Supplementary Materials of When Is Causal Inference Possible? A Statistical Test for Unmeasured Confounding

Muye Liu, Jun Xie

August 21, 2025

Contents

1 Proof of Lemma 3.1

Lemma 3.1. *The inverse probability weighted (IPW) estimator for the RCT and observational data have the following asymptotic distributions:*

$$\sqrt{m}(\hat{\omega}^o - \beta_A - \beta_U \delta_A) \xrightarrow{d} \mathcal{N}(0, V_o)$$

where $V_o/m = \mathrm{Var}(\hat{\omega}^o)$. *Analogously,*

$$\sqrt{n}(\hat{\omega}^r - \beta_A) \xrightarrow{d} \mathcal{N}(0, V_r)$$

where $V_r/n = \mathrm{Var}(\hat{\omega}^r)$.

When $e(X) \equiv p$ is known, the Hájek IPW estimator [1]

$$\hat{\omega} = \frac{\sum_i^n A_i Y_i / e(X_i)}{\sum_i^n A_i / e(X_i)} - \frac{\sum_i^n (1 - A_i) Y_i / (1 - e(X_i))}{\sum_i^n (1 - A_i)/(1 - e(X_i))}$$

reduces to the difference in sample means; therefore, by the classical central limit theorem, it is asymptotically normal. We then consider the case in which $e(\cdot)$ is estimated (e.g., via logistic regression) and derive the corresponding variances.

Proof. Consider the observational study with data $\mathcal{D}^o = (A_o, X, Y) \sim P^o$ and RCT with $\mathcal{D}^r = (A_r, X, Y) \sim P^r$.

Step 1 (Asymptotic distribution of the Hájek means with known $e \equiv p_r$) In Step 1, we use the notation of RCT to carry out the proof; The derivation for the observational data (OBS) is analogous. Define

$$\mu_1 = \mathbb{E}_{P^r}(Y \mid A_r = 1), \qquad \mu_0 = \mathbb{E}_{P^r}(Y \mid A_r = 0).$$

Therefore $\omega^r = \mu_1 - \mu_0 = \beta_A$. Because $e(X) \equiv p_r$,

$$\hat{\mu}_1 = \frac{\sum_{i=1}^n A_{ri} Y_i / p_r}{\sum_{i=1}^n A_{ri} / p_r}, \qquad \hat{\mu}_0 = \frac{\sum_{i=1}^n (1 - A_{ri}) Y_i / p_r}{\sum_{i=1}^n (1 - A_{ri}) / p_r},$$

$\hat{\omega}^r = \hat{\mu}_1 - \hat{\mu}_0$. Let $N_1 = \sum_{i=1}^n A_{ri} \sim \mathrm{Bin}(n, p_r)$ and $N_0 = n - N_1$, we have

$$\hat{\mu}_1 = \frac{1}{N_1} \sum_{i=1}^n A_{ri} Y_i, \qquad \hat{\mu}_0 = \frac{1}{N_0} \sum_{i=1}^n (1 - A_{ri}) Y_i.$$

Given (N_1, N_0), by CLT we have

$$\sqrt{N_1}(\hat{\mu}_1 - \mu_1) \xrightarrow{d} \mathcal{N}(0, S_{1r}), \qquad \sqrt{N_0}(\hat{\mu}_0 - \mu_0) \xrightarrow{d} \mathcal{N}(0, S_{0r}),$$

where $S_{1r} = \mathrm{Var}_{P^r}(Y \mid A_r = 1)$ and $S_{0r} = \mathrm{Var}_{P^r}(Y \mid A_r = 0)$. Because the treated and control subjects are independent, we have

$$\mathrm{Cov}(\hat{\mu}_1, \hat{\mu}_0 \mid N_1, N_0) = 0. \tag{1}$$

Combined Equation (1) with the fact that

$$\frac{N_1}{n} \xrightarrow{p} p_r, \qquad \frac{N_0}{n} \xrightarrow{p} 1 - p_r,$$

by the Slutsky's theorem we have

$$\sqrt{n}(\hat{\omega}^r - \omega^r) = \sqrt{\frac{n}{N_1}} \sqrt{N_1}(\hat{\mu}_1 - \mu_1) - \sqrt{\frac{n}{N_0}} \sqrt{N_0}(\hat{\mu}_0 - \mu_0) \xrightarrow{d} \mathcal{N}(0, V_r)$$

where $V_r = S_{1r}/p_r + S_{0r}/(1 - p_r)$.

In the OBS, analogously, with $e(X) \equiv p_o$ we have

$$\sqrt{m}(\hat{\omega}^o - \omega^o) \xrightarrow{d} \mathcal{N}(0, V_o)$$

where $V_o = S_{1o}/p_o + S_{0o}/(1 - p_o)$.

Step 2 (Compute S_0, S_1) Recall our underlying model that

$$Y = \beta_0 + \beta_A A + \beta_X X + \beta_U U + \epsilon,$$
$$U = \delta_0 + \delta_A A_o + \delta_X X + \nu$$

Let $\tilde{\beta}_X = \beta_X + \beta_U \delta_X$, we consider the RCT and OBS separately.

- In the OBS, substitute $U = \delta_0 + \delta_A A_o + \delta_X X + \nu$ into $Y = \beta_0 + \beta_A A_o + \beta_X X + \beta_U U + \epsilon$, we have

  $$Y = (\beta_0 + \beta_U \delta_0) + (\beta_A + \beta_U \delta_A) A_o + \tilde{\beta}_X X + \beta_U \nu + \epsilon,$$

  so

  $$\mathrm{Var}(Y \mid A_o = a) = \tilde{\beta}_X^2 \sigma_X^2 + \beta_U^2 \sigma_U^2 + \sigma_\epsilon^2 \equiv S_o$$

  for $a = 0, 1$. Therefore,

  $$V_o = \frac{S_o}{p_o(1 - p_o)} = \frac{\tilde{\beta}_X^2 \sigma_X^2 + \beta_U^2 \sigma_U^2 + \sigma_\epsilon^2}{p_o(1 - p_o)}$$

  when $e \equiv p_o$ is known.

- In the RCT, substitute $U = \delta_0 + \delta_A A_o + \delta_X X + \nu$ into $Y = \beta_0 + \beta_A A_r + \beta_X X + \beta_U U + \epsilon$, we have
$$Y = (\beta_0 + \beta_U \delta_0) + \beta_A A_r + \beta_U \delta_A A_o + \tilde{\beta}_X X + \beta_U \nu + \epsilon,$$
so
$$\mathrm{Var}(Y \mid A_r = a) = \tilde{\beta}_X^2 \sigma_X^2 + \beta_U^2 \left(\sigma_U^2 + \delta_A^2 p_o(1-p_o)\right) + \sigma_\epsilon^2 \equiv S_r$$
for $a = 0, 1$. Therefore,
$$V_r = \frac{S_r}{p_r(1-p_r)} = \frac{\tilde{\beta}_X^2 \sigma_X^2 + \beta_U^2 \left(\sigma_U^2 + \delta_A^2 p_o(1-p_o)\right) + \sigma_\epsilon^2}{p_r(1-p_r)}$$
when $e \equiv p_r$ is known.

Step 3 (Estimating propensity scores by intercept-only logistic (constant) does not change V_r) Based on our underlying model, the propensity score in the RCT is a constant and should be p_r, If we use an intercept-only logistic model to estimate the propensity score such as $e(X) = \hat{p}_r$, we still have

$$\hat{\mu}_1 = \frac{\sum_{i=1}^n A_{ri} Y_i / \hat{p}_r}{\sum_{i=1}^n A_{ri} / \hat{p}_r} = \frac{1}{N_1} \sum_{i=1}^n A_{ri} Y_i, \qquad \hat{\mu}_0 = \frac{\sum_{i=1}^n (1 - A_{ri}) Y_i / \hat{p}_r}{\sum_{i=1}^n (1 - A_{ri}) / \hat{p}_r} = \frac{1}{N_0} \sum_{i=1}^n (1 - A_{ri}) Y_i$$

which do not depend on $\hat{p}_r$. Therefore, the expression of V_r remains the same.

Step 4 (Estimating propensity scores by the logistic model $A \sim X$ changes V_o) For the OBS data, the propensity score is typically estimated by the logistic regression of A on X. This will change the form of the variance V_o. More specifically, denote

$$e_\gamma(x) = \mathrm{expit}(\gamma_0 + \gamma_1 x) = \frac{1}{1 + \mathrm{e}^{-(\gamma_0 + \gamma_1 \mathrm{x})}}.$$

Lunceford & Davidian [2] investigated the variance for an IPW estimator with estimated propensity scores by the Huber–White "sandwich" method [3]. Stack the two Hájek moment equations with the logistic scores for $\theta = (\mu_1, \mu_0, \gamma_0, \gamma_1)^\top$:

$$\psi_i(\theta) = \begin{pmatrix} \dfrac{A_i}{e_\gamma(X_i)} (Y_i - \mu_1) \\ \dfrac{1 - A_i}{1 - e_\gamma(X_i)} (Y_i - \mu_0) \\ A_i - e_\gamma(X_i) \\ X_i (A_i - e_\gamma(X_i)) \end{pmatrix}, \qquad \frac{1}{n_1} \sum_{i=1}^{n_1} \psi_i(\hat{\theta}) = 0.$$

At $\theta = \hat{\theta}$, the Jacobian ("bread") $A = \mathbb{E}[\partial_\theta \psi_i]$ and the covariance ("meat") $B = \mathbb{E}[\psi_i \psi_i^\top]$ are

$$A = \begin{pmatrix} -1 & 0 & 0 & -(1-p_o)\, \tilde{\beta}_X \sigma_X^2 \\ 0 & -1 & 0 & p_o\, \tilde{\beta}_X \sigma_X^2 \\ 0 & 0 & -p_o(1-p_o) & -p_o(1-p_o)\mu_x \\ 0 & 0 & -p_o(1-p_o)\mu_x & -p_o(1-p_o)(\mu_x^2 + \sigma_X^2) \end{pmatrix},$$

$$B = \begin{pmatrix} \frac{S_o}{p_o} & 0 & 0 & (1-p_o)\, \tilde{\beta}_X \sigma_X^2 \\ 0 & \frac{S_o}{1-p_o} & 0 & -p_o\, \tilde{\beta}_X \sigma_X^2 \\ 0 & 0 & p_o(1-p_o) & p_o(1-p_o)\mu_x \\ (1-p_o)\, \tilde{\beta}_X \sigma_X^2 & -p_o\, \tilde{\beta}_X \sigma_X^2 & p_o(1-p_o)\mu_x & p_o(1-p_o)(\mu_x^2 + \sigma_X^2) \end{pmatrix}.$$

The covariance is $\Sigma_\theta = \frac{1}{m}A^{-1}BA^{-\top}$. For $\omega^o = \mu_1 - \mu_0$, use $L = (1, -1, 0, 0)$, so $\mathrm{Var}(\hat{\omega}^o) \approx L\Sigma_\theta L^\top$ where

$$L\Sigma_\theta L^\top = \frac{1}{m}\left\{\frac{S_o}{p_o} + \frac{S_o}{1-p_o} - \frac{\tilde{\beta}_X^2\sigma_X^2}{p_o(1-p_o)}\right\} = \frac{1}{m}\,\frac{\beta_U^2\sigma_U^2 + \sigma_\epsilon^2}{p_o(1-p_o)}.$$

Therefore

$$V_o = \frac{\beta_U^2\sigma_U^2 + \sigma_\epsilon^2}{p_o(1-p_o)}$$

□

2 Proof of Theorem 3.2

Theorem 3.2. *We can perform a z-test at level α by rejecting H_0 when $|Z| > z_{1-\frac{\alpha}{2}}$. The power function of the test will have the form:*

$$\Phi\left(-z_{1-\frac{\alpha}{2}} - \frac{\delta_A}{h}\right) + \Phi\left(\frac{\delta_A}{h} - z_{1-\frac{\alpha}{2}}\right) + \mathcal{O}(n^{-\frac{1}{2}}),$$

where Φ is the CDF of the standard normal distribution and $h = h(\delta_A, \delta_X, \sigma_v, \sigma_\epsilon, n)$ is defined as follows with $\kappa = \frac{m}{n}$.

$$h(\delta_A, \delta_X, \sigma_v, \sigma_\epsilon, n) = \left\{\frac{(c+\delta_X)^2\sigma_X^2 + \delta_A^2\, p_o(1-p_o) + \sigma_U^2 + \eta}{np_r(1-p_r)} \;+\; \frac{\sigma_U^2 + \eta}{\kappa n p_o(1-p_o)}\right\}^{1/2}.$$

Proof. We will derive the finite-sample power by applying the Edgeworth expansion [4] on the CDF of the test statistics. We begin by recalling some notations:

$$\omega^r = \beta_A, \qquad \omega^o = \beta_A + \beta_U\,\delta_A \qquad \Delta = \omega^o - \omega^r = \beta_U\,\delta_A,$$

$$\begin{aligned}
c &= \frac{\beta_X}{\beta_U}, \qquad \eta = \frac{\sigma_\epsilon^2}{\beta_U^2},\\
V_r &= \frac{\tilde{\beta}_X^2\sigma_X^2 + \beta_U^2\left(\sigma_U^2 + \delta_A^2 p_o(1-p_o)\right) + \sigma_\epsilon^2}{p_r(1-p_r)}\\
&= \frac{\beta_U^2}{p_r(1-p_r)}\Big[(c+\delta_X)^2\sigma_X^2 + \delta_A^2\, p_o(1-p_o) + \sigma_U^2 + \eta\Big],\\
V_o &= \frac{\beta_U^2\sigma_U^2 + \sigma_\epsilon^2}{p_o(1-p_o)} = \frac{\beta_U^2}{p_o(1-p_o)}\Big(\sigma_U^2 + \eta\Big).
\end{aligned}$$

$$Z \;=\; \frac{\hat{\omega}^o - \hat{\omega}^r}{\sqrt{\dfrac{\hat{V}_r}{n} + \dfrac{\hat{V}_o}{m}}}, \qquad H_0:\ \omega^o = \omega^r, \quad \text{reject if } |Z| > z_{1-\alpha/2}.$$

Define

$$V = \frac{V_r}{n} + \frac{V_o}{m}, \qquad \xi = \frac{\Delta}{\sqrt{V}}.$$

Substituting the above variances gives,

$$V \;=\; \beta_U^2\Big\{\frac{(c+\delta_X)^2\sigma_X^2 + \delta_A^2\, p_o(1-p_o) + \sigma_U^2 + \eta}{np_r(1-p_r)} \;+\; \frac{\sigma_U^2 + \eta}{mp_o(1-p_o)}\Big\},$$

$$\xi = \frac{\beta_U\,\delta_A}{\sqrt{V}} = \frac{\delta_A}{\sqrt{\dfrac{(c+\delta_X)^2\sigma_X^2 + [\delta_A^2\,p_o(1-p_o)+\sigma_U^2]+\eta}{np_r(1-p_r)} + \dfrac{\sigma_U^2+\eta}{mp_o(1-p_o)}}}.$$

Note that β_U has been canceled between the numerator and denominator and thus the effect size ξ does not depend on β_U anymore. Define the skewnesses

$$\gamma_{1,r} = \frac{\mathbb{E}_r[(\hat{\omega}^r - \omega^r)^3]}{V_r^{3/2}}, \quad \gamma_{1,o} = \frac{\mathbb{E}_o[(\hat{\omega}^o - \omega^o)^3]}{V_o^{3/2}}.$$

Then the third order cumulant is

$$\begin{aligned}
\kappa_3(\hat{\omega}^o - \hat{\omega}^r) = \kappa_3(\hat{\omega}^o) - \kappa_3(\hat{\omega}^r) &= \frac{\gamma_{1,r}\,V_o^{3/2}}{n^2} - \frac{\gamma_{1,o}\,V_r^{3/2}}{m^2} = O(n^{-2}+m^{-2}),\\
V &= O(n^{-1}+m^{-1}),\\
\gamma_{\text{eff}} &\equiv \frac{\kappa_3(\hat{\omega}^o - \hat{\omega}^r)}{V^{3/2}} = O\big((n^{-1}+m^{-1})^{1/2}\big).
\end{aligned}$$

The Edgeworth expansion [4] for the CDF of Z is

$$F_Z(z) = \Phi(z) + \frac{\gamma_{\text{eff}}}{6}\,(1-z^2)\,\phi(z) + O(n^{-1}+m^{-1})$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution and $\phi(\cdot)$ is the density of standard normal distribution. Since the decision rule is to reject H_0 when $|Z| > z_{1-\alpha/2}$, the finite-sample power function can be define as $\pi_{n,m}(\Delta) = P_Z(|Z| > z_{1-\alpha/2} \mid \omega^o - \omega^r = \Delta)$, therefore

$$\begin{aligned}
\pi_{n,m}(\Delta) &= 1 - F_Z(z_{1-\alpha/2} - \xi) + F_Z(-z_{1-\alpha/2} - \xi)\\
&= \Phi(\xi - z_{1-\alpha/2}) \;+\; \Phi(-\xi - z_{1-\alpha/2}) + \frac{\gamma_{\text{eff}}}{6}\,\big[(\xi - z_{1-\alpha/2})^2 - 1\big]\,\phi(\xi - z_{1-\alpha/2})\\
&\quad - \frac{\gamma_{\text{eff}}}{6}\,\big[(\xi + z_{1-\alpha/2})^2 - 1\big]\,\phi(\xi + z_{1-\alpha/2}) + O(n^{-1}+m^{-1})\\
&= \Phi(\xi - z_{1-\alpha/2}) \;+\; \Phi(-\xi - z_{1-\alpha/2}) + O\big((n^{-1}+m^{-1})^{\frac{1}{2}}\big).
\end{aligned}$$

Let $\kappa = \frac{m}{n}$, and define

$$h = h(\delta_A, \delta_X, \sigma_v, \sigma_\epsilon, n) = \sqrt{\frac{(c+\delta_X)^2\sigma_X^2 + [\delta_A^2\,p_o(1-p_o)+\sigma_U^2]+\eta}{np_r(1-p_r)} + \frac{\sigma_U^2+\eta}{\kappa np_o(1-p_o)}},$$

we have the power function such that

$$\pi_{n,m}(\Delta) = \Phi\left(-z_{1-\frac{\alpha}{2}} - \frac{\delta_A}{h}\right) + \Phi\left(\frac{\delta_A}{h} - z_{1-\frac{\alpha}{2}}\right) + \mathcal{O}(n^{-\frac{1}{2}}).$$

□

3 Simulation Parameter Settings

3.1 Type I Error

Under the linear data-generating model in Section 4.1, the null hypothesis $H_0 : \omega^o = \omega^r$ holds when $\beta_U\delta_A = 0$. We examine two scenarios that satisfy the null hypothesis: $(\beta_U, \delta_A) = (2, 0)$ and

$(\beta_U, \delta_A) = (0, 2)$. For each scenario, we generate $1,000$ independent replicates; in each replicate we generate an RCT sample of size $n = 100$ and an observational sample of size $m = 2,000$. We then apply the proposed bootstrap test with $B = 1,000$ resamples and compute the empirical Type I error as the proportion of replicates is rejected at a significant level α in which H_0 holds. Remaining parameters follow the settings in the Table 1 below.

Table 1: Linear Model Parameter Setting.

Parameter	$(\beta_U, \delta_A) = (2, 0)$	$(\beta_U, \delta_A) = (0, 2)$
$(\beta_0, \beta_A, \beta_X)$	$(1, 2, 0.5)$	$(1, 2, 0.5)$
(δ_0, δ_X)	$(1, 1)$	$(1, 1)$
σ_ϵ	0.1	0.1
σ_U	0.5	0.5
(p_r, p_o)	(0.2, 0.5)	(0.2, 0.5)

Under the non-linear data-generating model in Section 4.1, the null hypothesis $H_0 : \omega^o = \omega^r$ holds when $\beta_U \left[\delta_{XA}^2 + \delta_A(2\delta_0 + 2\delta_X + \delta_A)\right] = 0$. We examine two scenarios that satisfy the null hypothesis:$(\beta_U, \delta_A, \delta_{XA}) = (0, 2, 2)$ and $(\beta_U, \delta_A, \delta_{XA})$=$(2, 0, 0)$. For each scenario, we generate $1,000$ independent replicates; in each replicate we generate an RCT sample of size $n = 100$ and an observational sample of size $m = 2,000$. We then apply the proposed bootstrap test with $B = 1,000$ resamples and compute the empirical Type I error as the proportion of replicates is rejected at a significant level α in which H_0 holds. Remaining parameters follow the settings in the Table 2 below.

Table 2: Non-linear Model Parameter Setting.

Parameter	$(\beta_U, \delta_A, \delta_{XA}) = (0, 2, 2)$	$(\beta_U, \delta_A, \delta_{XA})$=$(2, 0, 0)$
$(\gamma_0, \gamma_A, \gamma_X, \gamma_{XA})$	$(1, 2, 0.5, 1)$	$(1, 2, 0.5, 1)$
(δ_0, δ_X)	$(1, 1)$	$(1, 1)$
σ_ϵ	0.1	0.1
σ_U	0.5	0.5
(p_r, p_o)	(0.2, 0.5)	(0.2, 0.5)

3.2 Power

To demonstrate how β_U and δ_A affect the power of the test and validate Theorem 3.2, we generate both RCT data and observational data with an unmeasured confounder using the linear model described Section 4.1 under the alternative hypothesis H_a. Since $\omega^o - \omega^r = \beta_U \delta_A$, we vary either β_U or δ_A from 1 to 10 while fixing the other at 1. At the significant level $\alpha = 0.05$, we compute empirical power using 200 replicates for each value of β_U or δ_A with $B = 500$ bootstrap resamples per replicate. Additionally, we compute the power under different values of σ_U. Remaining parameters follow the settings in the Table 3 below.

Table 3: (β_U, δ_A) are varying as being described in Section 4.1.

Parameter	
(β_0, β_A)	$(1, 2)$
β_X/β_U	0.5
(δ_0, δ_X)	$(1, 1)$
σ_ϵ	0.1
(p_r, p_o)	(0.2, 0.5)

4 Additional Information of R Files

Type1.R: R scripts for generating data and assessing Type I error.
Power.R: R scripts for generating data and assessing power function.
STAR.R: R scripts for the real-world data experiments.
STAR.csv: The raw Tennessee Student/Teacher Achievement Ratio dataset.

References

[1] Paul R Rosenbaum and Donald B Rubin. “The central role of the propensity score in observational studies for causal effects”. In: *Biometrika* 70.1 (1983), pp. 41–55.

[2] Jared K. Lunceford and Marie Davidian. “Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study”. In: *Statistics in Medicine* 23 (2004). URL: https://api.semanticscholar.org/CorpusID:11912618.

[3] Halbert White. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity”. In: *Econometrica* 48.4 (1980), pp. 817–838. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1912934 (visited on 08/19/2025).

[4] Peter Hall. *The bootstrap and Edgeworth expansion*. Springer Science & Business Media, 2013.