1}^{N} A_i.$$

Here, $g_i(\lambda)$ is a function of λ determined from the above constraint. The resulting weight-calibrated (WC) estimator is

$$\hat{U}_{WC}(\theta) \;=\; \sum_{i=1}^{N} \delta_i \, \omega_i^0 \, g_i(\lambda) \, U_i(\theta). \tag{2.6}$$

To define a unique estimator that satisfies the calibration constraint, [5] introduced a distance function $d\big(\omega_i^0, \omega_i\big)$ to measure the discrepancy between ω_i^0 and ω_i. They showed that minimizing

$$\sum_{i=1}^{N} \delta_i \, d\big(\omega_i^0,\, \omega_i\big) \quad \text{subject to} \quad \sum_{i=1}^{N} \delta_i \, \omega_i \, A_i \;=\; \sum_{i=1}^{N} A_i$$

leads to an asymptotically unbiased and consistent estimator. In this study, we choose the following two examples of $d(\omega_i^0, \omega_i)$:

- Chi-squared distance:

$$d\big(\omega_i^0, \omega_i\big) = \frac{(\omega_i - \omega_i^0)^2}{2\,\omega_i^0}, \quad \text{which gives} \quad \omega_i = \omega_i^0\big[\,1 + \lambda^\top A_i\big].$$

- Log-distance (Raking):

$$d\big(\omega_i^0, \omega_i\big) = \omega_i \,\log\!\Big(\frac{\omega_i}{\omega_i^0}\Big) - \omega_i + \omega_i^0, \quad \text{which gives} \quad \omega_i = \omega_i^0 \exp\!\big(\lambda^\top A_i\big).$$

2.4.1 Doubly Robust Consistency

The weight-calibrated estimator $\hat{U}_{WC}(\theta)$ remains consistent if either the base weights $\omega_i^0 = 1/\pi_i$ are correctly specified or the auxiliary variables satisfy a correct outcome model. Specifically, let

$$\hat{U}_{WC}(\theta) \;=\; \sum_{i=1}^{N} \delta_i\, \omega_i^0\, g_i(\lambda)\, U_i(\theta),$$

as in (2.6), and suppose A_i is such that $A_i = E\big[U_i(\theta) \mid Z_i\big]$. We highlight two scenarios:

1. If $\omega_i^0 = 1/\pi_i$ is correctly specified, then the usual IPW estimator

   $$\hat{U}_{IPW}(\theta) \;=\; \sum_{i=1}^{N} \delta_i\, \omega_i^0\, U_i(\theta)$$

   is design-consistent in the sense that

   $$E\big[\hat{U}_{IPW}(\theta)\big] \;=\; E\Big[\sum_{i=1}^{N} U_i(\theta)\Big].$$

   By [5], multiplying these weights by the calibration factor $g_i(\lambda)$ does not affect first-order consistency. Formally,

   $$\hat{U}_{WC}(\theta) \;-\; \hat{U}_{IPW}(\theta) \;=\; O_p\big(n^{-1/2}\big).$$

Consequently,

$$E\big[\hat{U}_{WC}(\theta)\big] \;-\; E\big[\hat{U}_{IPW}(\theta)\big] \;=\; O\big(n^{-1/2}\big),$$

and hence $\hat{U}_{WC}(\theta)$ remains consistent at the $O\big(n^{-1/2}\big)$ scale.

2. If $A_i = E\big[U_i(\theta) \mid Z_i\big]$, then taking expectations of $\hat{U}_{WC}(\theta)$ yields

$$E\big[\hat{U}_{WC}(\theta)\big] \;=\; E\Big[\sum_{i=1}^{N} \delta_i\,\omega_i^0\,g_i(\lambda)\,E\big[U_i(\theta) \mid Z_i\big]\Big] \;=\; E\Big[\sum_{i=1}^{N} \delta_i\,\omega_i^0\,g_i(\lambda)\,A_i\Big].$$

By construction of the calibrated weights,

$$\sum_{i=1}^{N} \delta_i\,\omega_i^0\,g_i(\lambda)\,A_i \;=\; \sum_{i=1}^{N} A_i,$$

and

$$E\big[\sum_{i=1}^{N} A_i\big] \;=\; \sum_{i=1}^{N} E\big[E\big[\,U_i(\theta) \mid Z_i\big]\big] \;=\; E\big[\sum_{i=1}^{N} U_i(\theta)\big],$$

so

$$E\big[\hat{U}_{WC}(\theta)\big] \;=\; E\big[\sum_{i=1}^{N} U_i(\theta)\big].$$

Hence, $\hat{U}_{WC}(\theta)$ is consistent when at least one of these two models is specified correctly.

2.4.2 Variance Decomposition

We now analyze the variance of the weight-calibrated estimator

$$\hat{U}_{WC}(\theta) \;=\; \sum_{i=1}^{N} \delta_i\,\omega_i^0\,g_i(\lambda)\,U_i(\theta).$$

The calibration constraint imposes

$$\sum_{i=1}^{N} \delta_i \, \omega_i^0 \, g_i(\lambda) \, A_i \;=\; \sum_{i=1}^{N} A_i,$$

implying

$$\hat{U}_{WC}(\theta) \;=\; \sum_{i=1}^{N} \delta_i \, \omega_i^0 \, g_i(\lambda) \, \big[U_i(\theta) - A_i\big] \;+\; \sum_{i=1}^{N} A_i.$$

Define

$$V_i \;=\; \delta_i \, \omega_i^0 \, g_i(\lambda) \, \big(U_i(\theta) - A_i\big).$$

To derive $\mathrm{Var}[\hat{U}_{WC}(\theta)]$, we apply the law of total variance to each V_i:

$$\mathrm{Var}[V_i] \;=\; \mathrm{Var}\big(\mathrm{E}[V_i \mid Z_i, U_i(\theta)]\big) \;+\; \mathrm{E}\big[\mathrm{Var}(V_i \mid Z_i, U_i(\theta))\big].$$

Since $\delta_i \mid Z_i \sim \mathrm{Bernoulli}(\pi_i)$, we have $\mathrm{Var}(\delta_i \mid Z_i) = \pi_i(1 - \pi_i)$, leading to

$$V_i \;=\; \delta_i \, \frac{1}{\pi_i} \, g_i(\lambda) \, \big(U_i(\theta) - A_i\big).$$

Hence,

$$\mathrm{Var}[V_i \mid Z_i, U_i(\theta)] \;=\; \big(\tfrac{g_i(\lambda)}{\pi_i}\big)^2 \, (U_i(\theta) - A_i)^2 \, \pi_i(1 - \pi_i).$$

Next, we compute $\mathrm{Var}\big(\mathrm{E}[V_i \mid Z_i, U_i(\theta)]\big)$. Since $\mathrm{E}[\delta_i \mid Z_i] = \pi_i$, it follows that

$$\mathrm{E}[V_i \mid Z_i, U_i(\theta)] \;=\; \mathrm{E}\Big[\delta_i \, \tfrac{1}{\pi_i} \, g_i(\lambda) \, \big(U_i(\theta){-}A_i\big) \;\Big|\; Z_i, U_i(\theta)\Big] \;=\; g_i(\lambda) \, \big(U_i(\theta){-}A_i\big).$$

Therefore,

$$\mathrm{Var}\big(\mathrm{E}[V_i \mid Z_i, U_i(\theta)]\big) \;=\; \mathrm{Var}\Big[g_i(\lambda) \, (U_i(\theta) - A_i)\Big] \;=\; g_i(\lambda)^2 \, \mathrm{Var}\big(U_i(\theta) - A_i\big).$$

Putting these inner and outer parts together, we obtain

$$\begin{aligned} \mathrm{Var}[V_i] = \mathrm{E}\Big[&(U_i(\theta) - A_i)^2 \big(\tfrac{g_i(\lambda)}{\pi_i}\big)^2 \, \pi_i(1 - \pi_i)\Big] \\ &+\; g_i(\lambda)^2 \Big[\mathrm{Var}\big(U_i(\theta)\big) + \mathrm{Var}\big(A_i\big) - 2\,\mathrm{Cov}\big(U_i(\theta), A_i\big)\Big]. \quad (2.7) \end{aligned}$$

Summing over $i = 1, \ldots, N$ and assuming $\{V_i\}$ are independent, we obtain

$$\operatorname{Var}\big[\hat{U}_{WC}(\theta)\big] = \sum_{i=1}^{N} \operatorname{Var}[V_i].$$

Under certain orthogonality conditions—e.g. if A_i is the correct outcome model $E[\, U_i(\theta) \mid Z_i]$, or π_i is correctly specified—we reach the simpler form

$$\operatorname{Var}\big[\hat{U}_{WC}(\theta)\big] = \sum_{i=1}^{N} g_i(\lambda)^2 \Big\{ \operatorname{Var}[U_i(\theta)] \;+\; \mathrm{E}\Big[(U_i(\theta) - A_i)^2 \, \tfrac{1-\pi_i}{\pi_i} \Big] \Big\},$$

which can also be written as

$$\operatorname{Var}\big[\hat{U}_{WC}(\theta)\big] \;=\; \sum_{i=1}^{N} g_i(\lambda)^2 \operatorname{Var}\big(U_i(\theta)\big) \;+\; \sum_{i=1}^{N} g_i(\lambda)^2 \, \mathrm{E}\Big[\tfrac{1-\pi_i}{\pi_i} \, (U_i(\theta) - A_i)^2 \Big].$$

In this expression, the first summation represents the baseline variance under full observation, while the second summation captures the extra variance due to missingness or unequal sampling. If A_i is chosen as $E\big[\, U_i(\theta) \mid Z_i \big]$, then $(U_i(\theta) - A_i)$ is conditionally orthogonal to the sampling process and this additional term can be minimized.

Chapter 3

Equivalence of AIPW and Calibration Estimators

In this chapter, we investigate the relationship between the AIPW estimator and the weight calibration estimator. We first demonstrate that these two estimators are exactly equivalent in the well-known example of estimating the population total, and then extend the result to a general M-estimation framework.

3.1 Population Total

As a concrete illustration, consider estimating the population total

$$T = \sum_{i=1}^{N} y_i.$$

Define X as the $n \times 2$ matrix with rows $(1, x_i)$ for the observed units, and let $Y = (y_1, \ldots, y_n)^\top$. We set W to be the $n \times n$ diagonal matrix with entries

$1/\pi_i$. Using weighted least squares (WLS), one obtains

$$\hat{\beta} \;=\; (X^\top W X)^{-1} \,(X^\top W Y), \quad \text{and define} \quad \hat{\phi}_i(x) \;=\; x_i\,\hat{\beta}.$$

Substituting $\hat{\phi}_i(x)$ into the usual AIPW form yields

$$\hat{T}_{AIPW} = \sum_{i=1}^{N} \frac{\delta_i}{\pi_i}\, y_i \;-\; \sum_{i=1}^{N} \frac{\delta_i - \pi_i}{\pi_i}\,\bigl(x_i\,\hat{\beta}\bigr).$$

To simplify this expression, note that $T_x = \sum_{i=1}^{N} x_i$ and $\hat{T}_x = \sum_{i=1}^{n} \frac{\delta_i}{\pi_i}\, x_i$. After a short algebraic rearrangement involving these terms and $(X^\top W X)^{-1}$, we find

$$\hat{T}_{AIPW} = \sum_{i=1}^{N} \delta_i \,\frac{1 + (T_x - \hat{T}_x)\,(X^\top W X)^{-1}\, x_i}{\pi_i}\, y_i.$$

From a weight-calibration (WC) perspective, we take the same base weights $\omega_i^0 = 1/\pi_i$ and apply a chi-square distance function $d(a, b) = \frac{(a-b)^2}{2\,b}$, with auxiliary variable $A_i = x_i$. Solving the calibration equations yields

$$\omega_i \;=\; \omega_i^0\,\bigl(1 + \lambda\, x_i\bigr) \;=\; \frac{1 + (T_x - \hat{T}_x)\,(X^\top W X)^{-1}\, x_i}{\pi_i}.$$

The resulting WC estimator is

$$\hat{T}_{WC} = \sum_{i=1}^{N} \delta_i\,\omega_i\,y_i.$$

Consequently,, $\hat{T}_{WC}$ coincides exactly with the $\hat{T}_{AIPW}$ expression above. Therefore, under the assumptions of linear regression and chi-square calibration, both approaches produce the same point estimate for the population total. A detailed derivation of these steps can be found in [7].

3.2 General M-Estimation Framework

We now extend these ideas to more general parameters, such as regression coefficients obtained via estimating equations. Our goal is to show that the AIPW-based estimator and the calibration-based estimator remain asymptotically equivalent under suitable conditions on the auxiliary information.

Let $\theta \in \mathbb{R}^d$ be a finite-dimensional parameter of interest, and denote the full-data estimating function by

$$\sum_{i=1}^{N} U_i\big(\theta\big) \;=\; 0,$$

which would be solvable if all units $\{1, \ldots, N\}$ had fully observed data. However, only a subset is available in practice, necessitating a weighted estimating equation.

3.2.1 Inverse Probability Weighting (IPW)

If we adopt the IPW form, the estimator $\hat{\theta}_{IPW}$ satisfies

$$\sum_{i=1}^{N} \frac{\delta_i}{\pi_i} U_i\big(\hat{\theta}_{IPW}\big) \;=\; 0, \tag{3.1}$$

where $\delta_i \in \{0, 1\}$ indicates whether unit i is observed, and $\pi_i = P(\delta_i = 1 \mid Z_i)$.

3.2.2 Augmented Inverse Probability Weighting (AIPW)

The AIPW estimator $\hat{\theta}_{AIPW}$ solves

$$\sum_{i=1}^{N} \Big[\tfrac{\delta_i}{\pi_i} U_i\big(\hat{\theta}_{AIPW}\big) \;+\; \Big(1 - \tfrac{\delta_i}{\pi_i}\Big)\, \phi_i(Z) \Big] \;=\; 0, \tag{3.2}$$

where $\phi_i(Z)$ is an augmentation term. The optimal choice is $\phi_i(Z) = E\big[\,U_i(\theta) \mid Z_i\big]$, which minimizes variance and guarantees double robustness. Under these standard conditions, $\hat{\theta}_{AIPW}$ is consistent if either π_i or $\phi_i(Z)$ is correctly specified.

3.2.3 Weight Calibration (WC)

The calibration estimator $\hat{\theta}_{WC}$ solves

$$\sum_{i=1}^{N} \delta_i\, w_i(\lambda)\, U_i\big(\hat{\theta}_{WC}\big) \;=\; 0, \tag{3.3}$$

where $w_i(\lambda)$ are calibrated weights satisfying the constraint

$$\sum_{i=1}^{N} \delta_i\, w_i(\lambda)\, A_i \;=\; \sum_{i=1}^{N} A_i,$$

for some auxiliary variables A_i. Typically, one starts from $w_i^0 = 1/\pi_i$, and applies a minimal "distance" adjustment subject to the above calibration equation. Analogously, the optimal auxiliary vector is $A_i = E\big[\,U_i(\theta) \mid Z_i\big]$, which ensures double robustness.

3.3 Asymptotic Equivalence under Correct Specification

3.3.1 Assumptions

We assume the following throughout our proof:

(A1) Regularity conditions from [14] hold.

(A2) $\hat{\lambda} = O_p\big(n^{-1/2}\big)$, following [5].

(A3) $U_i(\theta_0)$ and $\phi_i(Z_i)$ each have finite second moments.

(A4) $\max \|z_i\| < \infty$, ensuring $\hat{\lambda}^\top z_i = O_p\big(n^{-1/2}\big)$.

(A5) $\pi_i > \sigma > 0$ with probability 1, preventing extreme weights.

In **(A1)**, we adopt standard M-estimation regularity conditions [14], ensuring $\sqrt{n}$-consistency and asymptotic linearity. Next, **(A2)** follows [5, Section 2], where $\hat{\lambda}$ converges at the $n^{-1/2}$ rate via minimal-distance calibration. For **(A3)**, we require $U_i(\theta_0)$ and $\phi_i(Z_i)$ to have finite second moments, aligning with AIPW-type arguments [4, Section 2.4]. Moving to **(A4)**, we assume $\{z_i\}$ is bounded so that $\hat{\lambda}^\top z_i$ remains $O_p(n^{-1/2})$ [5, Results 3–5], though unbounded z_i might still be feasible if $\hat{\lambda}$ converges to zero sufficiently fast. Finally, **(A5)** ensures δ_i/π_i does not blow up, preventing extreme inverse-probability weights.

We also note that the inclusion probability model is correct if $\pi_i = P(\delta_i = 1 \mid Z_i)$, while the outcome model is correct if, under AIPW, $\phi_i(Z_i) = E\big[U_i(\theta) \mid Z_i\big]$, or under WC, $A_i = E\big[U_i(\theta) \mid Z_i\big]$. In both cases, the additional information is incorporated as an influence function.

3.3.2 Asymptotic Equivalence

Under Assumptions (A1)–(A5), suppose that either the inclusion probability model or the outcome model is correctly specified for both the AIPW estimator and the WC estimator. Then they satisfy

$$\sqrt{n}\big(\hat{\theta}_{AIPW} - \hat{\theta}_{WC}\big) = o_p(1).$$

Proof. From (A1), both $\hat{\theta}_{AIPW}$ and $\hat{\theta}_{WC}$ can be viewed as M-estimators $\hat{\theta}$

satisfying

$$\frac{1}{n}\sum_{i=1}^{n} U\big(\hat{\theta}; z_i\big) = 0$$

for some estimating function $U(\theta; z_i)$. Each $\hat{\theta}$ has an asymptotic linear expansion of the form

$$\sqrt{n}\,\big(\hat{\theta}-\theta_0\big) \;=\; -\Big\{\partial_\theta\,\mathbb{E}\big[U(\theta_0; z)\big]\Big\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U\big(\theta_0; z_i\big) \;+\; o_p(1),$$

where θ_0 is the true parameter value. First, we aim to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_A(\theta_0; z_i) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_W(\theta_0; z_i) \;+\; o_p(1).$$

From Equations 3.2 and 3.3, the above can be written as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \delta_i\, w_i(\hat{\lambda})\, U_i(\theta_0) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\tfrac{\delta_i}{\pi_i}\, U_i(\theta_0) \;+\; \Big(1-\tfrac{\delta_i}{\pi_i}\Big)\,\phi_i(Z_i)\Big\} \;+\; o_p(1).$$

Under assumptions (A2) and (A4), Deville–Särndal's expansion [5] shows

$$w_i(\hat{\lambda}) \;=\; \frac{1}{\pi_i}\Big[1+\hat{\lambda}^\top z_i \;+\; O\big((\hat{\lambda}^\top z_i)^2\big)\Big].$$

Substituting this expansion and computing the difference between the two estimators, we get

$$\begin{aligned}
\Delta_i &= \Big[\frac{\delta_i}{\pi_i}U_i(\theta_0)\Big(1+\hat{\lambda}^\top z_i + O\big((\hat{\lambda}^\top z_i)^2\big)\Big)\Big] - \Big[\frac{\delta_i}{\pi_i}\,U_i(\theta_0) + \Big(1-\frac{\delta_i}{\pi_i}\Big)\,\phi_i(Z_i)\Big] \\
&= \underbrace{\frac{\delta_i}{\pi_i}\big(\hat{\lambda}^\top z_i\big)\,U_i(\theta_0)}_{\text{(a) Term}} \;-\; \underbrace{\Big(1-\frac{\delta_i}{\pi_i}\Big)\,\phi_i(Z_i)}_{\text{(b) Term}} \;+\; \underbrace{\frac{\delta_i}{\pi_i}\,O\big((\hat{\lambda}^\top z_i)^2\big)\,U_i(\theta_0)}_{\text{(c) Term}}.
\end{aligned}$$

(a) Term. Rewrite term as

$$\hat{\lambda}^\top\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\,z_i\,U_i(\theta_0)\Big).$$

Under **(A5)**, δ_i/π_i is bounded in probability, so

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\, z_i\, U_i(\theta_0) \;=\; O_p(1),$$

by a standard Central Limit Theorem (CLT) argument, provided z_i is bounded and $U_i(\theta_0)$ is mean-zero with finite variance. Including the factor δ_i/π_i (bounded by (A5)) does not affect the $\sqrt{n}$ rate in the sum, so dividing by $\sqrt{n}$ yields $O_p(1)$ overall. Next, by (A2) and (A4), we have $\|\hat{\lambda}\| = O_p(n^{-1/2})$. Therefore,

$$\hat{\lambda}^\top\Big(\tfrac{1}{\sqrt{n}}\sum_{i=1}^{n}\tfrac{\delta_i}{\pi_i}\, z_i\, U_i(\theta_0)\Big) \;=\; O_p\big(n^{-1/2}\big) \;\times\; O_p(1) \;=\; O_p\big(n^{-1/2}\big) \;=\; o_p(1).$$

(b) Term. We rewrite it as

$$\frac{1}{\sqrt{n}}\Big[\sum_{i=1}^{n}\phi_i(Z_i) \;-\; \sum_{i=1}^{n}\tfrac{\delta_i}{\pi_i}\,\phi_i(Z_i)\Big].$$

Because $\phi_i(Z_i)$ can serve as a calibration variable (see [5] for details), we have by weak calibration that

$$\sum_{i=1}^{n}\Big(\phi_i(Z_i) - \delta_i\, w_i(\hat{\lambda})\,\phi_i(Z_i)\Big) \;=\; o_p(\sqrt{n}).$$

Moreover, with baseline weights $w_i^0 = 1/\pi_i$ and the fact that $w_i(\hat{\lambda}) - w_i^0 = O_p(n^{-1/2})$, one shows

$$\sum_{i=1}^{n}\big\{\tfrac{\delta_i}{\pi_i}\,\phi_i(Z_i) - \delta_i\, w_i(\hat{\lambda})\,\phi_i(Z_i)\big\} \;=\; o_p(\sqrt{n}).$$

Hence, combining these two results,

$$\frac{1}{\sqrt{n}}\Big[\sum_{i=1}^{n}\phi_i(Z_i) - \sum_{i=1}^{n}\tfrac{\delta_i}{\pi_i}\,\phi_i(Z_i)\Big] \;=\; o_p(1).$$

(c) Term. Finally, for

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\pi_i} O\big((\hat{\lambda}^\top z_i)^2\big) U_i(\theta_0),$$

note that $(\hat{\lambda}^\top z_i)^2 = O_p(n^{-1})$ by (A2) and (A4). Hence

$$\frac{\delta_i}{\pi_i} O\big((\hat{\lambda}^\top z_i)^2\big) U_i(\theta_0) = O_p\big(n^{-1}\big) \times U_i(\theta_0),$$

and summing n such terms yields a total of $O_p(1)$. Dividing by $\sqrt{n}$ therefore gives $O_p(n^{-1/2})$, which is $o_p(1)$.

Combining (a), (b), and (c) shows that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Delta_i \ = \ o_p(1).$$

We now want to show that if either the inclusion probability model or the outcome model is correctly specified for both the AIPW estimator and the WC estimator, then

$$\partial_\theta \, \mathbb{E}\big[U_A(\theta; z)\big] \ = \ \partial_\theta \, \mathbb{E}\big[U_W(\theta; z)\big].$$

From Sections 2.3.1 and 2.4.1, we already have the following two key scenarios:

Case 1: Correct inclusion probability model. If the inclusion probability model π_i is correctly specified, then the AIPW estimator satisfies

$$E\big[\,U_A(\theta; z)\big] \ = \ E\Big[\sum_{i=1}^{N} U_i(\theta)\Big],$$

while for WC, we have

$$E\big[\hat{U}_{WC}(\theta)\big] \ - \ E\Big[\sum_{i=1}^{N} U_i(\theta)\Big] \ = \ O\big(n^{-1/2}\big),$$

Thus, to first order, the small calibration factor $g_i(\hat{\lambda})$ in $w(\hat{\lambda})$ does not change the derivative at θ_0. Consequently,

$$\frac{\partial}{\partial \theta}\, \mathbb{E}\big[U_A(\theta; z)\big]\Big|_{\theta_0} \;=\; \frac{\partial}{\partial \theta}\, \mathbb{E}\big[U_W(\theta; z)\big]\Big|_{\theta_0}.$$

Case 2: Correct outcome model. If the outcome model is correctly specified, then both

$$U_A(\theta; z) \quad \text{and} \quad U_W(\theta; z)$$

are unbiased for $\sum_{i=1}^{N} U_i(\theta)$ to first order. Indeed,

$$E\big[U_A(\theta; z)\big] \;=\; E\Big[\sum_{i=1}^{N} U_i(\theta)\Big], \quad E\big[U_W(\theta; z)\big] \;=\; E\Big[\sum_{i=1}^{N} U_i(\theta)\Big].$$

Hence again,

$$\frac{\partial}{\partial \theta}\, \mathbb{E}\big[U_A(\theta; z)\big]\Big|_{\theta_0} \;=\; \frac{\partial}{\partial \theta}\, \mathbb{E}\big[U_W(\theta; z)\big]\Big|_{\theta_0}.$$

Thus, either correct inclusion probability model or correct outcome model suffices to make AIPW and WC share the same first-order derivative in their M-estimation equations. By Slutsky's theorem, it then follows that

$$\sqrt{n}\,\big(\hat{\theta}_{AIPW} - \theta_0\big) \;-\; \sqrt{n}\,\big(\hat{\theta}_{WC} - \theta_0\big) \;=\; o_p(1),$$

and hence

$$\sqrt{n}\big(\hat{\theta}_{AIPW} - \hat{\theta}_{WC}\big) \;=\; o_p(1).$$

□

These results demonstrate that although $\hat{\theta}_{AIPW}$ and $\hat{\theta}_{WC}$ may not be exactly the same in finite samples due to their iterative construction, under either the correct inclusion probability model or the correct outcome model, they converge to the same distribution asymptotically. In particular, specifying the auxiliary variables as the projection of $U_i(\theta)$ onto the auxiliary information, that is $\phi_i(Z_i) = E\big[U_i(\theta) \mid Z_i\big]$, ensures that both $\hat{\theta}_{AIPW}$ and $\hat{\theta}_{WC}$ make optimal use of the observed data, and thus they coincide asymptotically.

Chapter 4

Simulation Study

4.1 Setup

We generate a population of size $N = 5000$. Each unit i has covariates (X_{1i}, Z_{1i}, Z_{2i}), which, for simplicity, are assumed to follow a trivariate normal distribution. Specifically, we set $(X_1,\ Z_1,\ Z_2) \sim \mathcal{N}(\mu,\ \Sigma)$, where the mean vector $\mu = (2,\ 0,\ 1)^\top$ and the covariance matrix Σ has unit variances on the diagonal and moderate correlations (e.g., around 0.1–0.5) on the off-diagonals. Thus, each of X_1, Z_1, Z_2 has variance 1, while their pairwise correlations lie in a moderate range. We generate a partially observed regressor X_2 via $X_2 = 0.8 + 0.1\,X_1 + \ Z_1 + \mathcal{N}(0,\ 0.2^2)$, and then define the outcome $Y = 1 + X_1 + X_2 + \mathcal{N}(0,\ 1)$.

To induce selective sampling, we assign each unit i an inclusion probability π_i according to a logistic model with the linear predictor $\operatorname{logit}(\pi_i) =$

$-0.8 + Z_{1i} - Z_{2i} + \mathcal{N}(0, 0.1^2)$. We then sample $\delta_i \sim \text{Bernoulli}(\pi_i)$. Here, $\delta_i = 1$ indicates that we observe (Y_i, X_{1i}, X_{2i}), whereas $\delta_i = 0$ means X_{2i} is unobserved. This design follows a Missing at Random (MAR) mechanism, since δ_i depends only on the fully observed variables (Z_{1i}, Z_{2i}).

We wish to study doubly robustness by allowing either the outcome model (generating X_2 and thus Y) or the inclusion-probability model (π) to be correct or misspecified.

- **Outcome Model:**
  - (i) *Correct:* $X_2 = \eta_0 + \eta_1 X_1 + \eta_2 Z_1 + \epsilon$.
  - (ii) *Misspecified:* $X_2^{\text{mis}} = \eta_0 + \eta_1 X_1 + \eta_2 (Z_1 \cdot X_1) + \epsilon$.
- **Inclusion Probability:**
  - (i) *Correct:* $\text{logit}(\pi) = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2$.
  - (ii) *Misspecified:* $\text{logit}(\pi) = \gamma_0 + \gamma_1 \, \text{sign}(Z_1) + \gamma_2 \, \text{sign}\big(Z_2 - 1\big)$.

We repeat this entire data generation and sampling procedure $B = 3000$ times to compare the estimators under four different scenarios. Specifically, we consider all combinations of whether the inclusion probability model π is correct or misspecified, and whether the outcome model is correct or misspecified.

4.2 Estimation Procedure

Once a sample is drawn, we observe $\delta_i = 1$ units (of size n) and fit a linear regression of Y on (X_1, X_2). Three estimators are considered:

- **IPW:** We perform a weighted least squares with weights $w_i = 1/\hat{\pi}_i$, where $\hat{\pi}_i$ is fitted from either the correct or misspecified logistic model. If $\hat{\pi}_i$ is correctly specified, IPW can be unbiased, but it may lose efficiency when additional variables are not fully exploited.

Both the AIPW and Weight Calibration estimators use imputation to handle missing X_2. To maximize efficiency, we incorporate the imputed $\hat{X}_2$ into an augmentation term designed to approximate $E[U_i(\beta) \mid Z_i]$.

- **AIPW:** We solve

$$\sum_{i=1}^{N} \left[\frac{\delta_i}{\hat{\pi}_i} U_i(\beta) \; + \; \left(1 - \frac{\delta_i}{\hat{\pi}_i}\right) \phi_i \right] = 0,$$

  where $U_i(\beta) = x_i\big(y_i - x_i^\top \beta\big)$ is the usual regression score function, where $x_i = (1, X_{1i}, X_{2i})^\top$. The $\phi_i = x_i^{\text{imp}}\Big(y_i \; - \; \big(x_i^{\text{imp}}\big)^\top \hat{\beta}\Big)$ is built from the imputed $\hat{X}_2$ as a pseudo-influence function. AIPW is doubly robust, meaning it remains consistent if either $\hat{\pi}_i$ is correctly specified or the outcome model is correctly specified.

- **WC:** We start with $w_i^0 = 1/\hat{\pi}_i$ and adjust them to w_i by imposing

$$\sum_{i=1}^{N} \delta_i \, w_i \, U_i(\beta) = 0, \quad \sum_{i=1}^{N} \delta_i \, w_i \, \tilde{IF}_i \; = \; \sum_{i=1}^{N} \tilde{IF}_i,$$

  where $\tilde{IF}_i$ is a pseudo-influence function derived from the same imputation. Minimizing $\sum \delta_i \, d(w_i, w_i^0)$ subject to these constraints yields a WC estimator that is also doubly robust. In large samples, AIPW and WC are asymptotically equivalent under the same conditions.

Table 4.1: Both correctly specified

	Metric	Full	IPW	WC	AIPW
β_0	Mean	0.000	0.004	0.001	0.001
	Bias	0.000	0.004	0.001	0.001
	SD	0.033	0.113	0.052	0.059
	RE	1.000	0.295	0.644	0.563
β_1	Mean	1.000	0.999	1.000	1.000
	Bias	0.000	-0.002	-0.001	0.000
	SD	0.015	0.058	0.025	0.030
	RE	1.000	0.251	0.587	0.482
β_2	Mean	2.000	1.999	2.000	1.999
	Bias	0.000	-0.001	0.000	0.000
	SD	0.014	0.062	0.027	0.038
	RE	1.000	0.228	0.515	0.374

Table 4.2: π: Misspecified, $E[U(\beta) \mid Z]$: Correct

	Metric	Full	IPW	WC	AIPW
β_0	Mean	0.000	0.002	0.000	0.000
	Bias	0.000	0.002	0.000	0.000
	SD	0.033	0.085	0.044	0.048
	RE	1.000	0.393	0.765	0.702
β_1	Mean	1.000	0.999	1.000	1.000
	Bias	0.000	-0.001	0.000	0.000
	SD	0.015	0.041	0.021	0.022
	RE	1.000	0.355	0.714	0.657
β_2	Mean	2.000	1.999	2.000	2.000
	Bias	0.000	-0.001	0.000	0.000
	SD	0.014	0.038	0.021	0.023
	RE	1.000	0.365	0.684	0.615

Table 4.3: π: Correct, $E[U(\beta) \mid Z]$: Misspecified

	Metric	Full	IPW	WC	AIPW
β_0	Mean	0.000	0.004	-0.004	-0.004
	Bias	0.000	0.004	-0.004	-0.004
	SD	0.033	0.113	0.096	0.191
	RE	1.000	0.295	0.349	0.175
β_1	Mean	1.000	0.999	1.008	1.003
	Bias	0.000	-0.002	0.007	0.003
	SD	0.015	0.058	0.045	0.079
	RE	1.000	0.251	0.324	0.184
β_2	Mean	2.000	1.999	1.992	1.996
	Bias	0.000	-0.001	-0.008	-0.003
	SD	0.014	0.062	0.049	0.116
	RE	1.000	0.228	0.288	0.121

Table 4.4: Both misspecified

	Metric	Full	IPW	WC	AIPW
β_0	Mean	0.000	0.002	-0.095	-0.363
	Bias	0.000	0.002	-0.095	-0.364
	SD	0.033	0.085	0.073	0.115
	RE	1.000	0.393	0.460	0.291
β_1	Mean	1.002	1.006	1.062	1.160
	Bias	0.000	0.004	0.060	0.158
	SD	0.019	0.056	0.046	0.044
	RE	1.000	0.338	0.410	0.427
β_2	Mean	1.997	1.986	1.971	1.946
	Bias	0.000	-0.011	-0.026	-0.051
	SD	0.013	0.049	0.033	0.049
	RE	1.000	0.267	0.394	0.264

4.3 Simulation Results

Tables 4.1–4.4 summarize the simulation results. As illustrated by Tables 4.2 and 4.3, whenever at least one of the two models—the inclusion probability model π or the outcome model—is correctly specified, the AIPW and WC estimators remain nearly unbiased. This observation confirms their well-known double-robust property. In contrast, the IPW estimator can be more sensitive to whether π is accurately modeled, because its construction relies solely on the correctness of the inclusion probability. At the same time, IPW may still perform reasonably if its implicit augmentation (treating $E[U_i(\beta) \mid Z_i]$ as zero) happens to be close to the truth in practice, for example if the true outcome structure inadvertently aligns with that simplification.

When both models are correctly specified, AIPW and WC not only eliminate bias but also exhibit an efficiency gain compared to IPW. Specifically, we observe lower standard deviations (SD) and higher relative efficiencies (RE) under these scenarios, reflecting the benefit of incorporating the correct conditional expectation $E[U_i(\beta) \mid Z_i]$ into the estimation process. Moreover, AIPW and WC produce almost identical point estimates and variances, illustrating their asymptotic equivalence.

When exactly one of the two models is misspecified, the pattern depends on which component is wrong. If the inclusion probability π is incorrect while the outcome model is correct, IPW becomes visibly biased or yields higher variance, because it has no reliable way to compensate for the erroneous propensity. By contrast, AIPW and WC still rely on the correct outcome specification, so they remain robust in terms of bias and maintain relatively stable efficiency. On the other hand, if π is correct but the outcome model is misspecified, AIPW and WC do not always surpass IPW. In some cases, their incorrect augmentation term might inflate the variance, underscoring that a poorly chosen ϕ_i can do more harm than good.

Finally, when both π and the outcome model are misspecified, the last columns of Table 4.4 show that all three estimators experience considerable bias. In such a fully misspecified environment, no method can consistently recover the true parameters, underscoring that double robustness cannot succeed if neither model is close to correct.

Chapter 5

Conclusion

5.1 Conclusion

In this study, we explored the theoretical and empirical connections between Augmented Inverse Probability Weighting (AIPW) and Weight Calibration (WC) methods. Our findings indicate that, under appropriate conditions—namely, when either the inclusion probability model or the outcome model is correctly specified—both AIPW and WC estimators achieve consistent and efficient estimation. Through simulation experiments, we demonstrated that these two approaches exhibit comparable performance in terms of bias and variance, and that they can even coincide asymptotically when the same auxiliary information is employed.

Despite these promising results, there remain certain limitations to our work. One important concern is that, in many practical scenarios—including both simulations and real-world applications—one must estimate the inclusion probabilities rather than assume they are known or perfectly modeled. This additional layer of estimation introduces extra variability that is not fully captured by the theoretical arguments presented here. Future research should therefore examine how to rigorously account for the variability arising from weight estimation processes, possibly through more comprehensive asymptotic analysis or resampling-based inference. Also, although our simulation designs featured moderate population sizes and correlations, further exploration under varying sample sizes, sampling designs, or strongly correlated covariates would be beneficial to corroborate the generality of our conclusions.

An important avenue for future work is to extend these methods to survival contexts. Previous studies have shown that AIPW can be employed to estimate regression coefficients in Cox proportional hazards models [15] and provided asymptotic results in survival analyses [9]. Meanwhile, [10, 11] have studied the asymptotic properties of weighted likelihood methods in more general semiparametric frameworks, particularly under two-phase sampling designs. Further research has also explored how calibration-type methods can be adapted to nested case-control studies [13], demonstrating notable efficiency gains over simpler inverse probability weighting approaches. Building on these lines of inquiry, it would be natural to investigate how the equivalences and robust properties discussed in this paper translate to survival data, especially in the presence of time-to-event outcomes, censoring mechanisms, and complex sampling schemes.

Supplementary Material

All R code used for our simulation studies is available in the public GitHub repository: `https://github.com/taeyeon98/GraduateThesis_KTY_SNU`.

Bibliography

[1] J. D. Y. Kang and J. L. Schafer, "Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data," *Statistical Science*, vol. 22, no. 4, pp. 523–539, 2007.

[2] D. G. Horvitz and D. J. Thompson, "A generalization of sampling without replacement from a finite universe," *Journal of the American Statistical Association*, vol. 47, p. 663, Dec 1952.

[3] P. R. Rosenbaum and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, vol. 70, no. 1, pp. 41–55, 1983.

[4] J. M. Robins, A. Rotnitzky, and L. P. Zhao, "Estimation of regression coefficients when some regressors are not always observed," *Journal of the American Statistical Association*, vol. 89, no. 427, p. 846, 1994.

[5] J.-C. Deville and C.-E. Sarndal, "Calibration estimators in survey sam-

pling," *Journal of the American Statistical Association*, vol. 87, no. 418, p. 376, 1992.

[6] J. K. Kim and D. Haziza, "Doubly robust inference under missing data in survey sampling," *Statistica Sinica*, vol. 24, no. 1, pp. 375–394, 2014.

[7] T. Lumley, P. A. Shaw, and J. Y. Dai, "Connections between survey calibration estimators and semiparametric models for incomplete data," *International Statistical Review*, vol. 79, no. 2, p. 200–220, 2011.

[8] J. K. Kim and J. N. K. Rao, "Combining data from two independent surveys: A review and some recent results," *TEST*, vol. 21, no. 2, pp. 151–170, 2012.

[9] L. Qi, C. Y. Wang, and R. L. Prentice, "Weighted estimators for proportional hazards regression with missing covariates," *Journal of the American Statistical Association*, vol. 100, no. 472, p. 1250–1263, 2005.

[10] N. E. Breslow and J. A. Wellner, "Weighted likelihood for semiparametric models and two-phase stratified samples, with application to cox regression," *Scandinavian Journal of Statistics*, vol. 34, no. 1, pp. 86–102, 2007.

[11] N. E. BRESLOW and J. A. WELLNER, "A z-theorem with estimated nuisance parameters and correction note for 'weighted likelihood for semiparametric models and two-phase stratified samples, with application to cox regression'," *Scandinavian Journal of Statistics*, vol. 35, no. 1, p. 186–192, 2008.

[12] T. Saegusa and J. A. Wellner, “Weighted likelihood estimation under two-phase sampling,” *Annals of Statistics*, vol. 41, no. 1, pp. 269–295, 2013.

[13] Y. E. Shin, M. H. Gail, and R. M. Pfeiffer, “Assessing risk model calibration with missing covariates,” *Biostatistics*, vol. 23, no. 3, p. 875–890, 2021.

[14] A. W. van der Vaart, *Asymptotic Statistics*. Cambridge University Press, 1998.

[15] C. Y. Wang and H. Y. Chen, “Augmented inverse probability weighted estimator for cox missing covariate regression,” *Biometrics*, vol. 57, no. 2, p. 414–419, 2001.

요약

본 논문은 부분 관측 혹은 선택적 관측 데이터에 적용되는 보강 역확률 가중 (AIPW) 추정량과 가중 보정(WC) 추정량의 이론적·실증적 관련성을 탐색하며, 이들의 이중견고성(double-robustness)을 강조한다. 두 방법은 각각 결과 모델(AIPW) 혹은 보정 제약(WC)을 통해 보조 정보를 활용하는 형태로 달라 보이지만, 관측확률 모델 또는 결과 모델 중 하나만 정확히 지정되어 있어도, AIPW와 WC 추정량은 동일한 영향함수를 갖고 1차 점근적 거동이 일치하여 일관적이고 효율적인 추정량을 제공함을 보인다. 이러한 통찰을 토대로, 일반적 M-추정 문제에서 두 방법의 동등성을 통합하는 이론적 증명을 제시한다. 완전 혹은 부분적으로 맞는 모델을 사용한 시뮬레이션 연구 결과, AIPW와 WC는 편향 및 분산 면에서 거의 동일한 유한표본 성능을 보여주어, 이들의 이중견고성이 확인되었다. 이는 동일한 보조 정보를 활용한다면, 연구자가 어느 방법을 택하더라도 최적성을 크게 훼손하지 않는다는 점을 시사한다. 아울러 보정 기반 혹은 보강된 방법이 이미 활용되고 있는 복잡한 설계나 생존분석(survival analysis) 분야에 이들을 확장할 가능성도 논의한다.

주요어: AIPW (보강된 역확률 가중), 가중보정, 이중견고성, M-추정, 점근적 동등성

학번: 2023-22449