Supplementary material for "Score test for unconfoundedness under a logistic treatment assignment model"

Hairu Wang[1], Yukun Liu*[2] and Haiying Zhou[3]

[1] School of Mathematical Sciences, Jiangsu University, Zhenjiang, Jiangsu 212013, China
[2] KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai 200062, China
[3]Research Center for Applied Mathematics and Interdisciplinary Sciences, Beijing Normal University, Zhuhai 519087, China

Abstract

This supplementary materials consists of three sections. In Section 1, we review the conditions and Theorem 1 in the main paper. Section 2 contains a lemma, which can ease much burden of our proof of Theorem 1, which is given in Section 3.

1 Conditions and Theorem 1

We begin by reviewing the conditions we made and Theorem 1 in the main paper.

Condition (C1) $\mathbb{E}\|\mathbf{X}\|^2 < \infty$ and $\mathbf{A} = \mathbb{E}[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\mathbf{X}\mathbf{X}^\top]$ is of full rank.

Condition (C2) (i) The parameter space Ω_1 of $\boldsymbol{\zeta}$ is independent of $(y, \mathbf{x})$ and compact. (ii) The true value $\boldsymbol{\zeta}_0$ of $\boldsymbol{\zeta}$ is an interior point of Ω_1. (iii) $\boldsymbol{\zeta}$ is identifiable, i.e. $\mathbb{E}\{\int |f\{y(1)|\mathbf{X},\boldsymbol{\zeta}\} - f\{y(1)|\mathbf{X},\boldsymbol{\zeta}'\}|dy\} > 0$ for any different elements $\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}'$ in Ω_1. (iv) $\mathbb{E}\{\sup_{\boldsymbol{\zeta}\in\Omega_1} |\log f\{y(1)|\mathbf{X},\boldsymbol{\zeta}\}|\} < \infty$. (v) $f\{y(1)|\mathbf{X},\boldsymbol{\zeta}\}$ is continuous in $\boldsymbol{\zeta}$ for almost all $(y, \mathbf{x})$.

Condition (C3) (i) The parameter space Ω_2 of $\boldsymbol{\eta}$ is independent of $(y, \mathbf{x})$ and compact. (ii) The true value $\boldsymbol{\eta}_0$ of $\boldsymbol{\eta}$ is an interior point of Ω_2. (iii) $\boldsymbol{\eta}$ is identifiable, i.e. $\mathbb{E}\{\int |f\{y(0)|\mathbf{X},\boldsymbol{\eta}\} - f\{y(0)|\mathbf{X},\boldsymbol{\eta}'\}|dy\} > 0$ for any different elements $\boldsymbol{\eta}$ and $\boldsymbol{\eta}'$ in Ω_2. (iv) $\mathbb{E}\{\sup_{\boldsymbol{\eta}\in\Omega_2} |\log f\{y(0)|\mathbf{X},\boldsymbol{\eta}\}|\} < \infty$. (v) $f\{y(0)|\mathbf{X},\boldsymbol{\eta}\}$ is continuous in $\boldsymbol{\eta}$ for almost all $(y, \mathbf{x})$.

*Corresponding author: ykliu@sfs.ecnu.edu.cn

Condition (C4) (i) $f\{y(1)|\mathbf{X},\boldsymbol{\zeta}\}$ is twice differentiable with respect to $\boldsymbol{\zeta}$ for almost all $(y,\mathbf{x})$, and $\nabla_{\boldsymbol{\zeta}\boldsymbol{\zeta}^\top} f\{y(1)|\mathbf{X},\boldsymbol{\zeta}\}$ is continuous at $\boldsymbol{\zeta}_0$. (ii) $\mathbf{B} = \mathbb{E}[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{\nabla_{\boldsymbol{\zeta}} \log f\{y(1)|\mathbf{X},\boldsymbol{\zeta}\}\}^{\otimes 2}]$ is positive definite. (iii) There exist a $\delta > 0$ and positive functions $M_1(\mathbf{x})$ and $M_2(y,\mathbf{x})$ such that $\mathbb{E}\{M_1(\mathbf{X})\} < \infty$ and $\mathbb{E}\{M_2(Y,\mathbf{X})\} < \infty$, and $\|\mathbf{x}\| \int |t|\{f(t|\mathbf{x},\boldsymbol{\zeta}) + \|\nabla_{\boldsymbol{\zeta}} f(t|\mathbf{x},\boldsymbol{\zeta})\|\}dt \le M_1(\mathbf{x})$ and $\|\nabla_{\boldsymbol{\zeta}\boldsymbol{\zeta}^\top} \log f(y|\mathbf{x},\boldsymbol{\zeta})\| \le M_2(y,\mathbf{x})$ for all $\boldsymbol{\zeta}$ satisfying $\|\boldsymbol{\zeta} - \boldsymbol{\zeta}_0\| \le \delta$.

Condition (C5) (i) $f\{y(0)|\mathbf{X},\boldsymbol{\eta}\}$ is twice differentiable with respect to $\boldsymbol{\eta}$ for almost all $(y,\mathbf{x})$, and $\nabla_{\boldsymbol{\eta}\boldsymbol{\eta}^\top} f\{y(0)|\mathbf{X},\boldsymbol{\eta}\}$ is continuous at $\boldsymbol{\eta}_0$. (ii) $\mathbf{C} = \mathbb{E}[\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\{\nabla_{\boldsymbol{\eta}} \log f\{y(0)|\mathbf{X},\boldsymbol{\eta}\}\}^{\otimes 2}]$ is positive definite. (iii) There exist a $\delta > 0$ and positive functions $M_3(\mathbf{x})$ and $M_4(y,\mathbf{x})$ such that $\mathbb{E}\{M_3(\mathbf{X})\} < \infty$ and $\mathbb{E}\{M_4(Y,\mathbf{X})\} < \infty$, and $\|\mathbf{x}\| \int |t|\{f(t|\mathbf{x},\boldsymbol{\eta}) + \|\nabla_{\boldsymbol{\eta}} f(t|\mathbf{x},\boldsymbol{\eta})\|\}dt \le M_3(\mathbf{x})$ and $\|\nabla_{\boldsymbol{\eta}\boldsymbol{\eta}^\top} \log f(y|\mathbf{x},\boldsymbol{\eta})\| \le M_4(y,\mathbf{x})$ for all $\boldsymbol{\eta}$ satisfying $\|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| \le \delta$.

Define

$$\begin{aligned} \ell_{0*}(\boldsymbol{\beta}) &= \mathbb{E}[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\log\pi(\mathbf{X}^\top\boldsymbol{\beta}) + \{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\log\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta})\}] \\ &= \mathbb{E}\{(\mathbf{X}^\top\boldsymbol{\beta})\pi(\mathbf{X}^\top\boldsymbol{\beta}_0) - \log\{1+\exp(\mathbf{X}^\top\boldsymbol{\beta})\}\}. \end{aligned} \tag{1}$$

This function is well defined for any $\boldsymbol{\beta}$ under Condition (C1), because

$$|\ell_{0*}(\boldsymbol{\beta})| \le \mathbb{E}\{|\mathbf{X}^\top\boldsymbol{\beta}| + \log\{1+\exp(|X^\top\boldsymbol{\beta}|)\}\} \le 3\mathbb{E}\{|\mathbf{X}^\top\boldsymbol{\beta}|\} \le 3\mathbb{E}(\|\mathbf{X}\|)\|\boldsymbol{\beta}\|,$$

and Condition (C1) implies $\mathbb{E}\|\mathbf{X}\| < \infty$.

Theorem 1 *Assume Conditions (C1)–(C5) and that $H_0 : \gamma_0 = \gamma_1 = \theta_0 = \theta_1 = 0$ is true. As n goes to infinity, $\sqrt{n}S_1(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\zeta}},\widehat{\boldsymbol{\eta}}) \xrightarrow{d} \mathcal{N}(0,\Sigma_1)$, where*

$$\Sigma_1 = \begin{pmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 & \sigma_{14}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \sigma_{23}^2 & \sigma_{24}^2 \\ \sigma_{31}^2 & \sigma_{32}^2 & \sigma_{33}^2 & \sigma_{34}^2 \\ \sigma_{41}^2 & \sigma_{42}^2 & \sigma_{43}^2 & \sigma_{44}^2 \end{pmatrix},$$

and

$$
\begin{aligned}
\sigma_{11}^2 &= A_{21} + B_{21} - \mathbf{A}_{11}^\top \mathbf{A}^{-1} \mathbf{A}_{11} - \mathbf{B}_{11}^\top \mathbf{C}^{-1} \mathbf{B}_{11}, \\
\sigma_{22}^2 &= A_{22} + B_{22} - \mathbf{A}_{12}^\top \mathbf{A}^{-1} \mathbf{A}_{12} - \mathbf{B}_{12}^\top \mathbf{B}^{-1} \mathbf{B}_{12}, \\
\sigma_{33}^2 &= A_{23} + B_{23} - \mathbf{A}_{13}^\top \mathbf{A}^{-1} \mathbf{A}_{13} - \mathbf{B}_{13}^\top \mathbf{C}^{-1} \mathbf{B}_{13}, \\
\sigma_{44}^2 &= A_{24} + B_{24} - \mathbf{A}_{14}^\top \mathbf{A}^{-1} \mathbf{A}_{14} - \mathbf{B}_{14}^\top \mathbf{B}^{-1} \mathbf{B}_{14}, \\
\sigma_{12}^2 &= \sigma_{21}^2 = A_2 - \mathbf{A}_{11}^\top \mathbf{A}^{-1} \mathbf{A}_{12}, \\
\sigma_{13}^2 &= \sigma_{31}^2 = A_{31} + \mathbf{B}_{31} - \mathbf{A}_{11}^\top \mathbf{A}^{-1} \mathbf{A}_{13} - \mathbf{B}_{11}^\top \mathbf{C}^{-1} \mathbf{B}_{13}, \\
\sigma_{14}^2 &= \sigma_{41}^2 = A_3 - \mathbf{A}_{11}^\top \mathbf{A}^{-1} \mathbf{A}_{14}, \\
\sigma_{23}^2 &= \sigma_{32}^2 = A_4 - \mathbf{A}_{12}^\top \mathbf{A}^{-1} \mathbf{A}_{13}, \\
\sigma_{24}^2 &= \sigma_{42}^2 = A_{32} + \mathbf{B}_{32} - \mathbf{A}_{12}^\top \mathbf{A}^{-1} \mathbf{A}_{14} - \mathbf{B}_{12}^\top \mathbf{B}^{-1} \mathbf{B}_{14}, \\
\sigma_{34}^2 &= \sigma_{43}^2 = A_5 - \mathbf{A}_{13}^\top \mathbf{A}^{-1} \mathbf{A}_{14},
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{A}_{11} &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}Y(0)\mathbf{X}\right],\\
\mathbf{A}_{12} &= \mathbb{E}\{\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}Y(1)\mathbf{X}\},\\
\mathbf{A}_{13} &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}Y^2(0)\mathbf{X}\right],\\
\mathbf{A}_{14} &= \mathbb{E}\{\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}Y^2(1)\mathbf{X}\},\\
A_{21} &= \mathbb{E}\left[\pi^2(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}Y^2(0)\right],\\
A_{22} &= \mathbb{E}\{\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2Y^2(1)\},\\
A_{23} &= \mathbb{E}\left[\pi^2(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2Y^4(0)\right],\\
A_{24} &= \mathbb{E}\{\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2Y^4(1)\},\\
A_{31} &= \mathbb{E}\left[\pi^2(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}Y^3(0)\right],\\
A_{32} &= \mathbb{E}\{\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}Y^3(1)\},\\
A_2 &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}Y(0)Y(1)\right],\\
A_3 &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}Y(0)Y^2(1)\right],\\
A_4 &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}Y^2(0)Y(1)\right],\\
A_5 &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2Y^2(0)Y^2(1)\right],\\
\mathbf{B}_{11} &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\int y\nabla_{\boldsymbol{\eta}}f\{y|\mathbf{X},\boldsymbol{\eta}_0\}dy\right],\\
\mathbf{B}_{12} &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\int y\nabla_{\boldsymbol{\zeta}}f\{y|\mathbf{X},\boldsymbol{\zeta}_0\}dy\right],\\
\mathbf{B}_{13} &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\int y^2\nabla_{\boldsymbol{\eta}}f\{y|\mathbf{X},\boldsymbol{\eta}_0\}dy\right],\\
\mathbf{B}_{14} &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\int y^2\nabla_{\boldsymbol{\zeta}}f\{y|\mathbf{X},\boldsymbol{\zeta}_0\}dy\right],\\
B_{21} &= \mathbb{E}[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2\mathbb{E}^2(Y(0)|X,\eta_0)],\\
B_{22} &= \mathbb{E}[\{\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\mathbb{E}^2(Y(1)|X,\zeta_0)],\\
B_{23} &= \mathbb{E}[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2\mathbb{E}^2\{Y^2(0)|X,\boldsymbol{\eta}_0\}],\\
B_{24} &= \mathbb{E}[\{\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2\mathbb{E}^2\{Y^2(1)|X,\boldsymbol{\zeta}_0\}],\\
B_{31} &= \mathbb{E}[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\mathbb{E}\{Y(0)|X,\boldsymbol{\eta}_0\}\mathbb{E}\{Y^2(0)|X,\boldsymbol{\zeta}_0\}],\\
B_{32} &= \mathbb{E}[\{\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\mathbb{E}\{Y(1)|X,\boldsymbol{\zeta}_0\}\mathbb{E}\{Y^2(1)|X,\boldsymbol{\zeta}_0\}].
\end{aligned}
$$

2 A lemma

The following lemma eases much of the burden in our proof of Theorem 1.

Lemma 1 *Assume Conditions (C1)–(C5) and that $H_0 : \gamma_0 = \gamma_1 = \theta_0 = \theta_1 = 0$ is true. As n goes to infinity,*

$$\sqrt{n}(\widehat{\boldsymbol\beta} - \boldsymbol\beta_0) = \mathbf{A}^{-1} n^{-1/2} \sum_{i=1}^{n} \{w_i - \pi(\mathbf{x}_i^\top \boldsymbol\beta_0)\}\mathbf{x}_i + o_p(1), \tag{2}$$

$$\sqrt{n}(\widehat{\boldsymbol\zeta} - \boldsymbol\zeta_0) = \mathbf{B}^{-1} n^{-1/2} \sum_{i=1}^{n} w_i \nabla_{\boldsymbol\zeta} \log f\{y_i(1)|\mathbf{x}_i, \boldsymbol\zeta_0\} + o_p(1), \tag{3}$$

$$\sqrt{n}(\widehat{\boldsymbol\eta} - \boldsymbol\eta_0) = \mathbf{C}^{-1} n^{-1/2} \sum_{i=1}^{n} (1 - w_i) \nabla_{\boldsymbol\eta} \log f\{y_i(0)|\mathbf{x}_i, \boldsymbol\eta_0\} + o_p(1), \tag{4}$$

and

$$\sqrt{n}\{(\widehat{\boldsymbol\beta} - \boldsymbol\beta_0)^\top, (\widehat{\boldsymbol\zeta} - \boldsymbol\zeta_0)^\top, (\widehat{\boldsymbol\eta} - \boldsymbol\eta_0)^\top\}^\top \xrightarrow{d} \mathcal{N}(\mathbf{0},\ \mathrm{diag}(\mathbf{A^{-1}},\ \mathbf{B^{-1}},\ \mathbf{C^{-1}})),$$

where the matrices $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *are defined in Conditions (C1), (C6) and (C7).*

Proof By definition, $\widehat{\boldsymbol\beta} = \arg\max_{\boldsymbol\beta} \frac{1}{n}\ell_0(\boldsymbol\beta)$, $\widehat{\boldsymbol\zeta} = \arg\max_{\boldsymbol\zeta} \frac{1}{n}\ell_1(\boldsymbol\zeta)$, and $\widehat{\boldsymbol\eta} = \arg\max_{\boldsymbol\eta} \frac{1}{n}\ell_2(\boldsymbol\eta)$, where $\ell_0(\boldsymbol\beta) = \sum_{i=1}^n [w_i \log \pi(\mathbf{x}_i^\top \boldsymbol\beta) + (1 - w_i) \log\{1 - \pi(\mathbf{x}_i^\top \boldsymbol\beta)\}]$, $\ell_1(\boldsymbol\zeta) = \sum_{i=1}^n \{w_i \log f\{y_i(1)|\mathbf{x}_i, \boldsymbol\zeta\}\}$ and $\ell_2(\boldsymbol\eta) = \sum_{i=1}^n \{(1 - w_i) \log f\{y_i(0)|\mathbf{x}_i, \boldsymbol\eta\}\}$. Because $\pi(\mathbf{x}_i^\top \boldsymbol\beta)$, $f\{y_i(1)|\mathbf{x}_i, \boldsymbol\zeta\}$ and $f\{y_i(0)|\mathbf{x}_i, \boldsymbol\eta\}$ are differentiable with respect to the underlying parameters, equivalently the estimators $\widehat{\boldsymbol\beta}$, $\widehat{\boldsymbol\zeta}$ and $\widehat{\boldsymbol\eta}$ satisfy

$$\sum_{i=1}^{n} \{w_i - \pi(\mathbf{x}_i^\top \widehat{\boldsymbol\beta})\}\mathbf{x}_i = 0, \tag{5}$$

$$\sum_{i=1}^{n} w_i \nabla_{\boldsymbol\zeta} \log f\{y_i(1)|\mathbf{x}_i, \boldsymbol\zeta\} = 0, \tag{6}$$

$$\sum_{i=1}^{n} (1 - w_i) \nabla_{\boldsymbol\eta} \log f\{y_i(0)|\mathbf{x}_i, \boldsymbol\eta\} = 0. \tag{7}$$

Both $\ell_0(\boldsymbol\beta)$ and $\ell_{0*}(\boldsymbol\beta)$ are concave functions, and $(1/n)\ell_0(\boldsymbol\beta)$ converges to $\ell_{0*}(\boldsymbol\beta)$ almost surely for each fixed $\boldsymbol\beta$. By the convexity lemma of **?**, $\sup_{\boldsymbol\beta} |(1/n)\ell_0(\boldsymbol\beta) - \ell_{0*}(\boldsymbol\beta)| = o_p(1)$. Condition (C1) implies that $\ell_{0*}(\boldsymbol\beta)$ is strictly concave at $\boldsymbol\beta_0$. This, together with $\nabla_{\boldsymbol\beta}\ell_{0*}(\boldsymbol\beta)|_{\boldsymbol\beta=\boldsymbol\beta_0} = 0$, implies that $\boldsymbol\beta_0$ is the unique maximizer of $\ell_{0*}(\boldsymbol\beta)$. Then by Theorem 5.7 of **?**, $\widehat{\boldsymbol\beta} - \boldsymbol\beta_0 = o_p(1)$. The linear approximation (2) follows by first-order Taylor expansion of the left-hand side of (5) at $\boldsymbol\beta_0$ and by Condition (C1).

Next, we derive a linear approximation of $\widehat{\boldsymbol\zeta}$ and $\widehat{\boldsymbol\eta}$. Let $\ell_{1*}(\boldsymbol\zeta) = \mathbb{E}[\pi(\mathbf{X}^\top \boldsymbol\beta_0) \log f\{Y(1)|\mathbf{X}, \boldsymbol\zeta_0\}]$ and $\ell_{2*}(\boldsymbol\eta) = \mathbb{E}[\{1 - \pi(\mathbf{X}^\top \boldsymbol\beta_0)\} \log f\{Y(0)|\mathbf{X}, \boldsymbol\eta_0\}]$. The identifiability in Condition (C4) guarantees that $\boldsymbol\zeta_0$ is the unique maximizer of $\ell_{1*}(\boldsymbol\zeta)$. Under Condition (C4), $f\{y(1)|\mathbf{x}, \boldsymbol\zeta\}$ is continuous in $\boldsymbol\zeta$ for almost all $(y, \mathbf{x})$. By Lemma 3.10 of **?**, the class $\{\log f\{y(1)|\mathbf{x}, \boldsymbol\zeta\} : \boldsymbol\zeta \in \Omega_1\}$ satisfies the uniform law of large numbers, in other words, $\sup_{\boldsymbol{\zeta}\in\Omega_1} |\ell_1(\boldsymbol{\zeta})/n - \ell_{1*}(\boldsymbol{\zeta})| = o_p(1)$. Therefore, the conditions of Theorem 5.7 of **?** are met, and $\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0 = o_p(1)$. The linear approximation (3) can be derived using similar techniques to those deriving the approximation (2). The linear approximation of $\widehat{\boldsymbol{\eta}}$ can be derived using similar processes to those deriving the approximation of $\widehat{\boldsymbol{\zeta}}$.

The asymptotic result follows immediately from the linear approximations in (2), (3) and (4) by the central limit theorem. Condition (C6) and (C7) guarantees that the remainder terms in the linear approximations are negligible and that their asymptotic variances are well defined. This completes the proof of Lemma 1.

3 Proof of Theorem 1

We have shown that $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\zeta}}, \widehat{\boldsymbol{\eta}})$ is $\sqrt{n}$ consistent with $(\boldsymbol{\beta}_0, \boldsymbol{\zeta}_0, \boldsymbol{\eta}_0)$. By first-order Taylor expansion of $S_1(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\zeta}}, \widehat{\boldsymbol{\eta}})$ at $(\boldsymbol{\beta}_0, \boldsymbol{\zeta}_0, \boldsymbol{\eta}_0)$, we have

$$\begin{aligned}\sqrt{n}S_1(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\zeta}}, \widehat{\boldsymbol{\eta}}) &= \sqrt{n}S_1(\boldsymbol{\beta}_0, \boldsymbol{\zeta}_0, \boldsymbol{\eta}_0) + \{\nabla_{\boldsymbol{\beta}} S_1(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\zeta}}, \widetilde{\boldsymbol{\eta}})\}^\top \cdot \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &\quad + \{\nabla_{\boldsymbol{\zeta}} S_1(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\zeta}}, \widetilde{\boldsymbol{\eta}})\}^\top \cdot \sqrt{n}(\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0) \\ &\quad + \{\nabla_{\boldsymbol{\eta}} S_1(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\zeta}}, \widetilde{\boldsymbol{\eta}})\}^\top \cdot \sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0), \end{aligned} \tag{8}$$

where $(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\zeta}}, \widetilde{\boldsymbol{\eta}})$ lies between $(\boldsymbol{\beta}_0, \boldsymbol{\zeta}_0, \boldsymbol{\eta}_0)$ and $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\zeta}}, \widehat{\boldsymbol{\eta}})$.

To proceed, we first need to derive the leading terms of $\nabla_{\boldsymbol{\beta}} S_1(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\zeta}}, \widetilde{\boldsymbol{\eta}})$, $\nabla_{\boldsymbol{\zeta}} S_1(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\zeta}}, \widetilde{\boldsymbol{\eta}})$ and $\nabla_{\boldsymbol{\eta}} S_1(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\zeta}}, \widetilde{\boldsymbol{\eta}})$. It follows from

$$S_1(\boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{\eta}) = \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} w_i\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\mathbb{E}\{Y(0)|\mathbf{x}_i, \boldsymbol{\eta}\} - (1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta})y_i(0) \\ w_i\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}y_i(1) - (1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\mathbb{E}\{Y(1)|\mathbf{x}_i, \boldsymbol{\zeta}\} \\ w_i\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\{1-2\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\mathbb{E}\{Y^2(0)|\mathbf{x}_i, \boldsymbol{\eta}\} - (1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\{1-2\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}y_i^2(0) \\ w_i\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\{1-2\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}y_i^2(1) - (1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\{1-2\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\mathbb{E}\{Y^2(1)|\mathbf{x}_i, \boldsymbol{\zeta}\} \end{pmatrix},$$

that

$$\begin{aligned}\nabla_{\boldsymbol{\beta}^\top} S_1(\boldsymbol{\beta}, \boldsymbol{\zeta}, \boldsymbol{\eta}) &= \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} -w_i\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\mathbb{E}\{Y(0)|\mathbf{x}_i, \boldsymbol{\eta}\}\mathbf{x}_i^\top \\ -w_i\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}y_i(1)\mathbf{x}_i^\top \\ -w_i\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\{3-4\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\mathbb{E}\{Y^2(0)|\mathbf{x}_i, \boldsymbol{\eta}\}\mathbf{x}_i^\top \\ -w_i\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\{3-4\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}y_i^2(1)\mathbf{x}_i^\top \end{pmatrix} \\ &\quad -\frac{1}{n}\sum_{i=1}^n \begin{pmatrix} (1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}y_i(0)\mathbf{x}_i^\top \\ (1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\mathbb{E}\{Y(1)|\mathbf{x}_i, \boldsymbol{\zeta}\}\mathbf{x}_i^\top \\ (1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\{1-4\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}y_i^2(0)\mathbf{x}_i^\top \\ (1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\{1-4\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\mathbb{E}\{Y^2(1)|\mathbf{x}_i, \boldsymbol{\zeta}\}\mathbf{x}_i^\top \end{pmatrix},\end{aligned}$$

$$\nabla_{\boldsymbol{\zeta}^\top} S_1(\boldsymbol{\beta},\boldsymbol{\zeta},\boldsymbol{\eta}) = \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} \mathbf{0}^\top \\ -(1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\nabla_{\boldsymbol{\zeta}^\top}\mathbb{E}\{Y(1)|\mathbf{x}_i,\boldsymbol{\zeta}\} \\ \mathbf{0}^\top \\ -(1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\{1-2\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\nabla_{\boldsymbol{\zeta}^\top}\mathbb{E}\{Y^2(1)|\mathbf{x}_i,\boldsymbol{\zeta}\} \end{pmatrix},$$

$$\nabla_{\boldsymbol{\eta}^\top} S_1(\boldsymbol{\beta},\boldsymbol{\zeta},\boldsymbol{\eta}) = \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} w_i\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\nabla_{\boldsymbol{\eta}^\top}\mathbb{E}\{Y(0)|\mathbf{x}_i,\boldsymbol{\eta}\} \\ \mathbf{0}^\top \\ w_i\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\{1-2\pi(\mathbf{x}_i^\top\boldsymbol{\beta})\}\nabla_{\boldsymbol{\eta}^\top}\mathbb{E}\{Y^2(0)|\mathbf{x}_i,\boldsymbol{\eta}\} \\ \mathbf{0}^\top \end{pmatrix}.$$

Under Conditions (C1), (C6) and (C7), by the weak law of large numbers, they converge in probability respectively to

$$\begin{aligned}\mathbb{E}[\nabla_{\boldsymbol{\beta}^\top} S_1(\boldsymbol{\beta},\boldsymbol{\zeta},\boldsymbol{\eta})] &= \mathbb{E}\begin{pmatrix} -\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\pi(\mathbf{X}^\top\boldsymbol{\beta})\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta})\}\mathbb{E}\{Y(0)|\mathbf{X},\boldsymbol{\eta}\}\mathbf{X}^\top \\ -\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\pi(\mathbf{X}^\top\boldsymbol{\beta})\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta})\}Y(1)\mathbf{X}^\top \\ -\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\pi(\mathbf{X}^\top\boldsymbol{\beta})\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta})\}\{3-4\pi(\mathbf{X}^\top\boldsymbol{\beta})\}\mathbb{E}\{Y^2(0)|\mathbf{X},\boldsymbol{\eta}\}\mathbf{X}^\top \\ -\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\pi(\mathbf{X}^\top\boldsymbol{\beta})\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta})\}\{3-4\pi(\mathbf{X}^\top\boldsymbol{\beta})\}Y^2(1)\mathbf{X}^\top \end{pmatrix} \\ &\quad -\mathbb{E}\begin{pmatrix} \{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\pi(\mathbf{X}^\top\boldsymbol{\beta})\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta})\}Y(0)\mathbf{X}^\top \\ \{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\pi(\mathbf{X}^\top\boldsymbol{\beta})\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta})\}\mathbb{E}\{Y(1)|\mathbf{X}^\top,\boldsymbol{\zeta}\}\mathbf{X}^\top \\ \{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\pi(\mathbf{X}^\top\boldsymbol{\beta})\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta})\}\{1-4\pi(\mathbf{X}^\top\boldsymbol{\beta})\}Y^2(0)\mathbf{X}^\top \\ \{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\pi(\mathbf{X}^\top\boldsymbol{\beta})\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta})\}\{1-4\pi(\mathbf{X}^\top\boldsymbol{\beta})\}\mathbb{E}\{Y^2(1)|\mathbf{X},\boldsymbol{\zeta}\}\mathbf{X}^\top \end{pmatrix},\end{aligned}$$

$$\mathbb{E}[\nabla_{\boldsymbol{\zeta}^\top} S_1(\boldsymbol{\beta},\boldsymbol{\zeta},\boldsymbol{\eta})] = \mathbb{E}\begin{pmatrix} \mathbf{0}^\top \\ -\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\pi(\mathbf{X}^\top\boldsymbol{\beta})\nabla_{\boldsymbol{\zeta}^\top}\mathbb{E}\{Y(1)|\mathbf{X},\boldsymbol{\zeta}\} \\ \mathbf{0}^\top \\ -\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\pi(\mathbf{X}^\top\boldsymbol{\beta})\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta})\}\nabla_{\boldsymbol{\zeta}^\top}\mathbb{E}\{Y^2(1)|\mathbf{X},\boldsymbol{\zeta}\} \end{pmatrix},$$

$$\mathbb{E}[\nabla_{\boldsymbol{\eta}^\top} S_1(\boldsymbol{\beta},\boldsymbol{\zeta},\boldsymbol{\eta})] = \mathbb{E}\begin{pmatrix} \pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta})\}\nabla_{\boldsymbol{\eta}^\top}\mathbb{E}\{Y(0)|\mathbf{X},\boldsymbol{\eta}\} \\ \mathbf{0}^\top \\ \pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta})\}\{1-2\pi(\mathbf{X}^\top\boldsymbol{\beta})\}\nabla_{\boldsymbol{\eta}^\top}\mathbb{E}\{Y^2(0)|\mathbf{X},\boldsymbol{\eta}\} \\ \mathbf{0}^\top \end{pmatrix}.$$

Conditions (C1), (C6) and (C7) also imply that the above convergences hold uniformly in $(\boldsymbol{\beta},\boldsymbol{\zeta},\boldsymbol{\eta})$, and therefore the two limit functions are continuous functions of $(\boldsymbol{\beta},\boldsymbol{\zeta},\boldsymbol{\eta})$. This, together with the consistency of $(\widehat{\boldsymbol\beta},\widehat{\boldsymbol\zeta},\widehat{\boldsymbol\eta})$, leads to

$$
\begin{aligned}
\nabla_{\boldsymbol\beta^\top} S_1(\widetilde{\boldsymbol\beta},\widetilde{\boldsymbol\zeta},\widetilde{\boldsymbol\eta}) &= -\begin{pmatrix}\mathbf{A}_{11}^\top\\ \mathbf{A}_{12}^\top\\ \mathbf{A}_{13}^\top\\ \mathbf{A}_{14}^\top\end{pmatrix} + o_p(1),\\
\nabla_{\boldsymbol\zeta^\top} S_1(\widetilde{\boldsymbol\beta},\widetilde{\boldsymbol\zeta},\widetilde{\boldsymbol\eta}) &= -\begin{pmatrix}\mathbf{0}^\top\\ \mathbf{B}_{12}^\top\\ \mathbf{0}^\top\\ \mathbf{B}_{14}^\top\end{pmatrix} + o_p(1),\\
\nabla_{\boldsymbol\eta^\top} S_1(\widetilde{\boldsymbol\beta},\widetilde{\boldsymbol\zeta},\widetilde{\boldsymbol\eta}) &= \begin{pmatrix}\mathbf{B}_{11}^\top\\ \mathbf{0}^\top\\ \mathbf{B}_{13}^\top\\ \mathbf{0}^\top\end{pmatrix} + o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{A}_{11} &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol\beta_0)\{1-\pi(\mathbf{X}^\top\boldsymbol\beta_0)\}Y(0)\mathbf{X}\right],\\
\mathbf{A}_{12} &= \mathbb{E}\{\pi(\mathbf{X}^\top\boldsymbol\beta_0)\{1-\pi(\mathbf{X}^\top\boldsymbol\beta_0)\}Y(1)\mathbf{X}\},\\
\mathbf{A}_{13} &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol\beta_0)\{1-\pi(\mathbf{X}^\top\boldsymbol\beta_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol\beta_0)\}Y^2(0)\mathbf{X}\right],\\
\mathbf{A}_{14} &= \mathbb{E}\{\pi(\mathbf{X}^\top\boldsymbol\beta_0)\{1-\pi(\mathbf{X}^\top\boldsymbol\beta_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol\beta_0)\}Y^2(1)\mathbf{X}\},\\
\mathbf{B}_{11} &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol\beta_0)\{1-\pi(\mathbf{X}^\top\boldsymbol\beta_0)\}\int y\nabla_{\boldsymbol\eta} f\{y|\mathbf{X},\boldsymbol\eta_0\}dy\right],\\
\mathbf{B}_{12} &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol\beta_0)\{1-\pi(\mathbf{X}^\top\boldsymbol\beta_0)\}\int y\nabla_{\boldsymbol\zeta} f\{y|\mathbf{X},\boldsymbol\zeta_0\}dy\right],\\
\mathbf{B}_{13} &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol\beta_0)\{1-\pi(\mathbf{X}^\top\boldsymbol\beta_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol\beta_0)\}\int y^2\nabla_{\boldsymbol\eta} f\{y|\mathbf{X},\boldsymbol\eta_0\}dy\right],\\
\mathbf{B}_{14} &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol\beta_0)\{1-\pi(\mathbf{X}^\top\boldsymbol\beta_0)\}\{1-2\pi(\mathbf{X}^\top\boldsymbol\beta_0)\}\int y^2\nabla_{\boldsymbol\zeta} f\{y|\mathbf{X},\boldsymbol\zeta_0\}dy\right].
\end{aligned}
$$

Consequently, the equality (8) reduces to

$$
\begin{aligned}
\sqrt{n}S_1(\widehat{\boldsymbol\beta},\widehat{\boldsymbol\zeta},\widehat{\boldsymbol\eta}) &= \sqrt{n}S_1(\boldsymbol\beta_0,\boldsymbol\zeta_0,\boldsymbol\eta_0) - \begin{pmatrix}\mathbf{A}_{11}^\top\\ \mathbf{A}_{12}^\top\\ \mathbf{A}_{13}^\top\\ \mathbf{A}_{14}^\top\end{pmatrix}\sqrt{n}(\widehat{\boldsymbol\beta}-\boldsymbol\beta_0) - \begin{pmatrix}\mathbf{0}^\top\\ \mathbf{B}_{12}^\top\\ \mathbf{0}^\top\\ \mathbf{B}_{14}^\top\end{pmatrix}\sqrt{n}(\widehat{\boldsymbol\zeta}-\boldsymbol\zeta_0)\\
&\quad + \begin{pmatrix}\mathbf{B}_{11}^\top\\ \mathbf{0}^\top\\ \mathbf{B}_{13}^\top\\ \mathbf{0}^\top\end{pmatrix}\sqrt{n}(\widehat{\boldsymbol\eta}-\boldsymbol\eta_0) + o_p(1). \qquad (9)
\end{aligned}
$$

Putting the linear approximations of $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\zeta}}$ and $\widehat{\boldsymbol{\eta}}$ from (2), (3) and (4) into (9), we have

$$\sqrt{n}S_1(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\zeta}},\widehat{\boldsymbol{\eta}}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} v_{i1}+v_{i2}+v_{i3}+v_{i4} \\ v_{i5}+v_{i6}+v_{i7}+v_{i8} \\ v_{i9}+v_{i10}+v_{i11}+v_{i12} \\ v_{i13}+v_{i14}+v_{i15}+v_{i16} \end{pmatrix} + o_p(1),$$

where

$$\begin{aligned}
&v_{i1} = w_i\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\}\mathbb{E}\{Y(0)|\mathbf{x}_i,\boldsymbol{\eta}_0\}, && v_{i2} = -(1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)y_i(0),\\
&v_{i3} = -\mathbf{A}_{11}^\top\mathbf{A}^{-1}\{w_i-\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\}\mathbf{x}_i, && v_{i4} = \mathbf{B}_{11}^\top\mathbf{C}^{-1}(1-w_i)\nabla_{\boldsymbol{\eta}}\log f\{y_i(0)|\mathbf{x}_i,\boldsymbol{\eta}_0\},\\
&v_{i5} = w_i\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\}y_i(1), && v_{i6} = -(1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\mathbb{E}\{Y(1)|\mathbf{x}_i,\boldsymbol{\zeta}_0\},\\
&v_{i7} = -\mathbf{A}_{12}^\top\mathbf{A}^{-1}\{w_i-\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\}\mathbf{x}_i, && v_{i8} = -\mathbf{B}_{12}^\top\mathbf{B}^{-1}w_i\nabla_{\boldsymbol{\zeta}}\log f(y_i(1)|\mathbf{x}_i,\boldsymbol{\zeta}_0),\\
&v_{i9} = w_i\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\}\{1-2\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\}\mathbb{E}\{Y^2(0)|\mathbf{x}_i,\boldsymbol{\eta}_0\}, && v_{i10} = -(1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\{1-2\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\}y_i^2(0),\\
&v_{i11} = -\mathbf{A}_{13}^\top\mathbf{A}^{-1}\{w_i-\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\}\mathbf{x}_i, && v_{i12} = \mathbf{B}_{13}^\top\mathbf{C}^{-1}(1-w_i)\nabla_{\boldsymbol{\eta}}\log f\{y_i(0)|\mathbf{x}_i,\boldsymbol{\eta}_0\},\\
&v_{i13} = w_i\{1-\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\}\{1-2\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\}y_i^2(1), && v_{i14} = -(1-w_i)\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\{1-2\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\}\mathbb{E}\{Y^2(1)|\mathbf{x}_i,\boldsymbol{\zeta}_0\},\\
&v_{i15} = -\mathbf{A}_{14}^\top\mathbf{A}^{-1}\{w_i-\pi(\mathbf{x}_i^\top\boldsymbol{\beta}_0)\}\mathbf{x}_i, && v_{i16} = -\mathbf{B}_{14}^\top\mathbf{B}^{-1}w_i\nabla_{\boldsymbol{\zeta}}\log f(y_i(1)|\mathbf{x}_i,\boldsymbol{\zeta}_0).
\end{aligned}$$

By the central limit theorem, as $n\to\infty$, $\sqrt{n}S_1(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\zeta}},\widehat{\boldsymbol{\eta}})\xrightarrow{d} N(0,\Sigma_1)$, with

$$\Sigma_1 = \mathbb{V}\mathrm{ar}\begin{pmatrix} v_{i1}+v_{i2}+v_{i3}+v_{i4} \\ v_{i5}+v_{i6}+v_{i7}+v_{i8} \\ v_{i9}+v_{i10}+v_{i11}+v_{i12} \\ v_{i13}+v_{i14}+v_{i15}+v_{i16} \end{pmatrix}.$$

Because $\mathbb{E}(v_{i1}+v_{i2}) = \mathbb{E}(v_{i5}+v_{i6}) = \mathbb{E}(v_{i9}+v_{i10}) = \mathbb{E}(v_{i13}+v_{i14}) = \mathbb{E}(v_{i3}) = \mathbb{E}(v_{i4}) = \mathbb{E}(v_{i7}) = \mathbb{E}(v_{i8}) = \mathbb{E}(v_{i11}) = \mathbb{E}(v_{i12}) = \mathbb{E}(v_{i15}) = \mathbb{E}(v_{i16}) = 0$, we have

$$\begin{gathered}
\sigma_{11}^2 = \mathbb{E}(v_{i1}+v_{i2}+v_{i3}+v_{i4})^2, \quad \sigma_{22}^2 = \mathbb{E}(v_{i5}+v_{i6}+v_{i7}+v_{i8})^2,\\
\sigma_{33}^2 = \mathbb{E}(v_{i9}+v_{i10}+v_{i11}+v_{i12})^2, \quad \sigma_{44}^2 = \mathbb{E}(v_{i13}+v_{i14}+v_{i15}+v_{i16})^2,\\
\sigma_{12}^2 = \sigma_{21}^2 = \mathbb{E}\{(v_{i1}+v_{i2}+v_{i3}+v_{i4})(v_{i5}+v_{i6}+v_{i7}+v_{i8})\},\\
\sigma_{13}^2 = \sigma_{31}^2 = \mathbb{E}\{(v_{i1}+v_{i2}+v_{i3}+v_{i4})(v_{i9}+v_{i10}+v_{i11}+v_{i12})\},\\
\sigma_{14}^2 = \sigma_{41}^2 = \mathbb{E}\{(v_{i1}+v_{i2}+v_{i3}+v_{i4})(v_{i13}+v_{i14}+v_{i15}+v_{i16})\},\\
\sigma_{23}^2 = \sigma_{32}^2 = \mathbb{E}\{(v_{i5}+v_{i6}+v_{i7}+v_{i8})(v_{i9}+v_{i10}+v_{i11}+v_{i12})\},\\
\sigma_{24}^2 = \sigma_{42}^2 = \mathbb{E}\{(v_{i5}+v_{i6}+v_{i7}+v_{i8})(v_{i13}+v_{i14}+v_{i15}+v_{i16})\},\\
\sigma_{34}^2 = \sigma_{43}^2 = \mathbb{E}\{(v_{i9}+v_{i10}+v_{i11}+v_{i12})(v_{i13}+v_{i14}+v_{i15}+v_{i16})\}.
\end{gathered}$$

With tedious algebra, it can be found that

$$\begin{aligned}
\mathbb{E}(v_{i1}^2) &= \mathbb{E}\left[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2\left\{\int y f_0(y|X,\eta_0)dy\right\}^2\right] = B_{21}, \\
\mathbb{E}(v_{i2}^2) &= \mathbb{E}\left[\pi^2(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}Y^2(0)\right] = A_{21}, \\
\mathbb{E}(v_{i3}^2) &= \mathbf{A}_{11}^\top\mathbf{A}^{-1}\mathbf{A}_{11}, \quad \mathbb{E}(v_{i4}^2) = \mathbf{B}_{11}^\top\mathbf{C}^{-1}\mathbf{B}_{11}
\end{aligned}$$

and that $\mathbb{E}(v_{i1}v_{i2}) = \mathbb{E}(v_{i1}v_{i4}) = \mathbb{E}(v_{i3}v_{i4}) = 0$,

$$\begin{aligned}
\mathbb{E}(v_{i1}v_{i3}) &= -\mathbf{A}_{11}^\top\mathbf{A}^{-1}\mathbb{E}[\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2\mathbf{X}Y(0)], \\
\mathbb{E}(v_{i2}v_{i3}) &= -\mathbf{A}_{11}^\top\mathbf{A}^{-1}\mathbb{E}\left[\{\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}^2\{1-\pi(\mathbf{X}^\top\boldsymbol{\beta}_0)\}\mathbf{X}Y(0)\right], \\
\mathbb{E}(v_{i2}v_{i4}) &= -\mathbf{B}_{11}^\top\mathbf{C}^{-1}\mathbf{B}_{11}.
\end{aligned}$$

Note that $\mathbb{E}(v_{i1}v_{i3}) + \mathbb{E}(v_{i2}v_{i3}) = -\mathbf{A}_{11}^\top\mathbf{A}^{-1}\mathbf{A}_{11}$. In summary, we have

$$\sigma_{11}^2 = A_{21} + B_{21} - \mathbf{A}_{11}^\top\mathbf{A}^{-1}\mathbf{A}_{11} - \mathbf{B}_{11}^\top\mathbf{C}^{-1}\mathbf{B}_{11}.$$

With the same tedious algebra process as σ_{11}^2, we have

$$\begin{aligned}
\sigma_{22}^2 &= A_{22} + B_{22} - \mathbf{A}_{12}^\top\mathbf{A}^{-1}\mathbf{A}_{12} - \mathbf{B}_{12}^\top\mathbf{B}^{-1}\mathbf{B}_{12}, \\
\sigma_{33}^2 &= A_{23} + B_{23} - \mathbf{A}_{13}^\top\mathbf{A}^{-1}\mathbf{A}_{13} - \mathbf{B}_{13}^\top\mathbf{C}^{-1}\mathbf{B}_{13}, \\
\sigma_{44}^2 &= A_{24} + B_{24} - \mathbf{A}_{14}^\top\mathbf{A}^{-1}\mathbf{A}_{14} - \mathbf{B}_{14}^\top\mathbf{B}^{-1}\mathbf{B}_{14}, \\
\sigma_{12}^2 &= \sigma_{21}^2 = A_2 - \mathbf{A}_{11}^\top\mathbf{A}^{-1}\mathbf{A}_{12}, \\
\sigma_{13}^2 &= \sigma_{31}^2 = A_{31} + \mathbf{B}_{31} - \mathbf{A}_{11}^\top\mathbf{A}^{-1}\mathbf{A}_{13} - \mathbf{B}_{11}^\top\mathbf{C}^{-1}\mathbf{B}_{13}, \\
\sigma_{14}^2 &= \sigma_{41}^2 = A_3 - \mathbf{A}_{11}^\top\mathbf{A}^{-1}\mathbf{A}_{14}, \\
\sigma_{23}^2 &= \sigma_{32}^2 = A_4 - \mathbf{A}_{12}^\top\mathbf{A}^{-1}\mathbf{A}_{13}, \\
\sigma_{24}^2 &= \sigma_{42}^2 = A_{32} + \mathbf{B}_{32} - \mathbf{A}_{12}^\top\mathbf{A}^{-1}\mathbf{A}_{14} - \mathbf{B}_{12}^\top\mathbf{B}^{-1}\mathbf{B}_{14}, \\
\sigma_{34}^2 &= \sigma_{43}^2 = A_5 - \mathbf{A}_{13}^\top\mathbf{A}^{-1}\mathbf{A}_{14}.
\end{aligned}$$

References

Pollard, D. (1990). *Empirical Processes: Theory and Applications.* Empirical Processes: Theory and Applications.

Van de Geer, S. (2000). *Empirical Processes in M-estimation*, volume 6. Cambridge university press.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.