

Causal Inference for Circular Data

Kuan-Hsun Wu¹

Department of Statistics, National Chengchi University

Abstract

In causal inference, a fundamental task is to estimate the effect resulting from a specific treatment, which is often handled with inverse probability weighting. Despite an abundance of attention to the advancement of this task, most articles have focused on linear data rather than circular data, which are measured in angles. In this article, we extend the causal inference framework to accommodate circular data. Specifically, two new treatment effects, average direction treatment effect (ADTE) and average length treatment effect (ALTE), are introduced to offer a proper causal explanation for these data. As the average direction and average length describe the location and concentration of a random sample of circular data, the ADTE and ALTE measure the change in direction and length between two counterfactual outcomes. With inverse probability weighting, we propose estimators that exhibit ideal theoretical properties, which are validated by a simulation study. To illustrate the practical utility of our estimator, we analyze the effect of different job types on dispatchers' sleep patterns using data from Federal Railroad Administration.

Keywords: Causal inference; circular statistics; directional data; mean direction; mean resultant length; inverse probability weighting.

¹Corresponding Author. Email: 110304015@g.nccu.edu.tw

1 Introduction

Circular statistics deals with the data which can be regarded as points on the circumference of unit circle and has been widely applied in various fields. For instance, Nagasaki et al. (2025) modeled the traffic volume data within a day using mixtures of Kato-Jones distributions (Kato and Jones, 2015); Mardia et al. (2007) explored the protein structure using bivariate circular statistics; and Hansen and Mount (1990) utilized smooth techniques on data of *stress field*, which describes how stress moves from point to point, and estimated the direction of stress. In these applications, data are typically observed rather than generated in a well-designed experiment. To correctly infer causality from such circular data, it is necessary to address the issue of confounding, which is a primary concern in the causal inference methodology.

Causal inference is the integration of methods exploring causality in observational studies. Specifically, it investigates the effect on an assigned outcome of interest brought by a treatment, which is usually characterized by the difference in two counterfactual outcomes. To broaden the general applicability of causal inference, a large body of literature has been contributed to accommodate data with complex features. For example, Yi and Chen (2023) and Chen (2020) handled the estimation of average treatment effect with measurement error. Since high dimensionality is common in the big data era, Huang and Chan (2017) focused on conditional average treatment effect where sufficient dimension reduction is taken into account. Moreover, Wu and Chen (2025) presented a flexible framework of estimating a diversity of casual effects where the dependence structure among confounders is described by graphical models. Despite abundant articles applying causal inference to complicated data, there is no special attention on the inferring causality from circular data.

In this article, we take the initiative in exploring causality for circular outcome. In particular, the location and dispersion of circular data are respectively measured by direction and length of its resultant vector. To capture the change in location and dispersion brought by a treatment, we apply inverse probability weighting (IPW) and define two new treatment effects: average length direction effect (ADTE) and average length treatment effect (ALTE). We derive the estimators for ADTE and ALTE under two IPW schemes, Horvitz and Thomp-

son's (1952) version and Hájek's (1971) version. Moreover, we show that the estimators are consistent and asymptotically normal, which is further supported by a simulation study.

We arrange the article as follows. In Section 2, we introduce conditions for circular data and causal inference and define the ADTE and ALTE. The two IPW approaches to estimating them are explained in Section 3, in where the theoretical properties are subsequently given. We also study the estimators numerically and display the simulation results in Section 4. In addition, we exemplify our estimators using Work Schedule and Sleep Patterns data of dispatchers in Section 5. Eventually, we summarize some significant results in Section 6.

2 Treatment Effects for Circular Outcome

In this section, we introduce some notions of circular statistics and discuss the proper accommodation of circular data in causal inference.

2.1 Circular Random Variables

Let Θ be a continuous random variable with probability density function (PDF) $f(\theta)$ and cumulative distribution function (CDF) $F(\theta)$. We say Θ is a circular random variable if the following conditions are satisfied (e.g., Fisher, 1993):

$$(C1) \quad f(\theta) \geq 0.$$

$$(C2) \quad f \text{ is } 2\pi\text{-periodic. That is, } f(\theta) = f(\theta + 2k\pi) \text{ for any integer } k.$$

$$(C3) \quad \int_0^{2\pi} f(\theta) d\theta = 1.$$

$$(C4) \quad F(\theta) = \int_0^\theta f(t) dt.$$

Condition (C1) assures that the PDF of Θ is non-negative, which is common even in a non-circular statistic context. Conditions (C2) and (C3) characterize the cyclic behavior of a circular random variable, which states the period of a cycle is of length 2π . Lastly, although any f meeting (C1)-(C3) are valid PDFs for circular random variables, it is convenient to regulate that the distribution starts at 0, which is stated by the condition on CDF, (C4).

Based on the target of analysis, one may change the starting point to other meaningful values.

A circular random variable is characterized by its p th trigonometric moment, which is defined as, for $p = 1, 2, \dots$,

$$\begin{aligned}\phi_p &\triangleq E\{\exp(i\Theta)\} \\ &= \alpha_p + i\beta_p,\end{aligned}\tag{1}$$

where

$$\alpha_p \triangleq E(\cos p\Theta)\tag{2}$$

is the p th cosine moment of Θ and

$$\beta_p \triangleq E(\sin p\Theta)\tag{3}$$

is the p th sine moment of Θ . Since (1) can be viewed as a vector in the complex plane \mathbb{C} , we defined its length as

$$\rho_p \triangleq (\alpha_p^2 + \beta_p^2)^{1/2}\tag{4}$$

and its direction as

$$\begin{aligned}\mu_p &\triangleq \text{atan2}(\beta_p, \alpha_p) \\ &= \begin{cases} \arctan(\beta_p/\alpha_p) & \text{if } \beta_p > 0, \alpha_p > 0; \\ \arctan(\beta_p/\alpha_p) + \pi, & \text{if } \alpha_p < 0; \\ \arctan(\beta_p/\alpha_p) + 2\pi, & \text{if } \beta_p > 0, \alpha_p < 0. \end{cases}\end{aligned}\tag{5}$$

The quantities (4) and (5) are called the p th mean length and mean direction, respectively. Since the case $p = 1$ is frequently encountered, we simply write $\phi, \alpha, \beta, \rho$ and μ to refer to (1), (2), (3), (4) and (5) with $p = 1$, respectively. In particular, we call ϕ the *mean resultant vector* with *mean resultant direction* $\mu \in [0, 2\pi)$ describing the location at where the distribution centers and *mean resultant length* $\rho \in [0, 1]$ representing the extent of concentration. As $\arctan(\cdot)$ returns values on $[-\pi/2, \pi/2)$, rather than $[0, 2\pi)$, a modified version of $\arctan(\cdot)$, $\text{atan2}(\cdot, \cdot)$, is used in (5) to return correct angles.

Additionally, to perform meaningful analysis of circular data, we require the underlying distribution of the data has a well-defined μ . Hence, the following condition is proposed:

(C5) Let Θ be a circular random variables satisfying (C1)-(C4). We have that $|\alpha| \in (0, 1]$.

2.2 Counterfactual Framework for Circular Outcomes

Let A be a binary indicator of treatment membership, where $A = 1$ suggests membership in the treated group and $A = 0$ suggests membership in the control group, and let $\mathbf{X} = (1, X_1, \dots, X_p)$ be the p -dimensional covariates with intercept and its realization is denoted by \mathbf{x} . We denote the outcome of interest by Θ . Let the observable random sample be $\mathcal{O} = (A_i, \mathbf{X}_i, \Theta_i)_{i=1}^n$. When $A = a$, the potential outcome under treatment condition a is denoted by $\Theta^{(a)}$. Specifically, the observable outcome Θ and potential outcomes $\Theta^{(1)}$ and $\Theta^{(0)}$ follow circular distributions meeting conditions (C1)-(C5). Moreover, for $a \in \{0, 1\}$, the first sine moment, first cosine moment, mean length and mean direction of $\Theta^{(a)}$ are denoted by $\alpha^{(a)}, \beta^{(a)}, \mu^{(a)}$ and $\rho^{(a)}$, respectively.

We define the *average direction treatment effect* (ADTE) as

$$\begin{aligned}\tau &\triangleq \text{atan2}(\beta^{(1)}, \alpha^{(1)}) - \text{atan2}(\beta^{(0)}, \alpha^{(0)}) \\ &= \mu^{(1)} - \mu^{(0)},\end{aligned}\tag{6}$$

which is the difference in the mean direction of two potential outcomes. Similarly, the *average length treatment effect* (ALTE), which captures the change in mean length between two counterfactual outcomes, is given by

$$\begin{aligned}\xi &\triangleq \{(\alpha^{(1)})^2 + (\beta^{(1)})^2\}^{1/2} - \{(\alpha^{(0)})^2 + (\beta^{(0)})^2\}^{1/2} \\ &= \rho^{(1)} - \rho^{(0)}.\end{aligned}\tag{7}$$

In this article, the issue is to estimate (6) and (7), quantities determined by the distribution $\Theta^{(1)}$ and $\Theta^{(0)}$, while the random sample \mathcal{O} always lack either $\Theta^{(1)}$ or $\Theta^{(0)}$ for each instance. To solve this, we apply the inverse probability weighting (IPW; Rosenbaum and Rubin, 1983) approach. It primarily relies on the use of propensity score, which is given by

$$\pi(\mathbf{x}) \triangleq P(A = 1 \mid \mathbf{X} = \mathbf{x}).\tag{8}$$

That is, the propensity score is the conditional probability of receiving treatment given $\mathbf{X} = \mathbf{x}$. To prepare for further developments in estimating (6) and (7), we introduce some common necessary assumptions for causal inference:

(C6) *Consistency*:

The observed outcome Θ is a mixture of the potential outcomes $(\Theta^{(1)}, \Theta^{(0)})$. In other words, we have $\Theta = A\Theta^{(1)} + (1 - A)\Theta^{(0)}$.

(C7) *Strong Ignorability:*

Conditioning on \mathbf{X} , the treatment A is independent of the joint distribution of $(\Theta^{(1)}, \Theta^{(0)})$.

(C8) $0 < \pi(\mathbf{x}) < 1$.

(C9) (A, \mathbf{X}) is independent and identically distributed.

(C10) \mathbf{X} is bounded.

With (C6) - (C10), numerous counterfactual quantities can be identified. The following lemma gives a broad class of identification results.

Lemma 2.1 (Inverse Probability Weighting.) *Let $\Theta, \Theta^{(1)}$ and $\Theta^{(0)}$ be circular random variables satisfying conditions (C1)-(C4) and assume that (C5) - (C9) hold. Suppose $g(\Theta)$ and $k(\mathbf{X})$ are functions depending only on Θ and \mathbf{X} , respectively. We have*

$$E\left(\frac{A}{\pi(\mathbf{X})}g(\Theta)k(\mathbf{X})\right) = E\left(g(\Theta^{(1)})k(\mathbf{X})\right) \quad (9)$$

and

$$E\left(\frac{1-A}{1-\pi(\mathbf{X})}g(\Theta)k(\mathbf{X})\right) = E\left(g(\Theta^{(0)})k(\mathbf{X})\right). \quad (10)$$

With $g(\Theta)$ specified $\cos \Theta$ or $\sin \Theta$ and $k(\mathbf{X}) = 1$, Lemma 2.1 gives the identification of $\alpha^{(a)}$ and $\beta^{(a)}$ for $a = 0, 1$. Since the pair of estimands, (τ, ξ) , is a unique function of $(\alpha^{(1)}, \beta^{(1)}, \alpha^{(0)}, \beta^{(0)})$, it is identifiable. In addition to identification, Lemma 2.1 is useful in the derivation of theoretical properties later shown in Section 3.1. The proof of Lemma 2.1 is placed in Appendix A.

3 Methodology

3.1 Inverse Probability Weighted Estimators

Motivated by Lemma 2.1, we can construct consistent estimators based on the random sample \mathcal{O} . However, in reality, the propensity score $\pi(\cdot)$ is rarely known and should be estimated. In this study, we model the propensity score by logistic regression, which is given by

$$\pi(\mathbf{x}) = \frac{\exp(\mathbf{x}^\top \boldsymbol{\eta})}{1 + \exp(\mathbf{x}^\top \boldsymbol{\eta})}, \quad (11)$$

where $\boldsymbol{\eta} \triangleq (\eta_0, \eta_1, \dots, \eta_p)^\top$. To estimate $\boldsymbol{\eta}$ in (11), we apply maximum likelihood estimation, which yields the estimator $\hat{\boldsymbol{\eta}}$ that solves the following estimating equation

$$\sum_{i=1}^n \psi_{\boldsymbol{\eta}}(A_i, \mathbf{X}_i; \boldsymbol{\eta}) \triangleq \sum_{i=1}^n \{A_i - \pi(\mathbf{X}_i)\} \mathbf{X}_i^\top = 0. \quad (12)$$

Therefore, we can estimate (8) with

$$\hat{\pi}(\mathbf{x}) = \frac{\exp(\mathbf{x}^\top \hat{\boldsymbol{\eta}})}{1 + \exp(\mathbf{x}^\top \hat{\boldsymbol{\eta}})}. \quad (13)$$

Subsequently, we define the inverse probability weight for the treated outcome on the i th instance as

$$w_{i,1}^{(1)} \triangleq \frac{A_i}{n\hat{\pi}(\mathbf{X}_i)} \quad (14)$$

and that for the controlled outcome on the i th instance as

$$w_{i,1}^{(0)} \triangleq \frac{1 - A_i}{n\{1 - \hat{\pi}(\mathbf{X}_i)\}}, \quad (15)$$

with $i = 1, \dots, n$. Based on Lemma 2.1, consistent estimators for $\alpha^{(a)}$ and $\beta^{(a)}$ are given by

$$\hat{\alpha}^{(a)} \triangleq \sum_{i=1}^n w_{i,1}^{(a)} \cos \Theta_i \quad (16)$$

and

$$\hat{\beta}^{(a)} \triangleq \sum_{i=1}^n w_{i,1}^{(a)} \sin \Theta_i, \quad (17)$$

for $a \in \{0, 1\}$, respectively. To obtain the estimator of (6) and (7), we directly plug in the estimators of $\alpha^{(a)}$ and $\beta^{(a)}$, which gives

$$\hat{\tau} \triangleq \text{atan2}(\hat{\beta}^{(1)}, \hat{\alpha}^{(1)}) - \text{atan2}(\hat{\beta}^{(0)}, \hat{\alpha}^{(0)}) \quad (18)$$

and

$$\widehat{\xi} \triangleq [\{\widehat{\alpha}^{(1)}\}^2 + \{\widehat{\beta}^{(1)}\}^2]^{1/2} - [\{\widehat{\alpha}^{(0)}\}^2 + \{\widehat{\beta}^{(0)}\}^2]^{1/2}, \quad (19)$$

respectively. Since the weights (14) and (15) may result in unstable estimators for extreme values of $\widehat{\pi}(\cdot)$, an alternative weight is given by normalizing them, which gives the following weight

$$w_{i,2}^{(a)} \triangleq \left(\sum_{j=1}^n w_{j,1}^{(a)} \right)^{-1} w_{i,1}^{(a)} \quad (20)$$

for $a \in \{0, 1\}$. Accordingly, we define the estimators of $\alpha^{(a)}$, $\beta^{(a)}$ using (20), which can be formulated as

$$\widetilde{\alpha}^{(a)} \triangleq \sum_{i=1}^n w_{i,2}^{(a)} \cos \Theta_i, \quad (21)$$

$$\widetilde{\beta}^{(a)} \triangleq \sum_{i=1}^n w_{i,2}^{(a)} \sin \Theta_i, \quad (22)$$

In a similar manner, (6) and (7) can be estimated with

$$\widetilde{\tau} \triangleq \text{atan2}(\widetilde{\beta}^{(1)}, \widetilde{\alpha}^{(1)}) - \text{atan2}(\widetilde{\beta}^{(0)}, \widetilde{\alpha}^{(0)}) \quad (23)$$

and

$$\widetilde{\xi} \triangleq [\{\widetilde{\alpha}^{(1)}\}^2 + \{\widetilde{\beta}^{(1)}\}^2]^{1/2} - [\{\widetilde{\alpha}^{(0)}\}^2 + \{\widetilde{\beta}^{(0)}\}^2]^{1/2}, \quad (24)$$

respectively. We call the weights (14) and (15) HT weights and the corresponding estimators (16), (17), (18) and (19) HT-type estimators since the use of this type of weight can be dated back to Horvitz and Thompson (1952). Likewise, the weight (20) can be attributed to Hájek (1971) and is referred to as Hájek weight. The estimators (21), (22), (23) and (24) which involves using Hájek weight are called Hájek-type estimators.

We observe that two different weighting schemes lead to identical estimators of τ defined

in (6), which is directly shown by

$$\begin{aligned}
\frac{\tilde{\beta}^{(a)}}{\tilde{\alpha}^{(a)}} &= \frac{\sum_{i=1}^n w_{i,2}^{(a)} \cos \Theta_i}{\sum_{i=1}^n w_{i,2}^{(a)} \sin \Theta_i} \\
&= \frac{\sum_{i=1}^n \left(\sum_{j=1}^n w_{j,1}^{(a)} \right)^{-1} w_{i,1}^{(a)} \cos \Theta_i}{\sum_{i=1}^n \left(\sum_{j=1}^n w_{j,1}^{(a)} \right)^{-1} w_{i,1}^{(a)} \sin \Theta_i} \\
&= \frac{\sum_{i=1}^n w_{i,1}^{(a)} \cos \Theta_i}{\sum_{i=1}^n w_{i,1}^{(a)} \sin \Theta_i} \\
&= \frac{\hat{\beta}^{(a)}}{\hat{\alpha}^{(a)}}.
\end{aligned} \tag{25}$$

for $a \in \{0, 1\}$. Since $\hat{\tau}$ and $\tilde{\tau}$ only depend on the ratios $\hat{\beta}^{(a)}/\hat{\alpha}^{(a)}$ and $\tilde{\beta}^{(a)}/\tilde{\alpha}^{(a)}$, for $a \in \{0, 1\}$, respectively, (25) is sufficient for the result $\hat{\tau} = \tilde{\tau}$. Hence, employing different weighting schemes only results in different estimator of ξ in (7).

3.2 Theoretical Results

To clearly address the properties of IPW estimators, we introduce some notations. Let the vector of nuisance parameters be $\boldsymbol{\omega} = (\alpha^{(1)}, \beta^{(1)}, \alpha^{(0)}, \beta^{(0)})^\top$ and the vector of parameters of interest be $\boldsymbol{\Delta} = (\tau, \xi)^\top$. The HT-type estimator of $\boldsymbol{\omega}$ and $\boldsymbol{\Delta}$ are denoted by $\hat{\boldsymbol{\omega}} \triangleq (\hat{\alpha}^{(1)}, \hat{\beta}^{(1)}, \hat{\alpha}^{(0)}, \hat{\beta}^{(0)})^\top$ and $\hat{\boldsymbol{\Delta}} = (\hat{\tau}, \hat{\xi})^\top$, respectively. Similarly, the Hájek-type estimator of $\boldsymbol{\omega}$ and $\boldsymbol{\Delta}$ are denoted by $\tilde{\boldsymbol{\omega}} \triangleq (\tilde{\alpha}^{(1)}, \tilde{\beta}^{(1)}, \tilde{\alpha}^{(0)}, \tilde{\beta}^{(0)})^\top$ and $\tilde{\boldsymbol{\Delta}} = (\tilde{\tau}, \tilde{\xi})^\top$, respectively. Given a positive integer q , we write the zero-valued column vector of length q as $\mathbf{0}_q$. Additionally, the Fisher information matrix of $\boldsymbol{\eta}$ is denoted by $\mathcal{A} \triangleq E[\pi(\mathbf{X})\{1 - \pi(\mathbf{X})\}\mathbf{X}^\top \mathbf{X}]$. We begin with stating the large sample property of nuisance parameters.

Theorem 3.1 *Under conditions (C1)-(C10), we have that, as $n \rightarrow \infty$,*

$$(i) \quad n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{d} N(\mathbf{0}_{1+p}, \mathcal{A}^{-1});$$

$$(ii) \quad n^{1/2}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}) \xrightarrow{d} N(\mathbf{0}_4, -\mathcal{B}\mathcal{A}^{-1}\mathcal{B}^\top + \mathcal{C});$$

$$(iii) \ n^{1/2}(\tilde{\omega} - \omega) \xrightarrow{d} N(\mathbf{0}_4, -\mathfrak{B}\mathcal{A}^{-1}\mathfrak{B}^\top + \mathfrak{C}),$$

where " \xrightarrow{d} " denotes convergence in distribution and the components of variance-covariance matrix $\mathcal{B}, \mathcal{C}, \mathfrak{B}$ and \mathfrak{C} are given in Appendix B.

Theorem 3.1 tackles the consistency and asymptotic normality of IPW estimators of nuisance parameters under two weighting schemes. Firstly, (i) states that the coefficients of logistic regression are consistently estimated. Secondly, HT estimators in (ii) and Hájek estimators in (iii), which use the estimated propensity score involving (i), are consistent and asymptotically normal. With nuisance parameters handled by Theorem 3.1, we can obtain the estimators of Δ by regarding it as a function of the vector of nuisance parameter ω . Accordingly, we arrive at the following theorem by applying delta method.

Theorem 3.2 *Under conditions (C1)-(C10), we have that, as $n \rightarrow \infty$,*

$$(i) \ n^{1/2}(\hat{\Delta} - \Delta) \xrightarrow{d} N(\mathbf{0}_2, \Sigma_{\text{HT}});$$

$$(ii) \ n^{1/2}(\tilde{\Delta} - \Delta) \xrightarrow{d} N(\mathbf{0}_2, \Sigma_{\text{Hájek}}),$$

where the variance-covariance matrices Σ_{HT} and $\Sigma_{\text{Hájek}}$ and the entries of them are shown in Appendix C.

With Theorem 3.2, we confirm that the proposed estimators form decent estimators of Δ and can be used to estimate ADTE (6) and ALTE in (7).

4 Numerical Studies

4.1 Simulation Design

We explore the numerical behaviors of estimators (18), (19), (23) and (24). Let the dimension of covariates be $p = 3$. We consider three different sample sizes $n = 250, 500$ or 1000 . The vector of covariate is generated as $\mathbf{X} = (1, X_1, X_2, X_3)^\top$, where X_1, X_2 and X_3 independently follow the Beta(2, 1) distribution.

When \mathbf{X} is generated, the treatment A is generated by the following model

$$\text{logit} \{P(A = 1|\mathbf{X})\} = 1 + X_1 + X_2 + X_3. \quad (26)$$

When generating potential outcomes, we consider four different scenarios where either ADTE or ALTE is subject to the confounders \mathbf{X} . Let $\text{WC}(\mu, \rho)$ denote the wrapped Cauchy distribution with parameters mean direction μ and mean resultant length ρ . The details of the parameter specification is described as follows.

(i) Scenario 1: Heterogeneous ADTE and constant ALTE.

$$\Theta^{(1)} \mid \mathbf{X} \sim \text{WC}\left(X_1 + X_2 + X_3, \frac{5}{6}\right); \Theta^{(0)} \mid \mathbf{X} \sim \text{WC}\left(\frac{1}{2}(X_1 + X_2 + X_3), \frac{2}{3}\right)$$

(ii) Scenario 2: Constant ADTE and heterogeneous ALTE.

$$\Theta^{(1)} \mid \mathbf{X} \sim \text{WC}\left(1, \frac{1}{3}(X_1 + X_2 + X_3)\right); \Theta^{(0)} \mid \mathbf{X} \sim \text{WC}\left(0, \frac{1}{4}(X_1 + X_2 + X_3)\right)$$

(iii) Scenario 3: Heterogeneous ADTE and heterogeneous ALTE.

$$\begin{aligned} \Theta^{(1)} \mid \mathbf{X} &\sim \text{WC}\left(X_1 + X_2 + X_3, \frac{1}{3}(X_1 + X_2 + X_3)\right); \\ \Theta^{(0)} \mid \mathbf{X} &\sim \text{WC}\left(\frac{1}{2}(X_1 + X_2 + X_3), \frac{1}{4}(X_1 + X_2 + X_3)\right) \end{aligned}$$

Finally, the observed random variable Θ_i is acquired by the consistency assumption, which gives

$$\Theta_i = A\Theta_i^{(1)} + (1 - A)\Theta_i^{(0)} \quad (27)$$

for $i = 1, \dots, n$. Consequently, we obtain the IID sample \mathcal{O} of size n . In all scenarios, the true ADTE and ALTE are set to 1 and 1/6, respectively. For demonstration purpose, Figure 1 illustrates the potential outcomes with different circular distribution.

For each setting, we generate 1000 datasets and obtain the estimate of (18), (19), (23) and (24). To examine the performance of the estimators derived by the proposed methods, we compute several indexes including the biases (BIAS), standard errors (S.E.), mean squared errors (MSE) and coverage rate (CR) using the produced 1000 estimates.

4.2 Simulation Result

Table 1 collects the simulation results across all four scenarios. Overall, the numerical results offer strong empirical evidence that the proposed estimators can consistently estimate the

parameter, regardless of heterogeneity. Moreover, since the coverage rates for all cases are close to 95%, the asymptotic normality shown in Theorem 3.2 is verified numerically.

For τ defined in (6), both the HT and Hájek estimators shown in (18) and (23) yield identical point estimates, which aligns with the theoretical finding established in (25). Across scenarios with either constant or heterogeneous directional treatment effects, the estimators of (6) exhibit decreasing bias and MSE as the sample size increases. As for ξ in (7), we observe that proposed estimators can offer precise estimates for all scenarios. Furthermore, the HT estimator (19) and the Hájek-type estimator (24) demonstrate opposite characteristics in small sample sizes. Although the HT-type estimator (19) tends to yield less biased estimates, it exhibits a larger variance. Conversely, the Hájek-type estimator (24) provides estimates with more stability, yet they are relatively more biased. To summarize, the simulation results confirm the equivalence of the HT and Hájek estimators for (6) and other large sample properties established in Theorem 3.2.

In addition, we comment that the characteristics of two types of estimators of (7) exhibit connections with the average treatment effect literature. Firstly, the Hájek estimator (24) generally shows greater stability than the HT-type estimator (19), which is in accordance with the numerical studies of Lunceford and Davidian (2004). Secondly, when estimating (7) in a small sample size, choosing one estimator over the other is a bias-variance tradeoff, which is a relationship stated in Khan and Ugander (2023). However, when the sample size n is sufficiently large, the performance of the HT-type estimator (19) resembles that of the Hájek-type estimator (24), and the bias-variance tradeoff becomes negligible.

5 Data Analysis

In this section, we focus on analyzing the Work Schedules and Sleep Pattern Survey Data of Railroad Dispatchers (Federal Railroad Administration, 2008) with the proposed methods in Section 3. Each participant contributes his/her demographic information and a daily record of sleep schedule for the study. For comprehensive description of the survey methodology, one can visit the website <https://railroads.dot.gov/program-areas/human-factors/work-schedules-and-sleep-pattern-survey-data> for official information and the access

of full dataset.

After necessary pre-processing, there are 439 participants at total. The dataset contains a variety of variables, including age (**Age_Group**, 1 as 20-29 years old, 2 as 30-39 years old, 3 as 40-49 years old, 4 as 50-59 years old and 6 as older than 59), sex (**Sex**, 1 as male and 2 as female), the count of total years in the present job (**Total_years_present_job**, in years), job types of dispatchers (**Job_type**), self-rated health status (**Health_Status**, 1 as poor, 2 as fair, 3 as good and 4 as excellent), marital status (**Marital_Status**, 1 as single and 2 as married), whether or not ingesting caffeinated drink (**Caff_Beverages**, 1 as yes and 2 as no), the time of falling asleep (**Time_fell_asleep**, in 24-hour clock form HH:MM with HH and MM being the hour and minute), self-rated frequency of feeling mentally or physically drained after work (**Mentally_Drained** and **Phys_Drained**, 1 as never, 2 as occasionally, 3 as frequently and 4 as always) and the count of life events in the last six months (**Total_life_events**), including birth of child, death of spouse, martial or financial woes, and other events.

We are interested in the effect of different job types of dispatchers on the time of falling asleep. In this dataset, there are two types of dispatchers: trick dispatchers and assistant chief dispatcher. The former regulates access to a designated territory and the latter is responsible of supervising and assisting the former. Moreover, we express the time of falling asleep as radians using the relationship 2π radians = 24 hours. Using the notation in Section 2, the treatment variable A is set as **Job_type** with 1 being assistant chief dispatcher and 0 being the trick dispatcher and the outcome variable Θ is set as **Time_fell_asleep**, which is transformed as radians. Other mentioned variables which are not selected as treatment or outcome are considered confounders \mathbf{X} .

Our goal is to estimate ADTE defined in (6) and ALTE defined in (7). The estimated ADTE using (18) and (23) is -0.243 radians, or equivalently, -55.691 minutes. Please note that we adopt the polar coordinate system in this article, which is counterclockwise, whereas the conventional 24-hour clock is clockwise. In this context, a negative ADTE estimate indicates that the treated group tends to fall asleep later. For ALTE, the HT-type estimator (19) yields an estimate of 0.445 while the Hájek type estimator gives an estimates 0.184 . Since both estimators produce positive values, this supports that the assistant chief dispatchers exhibit more similarity in their falling-asleep times. In conclusion, the estimators

suggests that the assistant chief dispatchers fall asleep approximately one hour later than trick dispatchers on average, and the falling-asleep times among the trick dispatchers are more dispersed than those among assistant chief dispatchers. For graphical illustration, the plots of the estimated resultant vector of two counterfactual distributions are placed at Figure 2.

6 Summary

In this article, we initiate to integrate circular statistics into the counterfactual framework. In particular, we introduce the ADTE and ALTE as population causal measures for circular data and, consequently, propose the corresponding estimators under HT and Hájek-type of inverse probability weighting. From the theoretical perspective, the proposed estimators enjoy several desirable properties such as consistency and asymptotic normality, which are further verified by the simulation study.

This article broadens the scope of causal inference by clarifying new types of causal effects for circular data as outcome variable. The contribution can be viewed as a critical advance and extension in both methodology and application. The proposed method is an extension of the commonly-used counterfactual framework and allows analysts to flexibly investigate causality in time, direction, or other types of data which can be placed on a unit circle.

Acknowledgement

The author has notified FRA for use of the Work schedules and sleep patterns of railroad dispatchers data via e-mail and appreciated Dr. Li-Pang Chen for illuminating discussions.

References

Chen, L. P. (2020). Causal inference for left-truncated and right-censored data with covariate measurement error. *Computational and Applied Mathematics*, 39, 1-27.

- Federal Railroad Administration. (2008). *Data Files: Work schedules and sleep patterns of railroad dispatchers*. Washington, DC: U.S. Department of Transportation.
- Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge University Press, Cambridge.
- Hájek, J. (1971). Comment on "An Essay on the Logical Foundations of Survey Sampling, Part One," in *The Foundations of Survey Sampling*, Godambe, V.P. and Sprott, D.A. eds., 236, Holt, Rinehart, and Winston.
- Hansen, K. M. and Mount, V. S. (1990). Smoothing and extrapolation of crustal stress orientation measurements. *Journal of Geophysical Research: Solid Earth*, 95(B2), 1155-1165.
- Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260), 663-685.
- Huang, M. Y. and Chan, K. C. G. (2017). Joint sufficient dimension reduction and estimation of conditional and average treatment effects. *Biometrika*, 104(3), 583-596.
- Kato, S. and Jones, M. C. (2015). A tractable and interpretable four-parameter family of unimodal distributions on the circle. *Biometrika*, 102(1), 181-190.
- Khan, S. and Ugander, J. (2023). Adaptive normalization for IPW estimation. *Journal of Causal Inference*, 11, 20220019.
- Lunceford, J. K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine*, 23, 2937-2960
- Nagasaki, K., Kato, S., Nakanishi, W. and Jones, M. C. (2025). Traffic count data analysis using mixtures of Kato–Jones distributions. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 74(2), 352-372.
- Mardia, K. V., Taylor, C. C. and Subramaniam, G. K. (2007). Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data. *Biometrics*, 63(2), 505-

- Rosenbaum, P. R. and Rubin, D. B. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
- Wu, K. H. and Chen, L. P. (2025). A unified approach for estimating various treatment effects in causal inference. arXiv preprint [arXiv:2503.22616](https://arxiv.org/abs/2503.22616).
- Yi, G. Y. and Chen, L. P. (2023). Estimation of the average treatment effect with variable selection and measurement error simultaneously addressed for potential confounders. *Statistical Methods in Medical Research*, 32, 691-711.

Table 1: Simulation results of ADTE and ALTE under Scenario 1-3.

n	Estimand	Method	Scenario 1				Scenario 2				Scenario 3			
			BIAS	S.E.	MSE	CR	BIAS	S.E.	MSE	CR	BIAS	S.E.	MSE	CR
250	τ	HT	0.006	0.369	0.136	0.949	-0.014	0.555	0.308	0.935	0.028	0.603	0.364	0.956
		Hájek	0.006	0.369	0.136	0.949	-0.014	0.555	0.308	0.935	0.028	0.603	0.364	0.956
	ξ	HT	-0.028	0.177	0.032	0.947	-0.056	0.210	0.047	0.941	-0.061	0.216	0.050	0.938
		Hájek	-0.036	0.161	0.027	0.969	-0.059	0.192	0.040	0.954	-0.065	0.197	0.043	0.956
500	τ	HT	-0.005	0.288	0.083	0.952	0.019	0.383	0.147	0.957	-0.016	0.414	0.172	0.950
		Hájek	-0.005	0.288	0.083	0.952	0.019	0.383	0.147	0.957	-0.016	0.414	0.172	0.950
	ξ	HT	-0.008	0.119	0.014	0.947	-0.030	0.141	0.021	0.950	-0.018	0.140	0.020	0.947
		Hájek	-0.010	0.117	0.014	0.960	-0.031	0.140	0.020	0.952	-0.020	0.137	0.019	0.951
1000	τ	HT	-0.006	0.245	0.060	0.952	0.000	0.230	0.053	0.951	-0.002	0.315	0.099	0.956
		Hájek	-0.006	0.245	0.060	0.952	0.000	0.230	0.053	0.951	-0.002	0.315	0.099	0.956
	ξ	HT	-0.005	0.087	0.008	0.948	-0.016	0.103	0.011	0.934	-0.011	0.104	0.011	0.945
		Hájek	-0.005	0.085	0.007	0.956	-0.016	0.103	0.011	0.939	-0.011	0.103	0.011	0.947

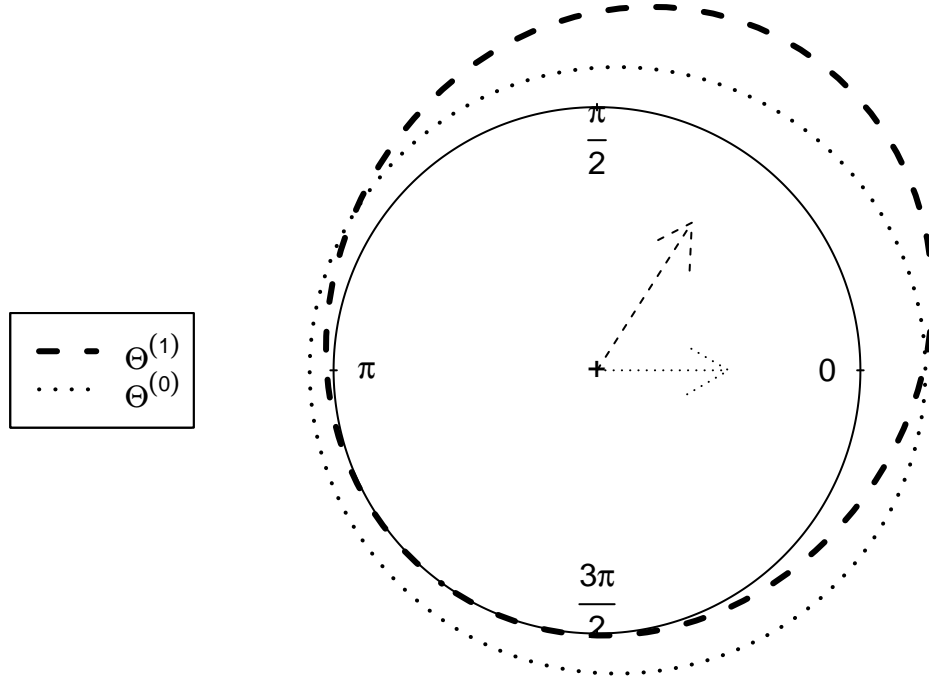


Figure 1: Illustration for simulation design. The ADTE and ALTE are the differences in the angles and lengths between two resultant vectors, respectively. The potential outcomes $\Theta^{(1)}$ and $\Theta^{(0)}$ have mean direction 1 and 0 and mean length $4/6$ and $1/2$, respectively.

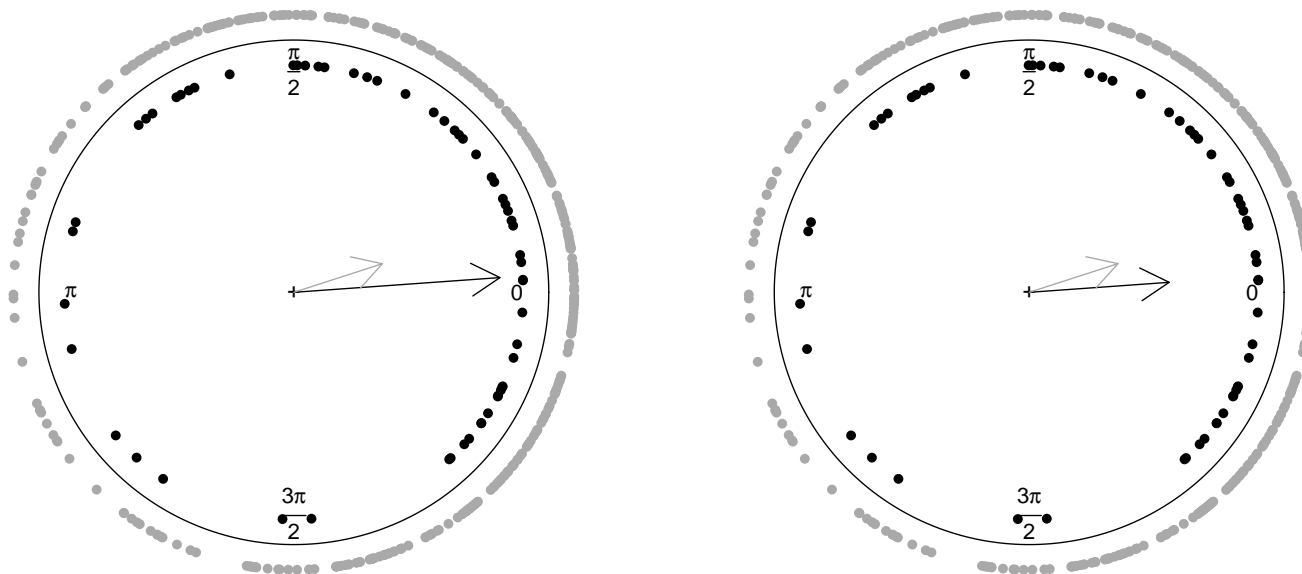


Figure 2: Left panel: the estimated resultant vectors of two counterfactual populations using HT weighting. Right panel: the estimated resultant vectors of two counterfactual populations using Hájek weighting. The black dots and arrows represent the participants with sleeping disorder and the gray dots and arrows represent those without disorder.

July 26, 2025

Supporting Information for “Causal Inference for Circular Data ”

Kuan-Hsun Wu¹

Department of Statistics, National Chengchi University

Abstract

The supporting information contains the details of the lemma and proofs of theorems in the manuscript, entitled “Causal Inference for Circular Data”.

Keywords: Causal inference; circular statistics; directional data; mean direction; mean resultant length; inverse probability weighting.

¹Corresponding Author. Email: 110304015@g.nccu.edu.tw

A Proof of Lemma 2.1

To show (9), we begin with observing that

$$\begin{aligned}
E\left(Ag(\Theta) \mid \mathbf{X}\right) &= E\left(1 \times g(\Theta) \mid \mathbf{X}, A = 1\right)P(A = 1 \mid \mathbf{X}) \\
&\quad + E\left(0 \times g(\Theta) \mid \mathbf{X}, A = 0\right)P(A = 0 \mid \mathbf{X}) \\
&= E\left(g(A\Theta^{(1)} + (1 - A)\Theta^{(0)}) \mid \mathbf{X}, A = 1\right)\pi(\mathbf{X}) \\
&= E\left(g(\Theta^{(1)}) \mid \mathbf{X}, A = 1\right)\pi(\mathbf{X}) \\
&= E\left(g(\Theta^{(1)}) \mid \mathbf{X}\right)\pi(\mathbf{X}),
\end{aligned} \tag{A.1}$$

where the first equality is obtained by partitioning the expectation with events $A = 1$ and $A = 0$, the second and third equality are applications of the consistency assumption (C5) and the fourth equality utilizes condition (C6). Moreover, applying the law of iterated expectations to the left-hand side of (9) yields

$$\begin{aligned}
E\left(\frac{A}{\pi(\mathbf{X})}g(\Theta)k(\mathbf{X})\right) &= E\left\{\frac{k(\mathbf{X})}{\pi(\mathbf{X})}E\left(Ag(\Theta) \mid \mathbf{X}\right)\right\} \\
&= E\left\{\frac{k(\mathbf{X})}{\pi(\mathbf{X})}E\left(g(\Theta^{(1)}) \mid \mathbf{X}\right)\pi(\mathbf{X})\right\} \\
&= E\left\{k(\mathbf{X})E\left(g(\Theta^{(1)}) \mid \mathbf{X}\right)\right\} \\
&= E\left\{E\left(g(\Theta^{(1)})k(\mathbf{X}) \mid \mathbf{X}\right)\right\} \\
&= E\left(g(\Theta^{(1)})k(\mathbf{X})\right),
\end{aligned} \tag{A.2}$$

where the second equality utilized (A.1). Hence, (9) follows. With $A/\pi(\mathbf{X})$ replaced by $(1 - A)/\{1 - \pi(\mathbf{X})\}$, (10) can be proved in a similar manner. This gives Lemma 2.1. \square

B Proof of Theorem 3.1

In this section, we prove the statement regarding the limiting distribution of estimators $\hat{\omega}$ and $\tilde{\omega}$ of the nuisance parameter ω using M-estimation (Stefanski and Boos, 2002). The

asymptotic results for the HT-type and Hájek-type estimators are handled separately. Since $\widehat{\omega}$ and $\widetilde{\omega}$ involve using $\widehat{\eta}$ as a part of estimate, we collectively tackle their asymptotic distribution. We define the true parameter as $\boldsymbol{\vartheta} \triangleq (\boldsymbol{\eta}^\top, \boldsymbol{\omega}^\top)^\top$ with its HT estimator defined as $\widehat{\boldsymbol{\vartheta}} \triangleq (\widehat{\boldsymbol{\eta}}^\top, \widehat{\boldsymbol{\omega}}^\top)^\top$ and its Hájek estimator defined as $\widetilde{\boldsymbol{\vartheta}} \triangleq (\widetilde{\boldsymbol{\eta}}^\top, \widetilde{\boldsymbol{\omega}}^\top)^\top$.

Part 1: HT-type estimators.

We represent the estimators in the form of estimating equations. The estimating equation for the coefficients of logistic regression is readily presented in (12). As for $\alpha^{(1)}, \beta^{(1)}, \alpha^{(0)}$ and $\beta^{(0)}$, the HT-type estimators in (16) and (17) with $a \in \{0, 1\}$ are the solutions to

$$\sum_{i=1}^n \psi_{\alpha^{(1)}}^{\text{HT}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \alpha^{(1)}) \triangleq \sum_{i=1}^n \left(\frac{A_i}{\pi(\mathbf{X}_i)} \cos \Theta_i - \alpha^{(1)} \right) = 0, \quad (\text{B.1})$$

$$\sum_{i=1}^n \psi_{\beta^{(1)}}^{\text{HT}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \beta^{(1)}) \triangleq \sum_{i=1}^n \left(\frac{A_i}{\pi(\mathbf{X}_i)} \sin \Theta_i - \beta^{(1)} \right) = 0, \quad (\text{B.2})$$

$$\sum_{i=1}^n \psi_{\alpha^{(0)}}^{\text{HT}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \alpha^{(0)}) \triangleq \sum_{i=1}^n \left(\frac{1 - A_i}{1 - \pi(\mathbf{X}_i)} \cos \Theta_i - \alpha^{(0)} \right) = 0 \quad (\text{B.3})$$

and

$$\sum_{i=1}^n \psi_{\beta^{(0)}}^{\text{HT}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \beta^{(0)}) \triangleq \sum_{i=1}^n \left(\frac{1 - A_i}{1 - \pi(\mathbf{X}_i)} \sin \Theta_i - \beta^{(0)} \right) = 0. \quad (\text{B.4})$$

Combining (12), (B.1) - (B.4) leads to

$$\sum_{i=1}^n \psi_{\boldsymbol{\vartheta}}^{\text{HT}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\vartheta}) \triangleq \begin{bmatrix} \sum_{i=1}^n \psi_{\boldsymbol{\eta}}(A_i, \mathbf{X}_i; \boldsymbol{\eta}) \\ \sum_{i=1}^n \psi_{\alpha^{(1)}}^{\text{HT}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \alpha^{(1)}) \\ \sum_{i=1}^n \psi_{\beta^{(1)}}^{\text{HT}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \beta^{(1)}) \\ \sum_{i=1}^n \psi_{\alpha^{(0)}}^{\text{HT}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \alpha^{(0)}) \\ \sum_{i=1}^n \psi_{\beta^{(0)}}^{\text{HT}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \beta^{(0)}) \end{bmatrix} = \mathbf{0}_{5+p}. \quad (\text{B.5})$$

Therefore, the HT-type estimator $\widehat{\boldsymbol{\vartheta}}$ is the solution to (B.5). By Stefanski and Boos (2002), as $n \rightarrow \infty$, the asymptotic distribution of this estimator is given by

$$n^{1/2}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \xrightarrow{d} N(\mathbf{0}_{5+p}, \gamma_{\text{HT}}(\boldsymbol{\vartheta})), \quad (\text{B.6})$$

where $\gamma_{\text{HT}}(\boldsymbol{\vartheta}) = A_{\text{HT}}(\boldsymbol{\vartheta})^{-1} B_{\text{HT}}(\boldsymbol{\vartheta})^{-1} \{A_{\text{HT}}(\boldsymbol{\vartheta})^{-1}\}^\top$ with matrices

$$\begin{aligned}
A_{\text{HT}}(\boldsymbol{\vartheta}) &= -E \left\{ \frac{\partial}{\partial \boldsymbol{\vartheta}^\top} \psi_{\boldsymbol{\vartheta}}^{\text{HT}}(A, \mathbf{X}, \Theta; \boldsymbol{\vartheta}) \right\} \\
&= \begin{bmatrix} -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\boldsymbol{\eta}} \right) & \mathbf{0}_{1+p} & \mathbf{0}_{1+p} & \mathbf{0}_{1+p} & \mathbf{0}_{1+p} \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\alpha^{(1)}}^{\text{HT}} \right) & -E \left(\frac{\partial}{\partial \alpha^{(1)}} \psi_{\alpha^{(1)}}^{\text{HT}} \right) & 0 & 0 & 0 \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\beta^{(1)}}^{\text{HT}} \right) & 0 & -E \left(\frac{\partial}{\partial \beta^{(1)}} \psi_{\beta^{(1)}}^{\text{HT}} \right) & 0 & 0 \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\alpha^{(0)}}^{\text{HT}} \right) & 0 & 0 & -E \left(\frac{\partial}{\partial \alpha^{(0)}} \psi_{\alpha^{(0)}}^{\text{HT}} \right) & 0 \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\beta^{(1)}}^{\text{HT}} \right) & 0 & 0 & 0 & E \left(\frac{\partial}{\partial \beta^{(0)}} \psi_{\beta^{(0)}}^{\text{HT}} \right) \end{bmatrix} \\
&= \begin{bmatrix} -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\boldsymbol{\eta}} \right) & \mathbf{0}_{1+p} & \mathbf{0}_{1+p} & \mathbf{0}_{1+p} & \mathbf{0}_{1+p} \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\alpha^{(1)}}^{\text{HT}} \right) & 1 & 0 & 0 & 0 \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\beta^{(1)}}^{\text{HT}} \right) & 0 & 1 & 0 & 0 \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\alpha^{(0)}}^{\text{HT}} \right) & 0 & 0 & 1 & 0 \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\beta^{(1)}}^{\text{HT}} \right) & 0 & 0 & 0 & 1 \end{bmatrix} \\
&\triangleq \begin{bmatrix} \mathbf{a}_{11} & \mathbf{0}_{(1+p) \times 4} \\ \mathbf{a}_{21} & \mathbf{I}_4 \end{bmatrix}, \tag{B.7}
\end{aligned}$$

$$\{A_{\text{HT}}(\boldsymbol{\vartheta})\}^{-1} = \begin{bmatrix} \mathbf{a}_{11}^{-1} & \mathbf{0}_{(1+p) \times 4} \\ -\mathbf{a}_{11}^{-1} \mathbf{a}_{21} & \mathbf{I}_4 \end{bmatrix}, \tag{B.8}$$

$$\begin{aligned}
B_{\text{HT}}(\boldsymbol{\vartheta}) &= E \left\{ \psi_{\boldsymbol{\vartheta}}^{\text{HT}}(A, \mathbf{X}, \Theta; \boldsymbol{\vartheta}) \psi_{\boldsymbol{\vartheta}}^{\text{HT}}(A, \mathbf{X}, \Theta; \boldsymbol{\vartheta})^{\top} \right\} \\
&= E \left[\begin{array}{c|cccc} \psi_{\boldsymbol{\eta}} \psi_{\boldsymbol{\eta}}^{\top} & \psi_{\boldsymbol{\eta}} \psi_{\alpha(1)}^{\text{HT}} & \psi_{\boldsymbol{\eta}} \psi_{\beta(1)}^{\text{HT}} & \psi_{\boldsymbol{\eta}} \psi_{\alpha(0)}^{\text{HT}} & \psi_{\boldsymbol{\eta}} \psi_{\beta(0)}^{\text{HT}} \\ \hline \psi_{\alpha(1)}^{\text{HT}} \psi_{\boldsymbol{\eta}}^{\top} & \psi_{\alpha(1)}^{\text{HT}} \psi_{\alpha(1)}^{\text{HT}} & \psi_{\alpha(1)}^{\text{HT}} \psi_{\beta(1)}^{\text{HT}} & \psi_{\alpha(1)}^{\text{HT}} \psi_{\alpha(0)}^{\text{HT}} & \psi_{\alpha(1)}^{\text{HT}} \psi_{\beta(0)}^{\text{HT}} \\ \psi_{\beta(1)}^{\text{HT}} \psi_{\boldsymbol{\eta}}^{\top} & \psi_{\beta(1)}^{\text{HT}} \psi_{\alpha(1)}^{\text{HT}} & \psi_{\beta(1)}^{\text{HT}} \psi_{\beta(1)}^{\text{HT}} & \psi_{\beta(1)}^{\text{HT}} \psi_{\alpha(0)}^{\text{HT}} & \psi_{\beta(1)}^{\text{HT}} \psi_{\beta(0)}^{\text{HT}} \\ \psi_{\alpha(0)}^{\text{HT}} \psi_{\boldsymbol{\eta}}^{\top} & \psi_{\alpha(0)}^{\text{HT}} \psi_{\alpha(1)}^{\text{HT}} & \psi_{\alpha(0)}^{\text{HT}} \psi_{\beta(1)}^{\text{HT}} & \psi_{\alpha(0)}^{\text{HT}} \psi_{\alpha(0)}^{\text{HT}} & \psi_{\alpha(0)}^{\text{HT}} \psi_{\beta(0)}^{\text{HT}} \\ \psi_{\beta(0)}^{\text{HT}} \psi_{\boldsymbol{\eta}}^{\top} & \psi_{\beta(0)}^{\text{HT}} \psi_{\alpha(1)}^{\text{HT}} & \psi_{\beta(0)}^{\text{HT}} \psi_{\beta(1)}^{\text{HT}} & \psi_{\beta(0)}^{\text{HT}} \psi_{\alpha(0)}^{\text{HT}} & \psi_{\beta(0)}^{\text{HT}} \psi_{\beta(0)}^{\text{HT}} \end{array} \right] \quad (\text{B.9}) \\
&\triangleq \begin{bmatrix} \mathbf{b}_{11} & \mathbf{b}_{21}^{\top} \\ \mathbf{b}_{21} & \mathbf{b}_{22} \end{bmatrix}
\end{aligned}$$

and \mathbf{I}_k being the $k \times k$ identity matrix for a positive integer k . A detailed investigation of the components of the submatrices is provided in Appendix E of Wu and Chen (2025). Here, we present some core results in it to simplify the variance-covariance matrix $\gamma_{\text{HT}}(\boldsymbol{\vartheta})$. Firstly, owing to Lemma 7.3.11 of Casella and Berger (2002), we have that $\mathbf{a}_{11} = \mathbf{b}_{11} = E[\pi(\mathbf{X}; \boldsymbol{\eta})\{1 - \pi(\mathbf{X}; \boldsymbol{\eta})\} \mathbf{X}^{\top} \mathbf{X}]$.

Secondly, we explore the relationship between \mathbf{a}_{21} and \mathbf{b}_{21} . On one hand, we have that

$$\begin{aligned}
\mathbf{a}_{21} &\triangleq - \left[E \left(\frac{\partial}{\partial \boldsymbol{\eta}^{\top}} \psi_{\alpha(1)}^{\text{HT}} \right) \quad E \left(\frac{\partial}{\partial \boldsymbol{\eta}^{\top}} \psi_{\beta(1)}^{\text{HT}} \right) \quad E \left(\frac{\partial}{\partial \boldsymbol{\eta}^{\top}} \psi_{\alpha(0)}^{\text{HT}} \right) \quad E \left(\frac{\partial}{\partial \boldsymbol{\eta}^{\top}} \psi_{\beta(0)}^{\text{HT}} \right) \right]^{\top} \\
&= \begin{bmatrix} E \left(\frac{1 - \pi(\mathbf{X})}{\pi(\mathbf{X})} A \cos \Theta \mathbf{X} \right) \\ E \left(\frac{1 - \pi(\mathbf{X})}{\pi(\mathbf{X})} A \sin \Theta \mathbf{X} \right) \\ E \left(\frac{\pi(\mathbf{X})}{1 - \pi(\mathbf{X})} (1 - A) \cos \Theta \mathbf{X} \right) \\ E \left(\frac{\pi(\mathbf{X})}{1 - \pi(\mathbf{X})} (1 - A) \sin \Theta \mathbf{X} \right) \end{bmatrix} \\
&= \begin{bmatrix} E \left[\{1 - \pi(\mathbf{X})\} \cos \Theta^{(1)} \mathbf{X}^{\top} \right] \\ E \left[\{1 - \pi(\mathbf{X})\} \sin \Theta^{(1)} \mathbf{X}^{\top} \right] \\ E \left\{ \pi(\mathbf{X}) \cos \Theta^{(0)} \mathbf{X}^{\top} \right\} \\ E \left\{ \pi(\mathbf{X}) \sin \Theta^{(0)} \mathbf{X}^{\top} \right\} \end{bmatrix},
\end{aligned}$$

where the second step is established by Lemma 2.1. On the other hand, for the submatrix

$\mathbf{b}_{21} \triangleq E \left[\psi_{\alpha^{(1)}}^{\text{HT}} \psi_{\boldsymbol{\eta}}^{\top} \quad \psi_{\beta^{(1)}}^{\text{HT}} \psi_{\boldsymbol{\eta}}^{\top} \quad \psi_{\alpha^{(0)}}^{\text{HT}} \psi_{\boldsymbol{\eta}}^{\top} \quad \psi_{\beta^{(0)}}^{\text{HT}} \psi_{\boldsymbol{\eta}}^{\top} \right]^{\top}$, we examine that

$$\begin{aligned}
E(\psi_{\alpha^{(1)}}^{\text{HT}} \psi_{\boldsymbol{\eta}}^{\top}) &= E \left[\left(\frac{A}{\pi(\mathbf{X}; \boldsymbol{\eta})} \cos \Theta - \alpha^{(1)} \right) \mathbf{X}^{\top} \{A - \pi(\mathbf{X}; \boldsymbol{\eta})\} \right] \\
&= E \left(\frac{A \mathbf{X}^{\top} \cos \Theta}{\pi(\mathbf{X}; \boldsymbol{\eta})} \right) - E(A \mathbf{X}^{\top} \cos \Theta) - \alpha^{(1)} E(A \mathbf{X}^{\top}) + \alpha^{(1)} E\{\mathbf{X}^{\top} \pi(\mathbf{X}; \boldsymbol{\eta})\} \\
&= E \left[\frac{A \mathbf{X}^{\top} \cos \Theta}{\pi(\mathbf{X}; \boldsymbol{\eta})} \{1 - \pi(\mathbf{X}; \boldsymbol{\eta})\} \right] - \alpha^{(1)} E\{\mathbf{X}^{\top} \pi(\mathbf{X}; \boldsymbol{\eta})\} + \alpha^{(1)} E\{\mathbf{X}^{\top} \pi(\mathbf{X}; \boldsymbol{\eta})\} \\
&= E \left[\{1 - \pi(\mathbf{X})\} \cos \Theta^{(1)} \mathbf{X}^{\top} \right],
\end{aligned} \tag{B.10}$$

where we utilizes the tower property of expectation on the third term of the second line and Lemma 2.1 on the first term of the third line. The remaining steps only involves basic calculations. With the similar derivation demonstrated in (B.10), we can show that other components in \mathbf{a}_{21} and \mathbf{b}_{21} are equal and hence $\mathbf{a}_{21} = \mathbf{b}_{21}$.

Finally, we thoroughly investigate the components in \mathbf{b}_{22} . Consider $a, t \in \{0, 1\}$. For $a = 1$ and $t = 0$, we have that

$$\begin{aligned}
E(\psi_{\alpha^{(1)}}^{\text{HT}} \psi_{\alpha^{(0)}}^{\text{HT}}) &= E \left\{ \left(\frac{A}{\pi(\mathbf{X})} \cos \Theta - \alpha^{(1)} \right) \left(\frac{1 - A}{1 - \pi(\mathbf{X})} \cos \Theta - \alpha^{(0)} \right) \right\} \\
&= E \left\{ \frac{A(1 - A)}{\pi(\mathbf{X}) \{1 - \pi(\mathbf{X})\}} \cos^2 \Theta \right\} - \alpha^{(0)} E \left(\frac{A}{\pi(\mathbf{X})} \cos \Theta \right) \\
&\quad - \alpha^{(1)} E \left(\frac{1 - A}{1 - \pi(\mathbf{X})} \cos \Theta \right) + \alpha^{(1)} \alpha^{(0)} \\
&= 0 - \alpha^{(0)} \alpha^{(1)} - \alpha^{(1)} \alpha^{(0)} + \alpha^{(1)} \alpha^{(0)} \\
&= -\alpha^{(1)} \alpha^{(0)},
\end{aligned} \tag{B.11}$$

where the third equality is a result of $A(1 - A) = 0$ and Lemma 2.1 with $g(\Theta) = \cos \Theta$ and other equalities only involve basic calculation of expectations. With similar use of $A(1 - A) = 0$ and Lemma 2.1, one can generalize equation (B.11) to the case $a \neq t$, which gives

$$E(\psi_{\alpha^{(a)}}^{\text{HT}} \psi_{\alpha^{(t)}}^{\text{HT}}) = -\alpha^{(a)} \alpha^{(t)}, \tag{B.12}$$

$$E(\psi_{\alpha^{(a)}}^{\text{HT}} \psi_{\beta^{(t)}}^{\text{HT}}) = -\alpha^{(a)} \beta^{(t)} \tag{B.13}$$

and

$$E(\psi_{\beta^{(a)}}^{\text{HT}} \psi_{\beta^{(t)}}^{\text{HT}}) = -\beta^{(a)} \beta^{(t)}. \quad (\text{B.14})$$

Moreover, we consider the case where $a = t = 1$.

$$\begin{aligned} E(\psi_{\alpha^{(1)}}^{\text{HT}} \psi_{\alpha^{(1)}}^{\text{HT}}) &= E \left\{ \left(\frac{A}{\pi(\mathbf{X})} \cos \Theta - \alpha^{(1)} \right)^2 \right\} \\ &= E \left(\frac{A}{\pi(\mathbf{X})^2} \cos^2 \Theta \right) - 2E \left\{ \left(\frac{A}{\pi(\mathbf{X})} \cos \Theta \right) \right\} \alpha^{(1)} + \{\alpha^{(1)}\}^2 \\ &= E \left(\frac{\cos^2 \Theta^{(1)}}{\pi(\mathbf{X})} \right) - 2\{\alpha^{(1)}\}^2 + \{\alpha^{(1)}\}^2 \\ &= E \left(\frac{\cos^2 \Theta^{(1)}}{\pi(\mathbf{X})} \right) - \{\alpha^{(1)}\}^2. \end{aligned} \quad (\text{B.15})$$

Following the similar procedure in (B.15), the results of other cases where $a = t$ are exhaustively given by

$$E(\psi_{\alpha^{(0)}}^{\text{HT}} \psi_{\alpha^{(0)}}^{\text{HT}}) = E \left(\frac{\sin^2 \Theta^{(1)}}{1 - \pi(\mathbf{X})} \right) - \{\alpha^{(0)}\}^2, \quad (\text{B.16})$$

$$E(\psi_{\alpha^{(1)}}^{\text{HT}} \psi_{\beta^{(1)}}^{\text{HT}}) = E \left(\frac{\cos \Theta^{(1)} \sin \Theta^{(1)}}{\pi(\mathbf{X})} \right) - \alpha^{(1)} \beta^{(1)}, \quad (\text{B.17})$$

$$E(\psi_{\alpha^{(0)}}^{\text{HT}} \psi_{\beta^{(0)}}^{\text{HT}}) = E \left(\frac{\cos \Theta^{(0)} \sin \Theta^{(0)}}{1 - \pi(\mathbf{X})} \right) - \alpha^{(0)} \beta^{(0)}, \quad (\text{B.18})$$

$$E(\psi_{\beta^{(1)}}^{\text{HT}} \psi_{\beta^{(1)}}^{\text{HT}}) = E \left(\frac{\sin^2 \Theta^{(1)}}{\pi(\mathbf{X})} \right) - \{\beta^{(1)}\}^2 \quad (\text{B.19})$$

and

$$E(\psi_{\beta^{(0)}}^{\text{HT}} \psi_{\beta^{(0)}}^{\text{HT}}) = E \left(\frac{\sin^2 \Theta^{(0)}}{1 - \pi(\mathbf{X})} \right) - \{\beta^{(0)}\}^2. \quad (\text{B.20})$$

With all components determined in (B.12)-(B.20), it follows that

$$\begin{aligned}
\mathbf{b}_{22} &= \begin{bmatrix} E\left(\frac{\cos^2 \Theta^{(1)}}{\pi(\mathbf{X})}\right) & E\left(\frac{\cos \Theta^{(1)} \sin \Theta^{(1)}}{\pi(\mathbf{X})}\right) & 0 & 0 \\ E\left(\frac{\cos \Theta^{(1)} \sin \Theta^{(1)}}{\pi(\mathbf{X})}\right) & E\left(\frac{\sin^2 \Theta^{(1)}}{\pi(\mathbf{X})}\right) & 0 & 0 \\ 0 & 0 & E\left(\frac{\sin^2 \Theta^{(1)}}{1 - \pi(\mathbf{X})}\right) & E\left(\frac{\cos \Theta^{(0)} \sin \Theta^{(0)}}{1 - \pi(\mathbf{X})}\right) \\ 0 & 0 & E\left(\frac{\cos \Theta^{(0)} \sin \Theta^{(0)}}{1 - \pi(\mathbf{X})}\right) & E\left(\frac{\sin^2 \Theta^{(0)}}{1 - \pi(\mathbf{X})}\right) \end{bmatrix} \\
&- \begin{bmatrix} \alpha^{(1)}\alpha^{(1)} & \alpha^{(1)}\beta^{(1)} & \alpha^{(1)}\alpha^{(0)} & \alpha^{(1)}\beta^{(0)} \\ \alpha^{(1)}\beta^{(1)} & \beta^{(1)}\beta^{(1)} & \beta^{(1)}\alpha^{(0)} & \beta^{(0)}\beta^{(1)} \\ \alpha^{(1)}\alpha^{(0)} & \alpha^{(1)}\beta^{(0)} & \alpha^{(0)}\alpha^{(0)} & \alpha^{(0)}\beta^{(0)} \\ \beta^{(1)}\alpha^{(0)} & \beta^{(0)}\beta^{(1)} & \alpha^{(0)}\beta^{(0)} & \beta^{(0)}\beta^{(0)} \end{bmatrix} \\
&= \begin{bmatrix} \mathbb{B}^{(1)} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathbb{B}^{(0)} \end{bmatrix} - \boldsymbol{\omega} \boldsymbol{\omega}^\top
\end{aligned} \tag{B.21}$$

where

$$\mathbb{B}^{(1)} \triangleq \begin{bmatrix} E\left(\frac{\cos^2 \Theta^{(1)}}{\pi(\mathbf{X})}\right) & E\left(\frac{\cos \Theta^{(1)} \sin \Theta^{(1)}}{\pi(\mathbf{X})}\right) \\ E\left(\frac{\cos \Theta^{(1)} \sin \Theta^{(1)}}{\pi(\mathbf{X})}\right) & E\left(\frac{\sin^2 \Theta^{(1)}}{\pi(\mathbf{X})}\right) \end{bmatrix}, \tag{B.22}$$

and

$$\mathbb{B}^{(0)} \triangleq \begin{bmatrix} E\left(\frac{\sin^2 \Theta^{(1)}}{1 - \pi(\mathbf{X})}\right) & E\left(\frac{\cos \Theta^{(0)} \sin \Theta^{(0)}}{1 - \pi(\mathbf{X})}\right) \\ E\left(\frac{\cos \Theta^{(0)} \sin \Theta^{(0)}}{1 - \pi(\mathbf{X})}\right) & E\left(\frac{\sin^2 \Theta^{(0)}}{1 - \pi(\mathbf{X})}\right) \end{bmatrix}. \tag{B.23}$$

With (B.8) and (B.9), we express the variance-covariance matrix as

$$\begin{aligned}
\gamma_{\text{HT}}(\boldsymbol{\vartheta}) &= \begin{bmatrix} \mathbf{a}_{11}^{-1} \mathbf{b}_{11} \mathbf{a}_{11}^{-1} & -\mathbf{a}_{11}^{-1} (\mathbf{a}_{21}^\top - \mathbf{b}_{21}^\top) \\ (-\mathbf{a}_{21} + \mathbf{b}_{21}) \mathbf{a}_{11}^{-1} & (\mathbf{a}_{21} - \mathbf{b}_{21}) \mathbf{a}_{11}^{-1} \mathbf{a}_{21}^\top - \mathbf{a}_{21} \mathbf{a}_{11}^{-1} \mathbf{b}_{21}^\top + \mathbf{b}_{22} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{a}_{11}^{-1} & \mathbf{0}_{(1+p) \times 4} \\ \mathbf{0}_{4 \times (1+p)} & -\mathbf{b}_{21} \mathbf{a}_{11}^{-1} \mathbf{b}_{21}^\top + \mathbf{b}_{22} \end{bmatrix}.
\end{aligned} \tag{B.24}$$

By the property of multivariate normal distribution, we have that

$$n^{1/2}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{d} \text{N}(\mathbf{0}_{1+p}, \mathcal{A}^{-1}),$$

and

$$n^{1/2}(\widehat{\boldsymbol{\omega}} - \boldsymbol{\omega}) \xrightarrow{d} N(\mathbf{0}_4, -\mathcal{B}\mathcal{A}\mathcal{B}^\top + \mathcal{C})$$

where $\mathcal{A} \triangleq \mathbf{a}_{11}^{-1}$, $\mathcal{B} \triangleq \mathbf{b}_{21}$ and $\mathcal{C} \triangleq \mathbf{b}_{22}$. This gives results (i) and (ii).

Part 2: Hájek-type estimators

The procedures of deriving asymptotics for Hájek-type estimators are similar with Part 1 of the proof. The only difference is that the estimating equations are different. For Hájek-type estimators (21) and (22) with $a \in \{0, 1\}$, they satisfies the following equations

$$\sum_{i=1}^n \psi_{\alpha^{(1)}}^{\text{Hájek}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \alpha^{(1)}) \triangleq \sum_{i=1}^n \frac{A_i}{\pi(\mathbf{X}_i)} (\cos \Theta_i - \alpha^{(1)}) = 0, \quad (\text{B.25})$$

$$\sum_{i=1}^n \psi_{\beta^{(1)}}^{\text{Hájek}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \beta^{(1)}) \triangleq \sum_{i=1}^n \frac{A_i}{\pi(\mathbf{X}_i)} (\sin \Theta_i - \beta^{(1)}) = 0, \quad (\text{B.26})$$

$$\sum_{i=1}^n \psi_{\alpha^{(0)}}^{\text{Hájek}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \alpha^{(0)}) \triangleq \sum_{i=1}^n \frac{1 - A_i}{1 - \pi(\mathbf{X}_i)} (\cos \Theta_i - \alpha^{(0)}) = 0 \quad (\text{B.27})$$

and

$$\sum_{i=1}^n \psi_{\beta^{(0)}}^{\text{Hájek}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \beta^{(0)}) \triangleq \sum_{i=1}^n \frac{1 - A_i}{1 - \pi(\mathbf{X}_i)} (\sin \Theta_i - \beta^{(0)}) = 0. \quad (\text{B.28})$$

Since the treatment model is handled by logistic regression, the equation (12) is directly adopted in forming the estimating equations. Combining (12) and (B.25) - (B.28) gives

$$\sum_{i=1}^n \psi_{\boldsymbol{\vartheta}}^{\text{Hájek}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\vartheta}) = \begin{bmatrix} \sum_{i=1}^n \psi_{\boldsymbol{\eta}}(A_i, \mathbf{X}_i; \boldsymbol{\eta}) \\ \sum_{i=1}^n \psi_{\alpha^{(1)}}^{\text{Hájek}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \alpha^{(1)}) \\ \sum_{i=1}^n \psi_{\beta^{(1)}}^{\text{Hájek}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \beta^{(1)}) \\ \sum_{i=1}^n \psi_{\alpha^{(0)}}^{\text{Hájek}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \alpha^{(0)}) \\ \sum_{i=1}^n \psi_{\beta^{(0)}}^{\text{Hájek}}(A_i, \mathbf{X}_i, \Theta_i; \boldsymbol{\eta}, \beta^{(0)}) \end{bmatrix} = \mathbf{0}_{5+p}, \quad (\text{B.29})$$

to which $\widetilde{\boldsymbol{\vartheta}}$ is a solution. According to Stefanski and Boos (2002), we have the asymptotic distribution as

$$n^{1/2}(\widetilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}) \xrightarrow{d} N(\mathbf{0}_{5+p}, \gamma_{\text{Hájek}}(\boldsymbol{\vartheta})), \quad (\text{B.30})$$

where $\gamma_{\text{Hájek}}(\boldsymbol{\vartheta}) = A_{\text{Hájek}}(\boldsymbol{\vartheta})^{-1} B_{\text{Hájek}}(\boldsymbol{\vartheta}) \{A_{\text{Hájek}}(\boldsymbol{\vartheta})^{-1}\}^\top$ with matrices given by

$$\begin{aligned}
A_{\text{Hájek}}(\boldsymbol{\vartheta}) &= -E \left\{ \frac{\partial}{\partial \boldsymbol{\vartheta}^\top} \psi_{\boldsymbol{\vartheta}}^{\text{Hájek}}(A, \mathbf{X}, \Theta; \boldsymbol{\vartheta}) \right\} \\
&= \begin{bmatrix} -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\boldsymbol{\eta}} \right) & \mathbf{0}_{1+p} & \mathbf{0}_{1+p} & \mathbf{0}_{1+p} & \mathbf{0}_{1+p} \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\alpha^{(1)}}^{\text{Hájek}} \right) & -E \left(\frac{\partial}{\partial \alpha^{(1)}} \psi_{\alpha^{(1)}}^{\text{Hájek}} \right) & 0 & 0 & 0 \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\beta^{(1)}}^{\text{Hájek}} \right) & 0 & -E \left(\frac{\partial}{\partial \beta^{(1)}} \psi_{\beta^{(1)}}^{\text{Hájek}} \right) & 0 & 0 \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\alpha^{(0)}}^{\text{Hájek}} \right) & 0 & 0 & -E \left(\frac{\partial}{\partial \alpha^{(0)}} \psi_{\alpha^{(0)}}^{\text{Hájek}} \right) & 0 \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\beta^{(1)}}^{\text{Hájek}} \right) & 0 & 0 & 0 & E \left(\frac{\partial}{\partial \beta^{(0)}} \psi_{\beta^{(0)}}^{\text{Hájek}} \right) \end{bmatrix} \\
&= \begin{bmatrix} -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\boldsymbol{\eta}} \right) & \mathbf{0}_{1+p} & \mathbf{0}_{1+p} & \mathbf{0}_{1+p} & \mathbf{0}_{1+p} \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\alpha^{(1)}}^{\text{Hájek}} \right) & 1 & 0 & 0 & 0 \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\beta^{(1)}}^{\text{Hájek}} \right) & 0 & 1 & 0 & 0 \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\alpha^{(0)}}^{\text{Hájek}} \right) & 0 & 0 & 1 & 0 \\ -E \left(\frac{\partial}{\partial \boldsymbol{\eta}^\top} \psi_{\beta^{(1)}}^{\text{Hájek}} \right) & 0 & 0 & 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{a}_{11} & \mathbf{0}_{4 \times (1+p)} \\ \mathbf{a}_{21} & \mathbf{I}_4 \end{bmatrix}, \tag{B.31}
\end{aligned}$$

$$A_{\text{Hájek}}(\boldsymbol{\vartheta})^{-1} = \begin{bmatrix} \mathbf{a}_{11}^{-1} & \mathbf{0}_{4 \times (1+p)} \\ -\mathbf{a}_{11}^{-1} \mathbf{a}_{21} & \mathbf{I}_4 \end{bmatrix} \tag{B.32}$$

and

$$\begin{aligned}
B_{\text{Hájek}}(\boldsymbol{\vartheta}) &= E \left\{ \psi_{\boldsymbol{\vartheta}}^{\text{Hájek}}(A, \mathbf{X}, \Theta; \boldsymbol{\vartheta}) \psi_{\boldsymbol{\vartheta}}^{\text{Hájek}}(A, \mathbf{X}, \Theta; \boldsymbol{\vartheta})^\top \right\} \\
&= E \left[\begin{array}{c|cccc}
\psi_{\boldsymbol{\eta}} \psi_{\boldsymbol{\eta}}^\top & \psi_{\boldsymbol{\eta}} \psi_{\alpha(1)}^{\text{Hájek}} & \psi_{\boldsymbol{\eta}} \psi_{\beta(1)}^{\text{Hájek}} & \psi_{\boldsymbol{\eta}} \psi_{\alpha(0)}^{\text{Hájek}} & \psi_{\boldsymbol{\eta}} \psi_{\beta(0)}^{\text{Hájek}} \\
\hline
\psi_{\alpha(1)}^{\text{Hájek}} \psi_{\boldsymbol{\eta}}^\top & \psi_{\alpha(1)}^{\text{Hájek}} \psi_{\alpha(1)}^{\text{Hájek}} & \psi_{\alpha(1)}^{\text{Hájek}} \psi_{\beta(1)}^{\text{Hájek}} & \psi_{\alpha(1)}^{\text{Hájek}} \psi_{\alpha(0)}^{\text{Hájek}} & \psi_{\alpha(1)}^{\text{Hájek}} \psi_{\beta(0)}^{\text{Hájek}} \\
\psi_{\beta(1)}^{\text{Hájek}} \psi_{\boldsymbol{\eta}}^\top & \psi_{\beta(1)}^{\text{Hájek}} \psi_{\alpha(1)}^{\text{Hájek}} & \psi_{\beta(1)}^{\text{Hájek}} \psi_{\beta(1)}^{\text{Hájek}} & \psi_{\beta(1)}^{\text{Hájek}} \psi_{\alpha(0)}^{\text{Hájek}} & \psi_{\beta(1)}^{\text{Hájek}} \psi_{\beta(0)}^{\text{Hájek}} \\
\psi_{\alpha(0)}^{\text{Hájek}} \psi_{\boldsymbol{\eta}}^\top & \psi_{\alpha(0)}^{\text{Hájek}} \psi_{\alpha(1)}^{\text{Hájek}} & \psi_{\alpha(0)}^{\text{Hájek}} \psi_{\beta(1)}^{\text{Hájek}} & \psi_{\alpha(0)}^{\text{Hájek}} \psi_{\alpha(0)}^{\text{Hájek}} & \psi_{\alpha(0)}^{\text{Hájek}} \psi_{\beta(0)}^{\text{Hájek}} \\
\psi_{\beta(1)}^{\text{Hájek}} \psi_{\boldsymbol{\eta}}^\top & \psi_{\beta(0)}^{\text{Hájek}} \psi_{\alpha(1)}^{\text{Hájek}} & \psi_{\beta(0)}^{\text{Hájek}} \psi_{\beta(1)}^{\text{Hájek}} & \psi_{\beta(0)}^{\text{Hájek}} \psi_{\alpha(0)}^{\text{Hájek}} & \psi_{\beta(0)}^{\text{Hájek}} \psi_{\beta(0)}^{\text{Hájek}}
\end{array} \right] \quad (\text{B.33}) \\
&= \begin{bmatrix} \mathbf{b}_{11} & \mathbf{b}_{21}^\top \\ \mathbf{b}_{21} & \mathbf{b}_{22} \end{bmatrix}.
\end{aligned}$$

In the spirit of Wu and Chen (2025), we have that

$$\mathbf{a}_{21} = \mathbf{b}_{21} = \begin{bmatrix} E \left[\{1 - \pi(\mathbf{X})\} \cos(\Theta^{(1)} - \alpha^{(1)}) \mathbf{X} \right] \\ E \left[\{1 - \pi(\mathbf{X})\} \sin(\Theta^{(1)} - \beta^{(1)}) \mathbf{X} \right] \\ E \left[\pi(\mathbf{X}) \cos(\Theta^{(0)} - \alpha^{(0)}) \mathbf{X} \right] \\ E \left[\pi(\mathbf{X}) \sin(\Theta^{(0)} - \beta^{(0)}) \mathbf{X} \right] \end{bmatrix}. \quad (\text{B.34})$$

In Part 1, the result $\mathbf{a}_{11} = \mathbf{b}_{11}$ is already given. Combining these results gives the simplification of the asymptotic variance-covariance matrix, which is expressed as

$$\gamma_{\text{Hájek}}(\boldsymbol{\vartheta}) = \begin{bmatrix} \mathbf{a}_{11}^{-1} & \mathbf{0}_{(1+p) \times 4} \\ \mathbf{0}_{4 \times (1+p)} & -\mathbf{b}_{21} \mathbf{a}_{11}^{-1} \mathbf{b}_{21}^\top + \mathbf{b}_{22} \end{bmatrix}. \quad (\text{B.35})$$

We take a closer look at \mathbf{b}_{22} , which is defined as

$$\begin{aligned}
\mathbf{b}_{22} &\triangleq \begin{bmatrix} \psi_{\alpha(1)}^{\text{Hájek}} \psi_{\alpha(1)}^{\text{Hájek}} & \psi_{\alpha(1)}^{\text{Hájek}} \psi_{\beta(1)}^{\text{Hájek}} & \psi_{\alpha(1)}^{\text{Hájek}} \psi_{\alpha(0)}^{\text{Hájek}} & \psi_{\alpha(1)}^{\text{Hájek}} \psi_{\beta(0)}^{\text{Hájek}} \\ \psi_{\beta(1)}^{\text{Hájek}} \psi_{\alpha(1)}^{\text{Hájek}} & \psi_{\beta(1)}^{\text{Hájek}} \psi_{\beta(1)}^{\text{Hájek}} & \psi_{\beta(1)}^{\text{Hájek}} \psi_{\alpha(0)}^{\text{Hájek}} & \psi_{\beta(1)}^{\text{Hájek}} \psi_{\beta(0)}^{\text{Hájek}} \\ \psi_{\alpha(0)}^{\text{Hájek}} \psi_{\alpha(1)}^{\text{Hájek}} & \psi_{\alpha(0)}^{\text{Hájek}} \psi_{\beta(1)}^{\text{Hájek}} & \psi_{\alpha(0)}^{\text{Hájek}} \psi_{\alpha(0)}^{\text{Hájek}} & \psi_{\alpha(0)}^{\text{Hájek}} \psi_{\beta(0)}^{\text{Hájek}} \\ \psi_{\beta(0)}^{\text{Hájek}} \psi_{\alpha(1)}^{\text{Hájek}} & \psi_{\beta(0)}^{\text{Hájek}} \psi_{\beta(1)}^{\text{Hájek}} & \psi_{\beta(0)}^{\text{Hájek}} \psi_{\alpha(0)}^{\text{Hájek}} & \psi_{\beta(0)}^{\text{Hájek}} \psi_{\beta(0)}^{\text{Hájek}} \end{bmatrix} \\
&= \begin{bmatrix} \mathfrak{B}^{(1)} & \mathbf{0}_{2 \times 2} \\ \mathbf{0}_{2 \times 2} & \mathfrak{B}^{(0)} \end{bmatrix}, \quad (\text{B.36})
\end{aligned}$$

where

$$\mathfrak{B}^{(1)} \triangleq \begin{bmatrix} E\left\{\frac{(\cos \Theta^{(1)} - \alpha^{(1)})^2}{\pi(\mathbf{X})}\right\} & E\left\{\frac{(\cos \Theta^{(1)} - \alpha^{(1)})(\sin \Theta^{(1)} - \beta^{(1)})}{\pi(\mathbf{X})}\right\} \\ E\left\{\frac{(\cos \Theta^{(1)} - \alpha^{(1)})(\sin \Theta^{(1)} - \beta^{(1)})}{\pi(\mathbf{X})}\right\} & E\left\{\frac{(\sin \Theta^{(1)} - \beta^{(1)})^2}{\pi(\mathbf{X})}\right\} \end{bmatrix} \quad (\text{B.37})$$

and

$$\mathfrak{B}^{(0)} \triangleq \begin{bmatrix} E\left\{\frac{(\cos \Theta^{(0)} - \alpha^{(0)})^2}{1 - \pi(\mathbf{X})}\right\} & E\left\{\frac{(\cos \Theta^{(0)} - \alpha^{(0)})(\sin \Theta^{(0)} - \beta^{(0)})}{1 - \pi(\mathbf{X})}\right\} \\ E\left\{\frac{(\cos \Theta^{(0)} - \alpha^{(0)})(\sin \Theta^{(0)} - \beta^{(0)})}{1 - \pi(\mathbf{X})}\right\} & E\left\{\frac{(\sin \Theta^{(0)} - \beta^{(0)})^2}{1 - \pi(\mathbf{X})}\right\} \end{bmatrix}. \quad (\text{B.38})$$

Owing to the property of multivariate normal distribution, we obtain that

$$n^{1/2}(\tilde{\boldsymbol{\omega}} - \boldsymbol{\omega}) \xrightarrow{d} \text{N}(\mathbf{0}_4, -\mathbf{b}_{21}\mathbf{a}_{11}^{-1}\mathbf{b}_{21}^\top + \mathbf{b}_{22})$$

with the components of variance-covariance given in (B.34), (B.36)-(B.38). Hence, the result (ii) is proved. \square

C Proof of Theorem 3.2

Define

$$\mathbf{f}(\boldsymbol{\omega}) \triangleq \begin{bmatrix} \text{atan2}(\beta^{(1)}, \alpha^{(1)}) - \text{atan2}(\beta^{(0)}, \alpha^{(0)}) \\ [\{\alpha^{(1)}\}^2 + \{\beta^{(1)}\}^2]^{1/2} - [\{\alpha^{(0)}\}^2 + \{\beta^{(0)}\}^2]^{1/2} \end{bmatrix}. \quad (\text{C.1})$$

Note that (C.1) gives $\hat{\Delta} = \mathbf{f}(\hat{\boldsymbol{\omega}})$, $\tilde{\Delta} = \mathbf{f}(\tilde{\boldsymbol{\omega}})$ and $\Delta = \mathbf{f}(\boldsymbol{\omega})$. The Jacobian matrix of (C.1) is derived by

$$\begin{aligned} \mathbf{J} &\triangleq \begin{bmatrix} \frac{-\beta^{(1)}}{(\alpha^{(1)})^2 + (\beta^{(1)})^2} & \frac{\alpha^{(1)}}{(\alpha^{(1)})^2 + (\beta^{(1)})^2} & \frac{\beta^{(0)}}{(\alpha^{(0)})^2 + (\beta^{(0)})^2} & \frac{-\alpha^{(0)}}{(\alpha^{(0)})^2 + (\beta^{(0)})^2} \\ \frac{\alpha^{(1)}}{\{(\alpha^{(1)})^2 + (\beta^{(1)})^2\}^{0.5}} & \frac{\beta^{(1)}}{\{(\alpha^{(1)})^2 + (\beta^{(1)})^2\}^{0.5}} & \frac{-\alpha^{(0)}}{\{(\alpha^{(0)})^2 + (\beta^{(0)})^2\}^{0.5}} & \frac{-\beta^{(0)}}{\{(\alpha^{(0)})^2 + (\beta^{(0)})^2\}^{0.5}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{-\beta^{(1)}}{\{\rho^{(1)}\}^2} & \frac{\alpha^{(1)}}{\{\rho^{(1)}\}^2} & \frac{\beta^{(0)}}{\{\rho^{(0)}\}^2} & \frac{-\alpha^{(0)}}{\{\rho^{(0)}\}^2} \\ \frac{\alpha^{(1)}}{\rho^{(1)}} & \frac{\beta^{(1)}}{\rho^{(1)}} & \frac{-\alpha^{(0)}}{\rho^{(0)}} & \frac{-\beta^{(0)}}{\rho^{(0)}} \end{bmatrix} \\ &\triangleq \begin{bmatrix} \mathbf{D}^{(1)}\mathbb{J}^{(1)} & \mathbf{D}^{(0)}\mathbb{J}^{(0)} \end{bmatrix}, \end{aligned}$$

where

$$\mathbb{J}^{(1)} \triangleq \begin{bmatrix} -\beta^{(1)} & \alpha^{(1)} \\ \alpha^{(1)} & \beta^{(1)} \end{bmatrix},$$

$$\mathbb{J}^{(0)} \triangleq \begin{bmatrix} \beta^{(0)} & -\alpha^{(0)} \\ -\alpha^{(0)} & -\beta^{(0)} \end{bmatrix}$$

and, for $a \in \{0, 1\}$,

$$\mathbf{D}^{(a)} \triangleq \begin{bmatrix} \{\rho^{(a)}\}^{-2} & 0 \\ 0 & \{\rho^{(a)}\}^{-1} \end{bmatrix}.$$

Directly applying multivariate delta method on HT-type estimators yields

$$n^{1/2}(\widehat{\Delta} - \Delta) \xrightarrow{d} N(\mathbf{0}_2, \Sigma_{\text{HT}}),$$

where

$$\begin{aligned} \Sigma_{\text{HT}} &\triangleq \mathbf{J}(-\mathbf{b}_{21}\mathbf{a}_{11}^{-1}\mathbf{b}_{21}^\top + \mathbf{b}_{22})\mathbf{J}^\top \\ &= -\mathbf{J}\mathbf{b}_{21}\mathbf{a}_{11}^{-1}\mathbf{b}_{21}^\top\mathbf{J}^\top + \mathbf{J}\mathbf{b}_{22}\mathbf{J}^\top \\ &= -\mathcal{J}_{\text{HT}}\mathcal{A}^{-1}\mathcal{J}_{\text{HT}}^\top + \mathbf{D}^{(1)}\mathbb{J}^{(1)}(\mathbb{B}^{(1)} - \boldsymbol{\omega}_1\boldsymbol{\omega}_1^\top)\mathbb{J}^{(1)}\mathbf{D}^{(1)} \\ &\quad + \mathbf{D}^{(0)}\mathbb{J}^{(0)}(\mathbb{B}^{(0)} - \boldsymbol{\omega}_0\boldsymbol{\omega}_0^\top)\mathbb{J}^{(0)}\mathbf{D}^{(0)} - 2\mathbf{D}^{(1)}\mathbb{J}^{(1)}\boldsymbol{\omega}_1\boldsymbol{\omega}_0^\top\mathbb{J}^{(0)}\mathbf{D}^{(0)} \end{aligned}$$

with $\mathcal{J}_{\text{HT}} \triangleq \mathbf{J}\mathbf{b}_{21}$, $\boldsymbol{\omega}_1 \triangleq (\alpha^{(1)}, \beta^{(1)})^\top$ and $\boldsymbol{\omega}_0 \triangleq (\alpha^{(0)}, \beta^{(0)})^\top$. This leads to result (i). Again, applying multivariate delta method on Hájek-type estimators gives

$$n^{1/2}(\widetilde{\Delta} - \Delta) \xrightarrow{d} N(\mathbf{0}_2, \Sigma_{\text{Hájek}}),$$

where

$$\begin{aligned} \Sigma_{\text{Hájek}} &\triangleq \mathbf{J}(-\mathbf{b}_{21}\mathbf{a}_{11}^{-1}\mathbf{b}_{21}^\top + \mathbf{b}_{22})\mathbf{J}^\top \\ &= -\mathbf{J}\mathbf{b}_{21}\mathbf{a}_{11}^{-1}\mathbf{b}_{21}^\top\mathbf{J}^\top + \mathbf{J}\mathbf{b}_{22}\mathbf{J}^\top \\ &= -\mathcal{J}_{\text{Hájek}}\mathcal{A}\mathcal{J}_{\text{Hájek}}^\top + \mathbf{D}^{(1)}\mathbb{J}^{(1)}\mathfrak{B}^{(1)}\mathbb{J}^{(1)}\mathbf{D}^{(1)} + \mathbf{D}^{(0)}\mathbb{J}^{(0)}\mathfrak{B}^{(0)}\mathbb{J}^{(0)}\mathbf{D}^{(0)} \end{aligned}$$

with $\mathcal{J}_{\text{Hájek}} \triangleq \mathbf{J}\mathbf{b}_{21}$. This brings about result (ii). □

References

Casella, G. and Berger, R. L. (2002). *Statistical Inference, Second Edition*. Duxbury Pacific Grove, CA.

- Stefanski, L. A. and Boos, D. D. (2002). The calculus of M-estimation. *The American Statistician*, 56(1), 29-38.
- Wu, K. H. and Chen, L. P. (2025). A unified approach for estimating various treatment effects in causal inference. arXiv preprint [arXiv:2503.22616](https://arxiv.org/abs/2503.22616).