Supplementary Material for " A Unified Framework for Specification Tests of Continuous Treatment Effect Models"

Wei Huang
School of Mathematics and Statistics and Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS),
University of Melbourne, Australia
and
Oliver Linton
Faculty of Economics, University of Cambridge
and
Zheng Zhang
Institute of Statistics & Big Data,
Renmin University of China

A Some preliminary results

We recall some preliminary results which have been established in Ai et al. (2021). The following conditions are inherited from Ai et al. (2021):

Assumption A.1. *(i) The support $\mathcal{X}$ of $\boldsymbol{X}$ is a compact subset of $\mathbb{R}^r$. The support $\mathcal{T}$ of the treatment variable T is a compact subset of $\mathbb{R}$. (ii) There exist two positive constants η_1 and η_2 such that*

$$0 < \eta_1 \leq \pi_0(t, \boldsymbol{x}) \leq \eta_2 < \infty \ , \ \forall (t, \boldsymbol{x}) \in \mathcal{T} \times \mathcal{X} \ .$$

Assumption A.2. *There exist $\Lambda_{K_1 \times K_2} \in \mathbb{R}^{K_1 \times K_2}$ and a positive constant $\alpha > 0$ such that*

$$\sup_{(t,\boldsymbol{x}) \in \mathcal{T} \times \mathcal{X}} \left| \rho'^{-1} \left\{ \pi_0(t, \boldsymbol{x}) \right\} - u_{K_1}(t)^\top \Lambda_{K_1 \times K_2} v_{K_2}(\boldsymbol{x}) \right| = O(K^{-\alpha}) \, ,$$

where $\rho(u) = -\exp(-u-1)$ and ρ'^{-1} is the inverse function of ρ'.

Assumption A.3. *(i) For every K_1 and K_2, the smallest eigenvalues of $\mathbb{E}\left[u_{K_1}(T) u_{K_1}(T)^\top\right]$ and $\mathbb{E}\left[v_{K_2}(\boldsymbol{X}) v_{K_2}(\boldsymbol{X})^\top\right]$ are bounded away from zero uniformly in K_1 and K_2. (ii) There are two sequences of constants $\zeta_1(K_1)$ and $\zeta_2(K_2)$ satisfying $\sup_{t \in \mathcal{T}} \|u_{K_1}(t)\| \leq \zeta_1(K_1)$ and $\sup_{\boldsymbol{x} \in \mathcal{X}} \|v_{K_2}(\boldsymbol{x})\| \leq \zeta_2(K_2)$, $K = K_1(N) K_2(N)$ and $\zeta(K) := \zeta_1(K_1) \zeta_2(K_2)$, such that $\zeta(K) K^{-\alpha} \to 0$ and $\zeta(K) \sqrt{K/N} \to 0$ as $N \to \infty$.*

Assumption A.4. *$\zeta(K)\sqrt{K^2/N} \to 0$ and $\sqrt{N}K^{-\alpha} \to 0$.*

See Ai et al. (2021) for a detailed discussion on Assumptions A.1 -A.4. Under these conditions, Ai et al. (2021, Theorem 3) established the following results:

Proposition 1. *Suppose that Assumptions A.1-A.3 hold. Then, we obtain the following:*

$$\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\hat{\pi}_K(t,\boldsymbol{x}) - \pi_0(t,\boldsymbol{x})| = O_p\left[\max\left\{\zeta(K)K^{-\alpha}, \zeta(K)\sqrt{\frac{K}{N}}\right\}\right],$$

$$\int_{\mathcal{T}\times\mathcal{X}} |\hat{\pi}_K(t,\boldsymbol{x}) - \pi_0(t,\boldsymbol{x})|^2 dF_{T,X}(t,\boldsymbol{x}) = O_p\left\{\max\left(K^{-2\alpha}, \frac{K}{N}\right)\right\},$$

$$\frac{1}{N}\sum_{i=1}^{N} |\hat{\pi}_K(T_i,\boldsymbol{X}_i) - \pi_0(T_i,\boldsymbol{X}_i)|^2 = O_p\left\{\max\left(K^{-2\alpha}, \frac{K}{N}\right)\right\}.$$

Furthermore, for any estimand with the form of $\mathbb{E}\{\pi_0(T,\boldsymbol{X})R(T,\boldsymbol{X},Y)\}$, where $R(T,\boldsymbol{X},Y) \in L^1(dF_{T,X,Y})$, Proposition 2 of Ai et al. (2021) provides an asymptotically equivalent representation for the plug-in estimator $N^{-1}\sum_{i=1}^N \widehat{\pi}_K(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)$:

Proposition 2. *Suppose that Assumptions A.1-A.4 hold. For any integrable function $R(T,\boldsymbol{X},Y)$ where $\mathbb{E}\{R(T,\boldsymbol{X},Y)|T=t,\boldsymbol{X}=\boldsymbol{x}\}$ is continuously differentiable. Then,*

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[\widehat{\pi}_K(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i) - \mathbb{E}\{\pi_0(T,\boldsymbol{X})R(T,\boldsymbol{X},Y)\}\right] \\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big[\pi_0(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i) - \mathbb{E}\{\pi_0(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)|T_i,\boldsymbol{X}_i\} \\
&\qquad + \mathbb{E}\{\pi_0(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)|T_i\} - \mathbb{E}\{\pi_0(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)\} \\
&\qquad + \mathbb{E}\{\pi_0(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)|\boldsymbol{X}_i\} - \mathbb{E}\{\pi_0(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)\}\Big] + o_p(1).
\end{aligned}$$

The following conditions are restatements of Assumptions 1-5 listed in the main paper:

Assumption 1. *For all $t \in \mathcal{T}$, given $\boldsymbol{X}$, T is independent of $Y^*(t)$, that is, $Y^*(t) \perp T|\boldsymbol{X}$, for all $t \in \mathcal{T}$.*

Assumption 2. *Under H_0, (i) $\boldsymbol{\theta}^*$ is an interior point of Θ, where Θ is a compact set in $\mathbb{R}^p$; (ii) $\|M_N(\widehat{\boldsymbol{\theta}},\widehat{\pi}_K)\| = \inf_{\boldsymbol{\theta}\in\Theta_\delta}\|M_N(\boldsymbol{\theta},\widehat{\pi}_K)\| + o_P(N^{-1/2})$, where $\Theta_\delta := \{\boldsymbol{\theta}\in\Theta : \|\boldsymbol{\theta}-\boldsymbol{\theta}^*\| \le \delta\}$.*

Assumption 3. *Let $\eta(T,\boldsymbol{X},Y;t)$ be defined in (4.2), $Var\{\eta(T,\boldsymbol{X},Y;t)\} < \infty$ for all $t \in \mathcal{T}$.*

Assumption 4. *(i) $w(t;\boldsymbol{\theta})$ is continuously differentiable in $\boldsymbol{\theta} \in \Theta$ and continuous in $t \in \mathcal{T}$;*

(ii) $g(t;\boldsymbol{\theta})$ is twice continuously differentiable in $\boldsymbol{\theta} \in \Theta$ and $\nabla_{\boldsymbol{\theta}} g(t;\boldsymbol{\theta})$ is continuous in $t \in \mathcal{T}$;

(iii) $\mathbb{E}[m\{Y; g(T; \boldsymbol{\theta}^*)\}|T = t, \boldsymbol{X} = \boldsymbol{x}]$ *is continuously differentiable in* $(t, \boldsymbol{x})$*;*

(iv) $\mathbb{E}\left[\pi_0(T, \boldsymbol{X})m\{Y; g(T; \boldsymbol{\theta})\}w(T; \boldsymbol{\theta})|T = t, \boldsymbol{X} = \boldsymbol{x}\right]$ *is differentiable w.r.t.* $\boldsymbol{\theta}$ *and* $\nabla_{\boldsymbol{\theta}}\mathbb{E}[\pi_0(T, \boldsymbol{X})m\{Y; g(T; \boldsymbol{\theta})\}w(T; \boldsymbol{\theta})]\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$ *is of full (column) rank.*

Assumption 5. *(i)* $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\Theta}|m\{Y; g(T; \boldsymbol{\theta})\}|^{2+\delta}\right] < \infty$ *for some* $\delta > 0$*; (ii) The function class* $\Big\{m\{Y; g(T; \boldsymbol{\theta})\} : \boldsymbol{\theta} \in \Theta\Big\}$ *satisfies:*

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}_1:\|\boldsymbol{\theta}_1-\boldsymbol{\theta}\|<\delta}|m\{Y; g(T; \boldsymbol{\theta}_1)\} - m\{Y; g(T; \boldsymbol{\theta})\}|^2\right]^{1/2} \le C \cdot \delta$$

for any $\boldsymbol{\theta} \in \Theta$ *and any small* $\delta > 0$ *and for some finite positive constant* C*.*

B Proof of Theorem 1

Proof. We first show that, under H_0, $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \xrightarrow{P} 0$. Because, under H_0, $\widehat{\boldsymbol{\theta}}$ (resp. $\boldsymbol{\theta}^*$) is a unique minimizer of $\|N^{-1}\sum_{i=1}^N \widehat{\pi}_K(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta})\|$ (resp. $\|\mathbb{E}[\pi_0(T, \boldsymbol{X})m\{Y; g(T; \boldsymbol{\theta})\}w(T; \boldsymbol{\theta})]\|$), from the theory of M-estimation (van der Vaart, 1998, Theorem 5.7), if the following condition holds:

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{N}\sum_{i=1}^N \widehat{\pi}_K(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta}) - \mathbb{E}\left[\pi_0(T, \boldsymbol{X})m\{Y; g(T; \boldsymbol{\theta})\}w(T; \boldsymbol{\theta})\right]\right\| \xrightarrow{p} 0.$$

Then $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^*$. Note that

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{N}\sum_{i=1}^N \widehat{\pi}_K(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta}) - \mathbb{E}\left[\pi_0(T, \boldsymbol{X})m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta})\right]\right\| \\
&\le \sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{N}\sum_{i=1}^N \left\{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\right\} m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta})\right\| \qquad \text{(B.1)} \\
&+ \sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{N}\sum_{i=1}^N \pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta}) - \mathbb{E}\left[\pi_0(T, \boldsymbol{X})m\{Y; g(T; \boldsymbol{\theta})\}w(T; \boldsymbol{\theta})\right]\right\|. \qquad \text{(B.2)}
\end{aligned}$$

We first show (B.1) is of $o_p(1)$. Using Assumptions 4 and 5, the Cauchy-Schwarz inequality and Proposition 1, we have that

$$\begin{aligned}
|(\text{B.1})| &\le \left\{\frac{1}{N}\sum_{i=1}^N \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\}^2\right\}^{1/2} \cdot \sup_{\boldsymbol{\theta}\in\Theta}\left\{\frac{1}{N}\sum_{i=1}^N \|m\{Y; g(T; \boldsymbol{\theta})\}w(T; \boldsymbol{\theta})\|^2\right\}^{1/2} \\
&\le o_p(1).
\end{aligned}$$

We next show (B.2) is of $o_p(1)$. Note that, by the law of large numbers, for every $\boldsymbol{\theta} \in \Theta$, $\left\|N^{-1}\sum_{i=1}^N \pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta}) - \mathbb{E}\left[\pi_0(T, \boldsymbol{X})m\{Y; g(T; \boldsymbol{\theta})\}w(T; \boldsymbol{\theta})\right]\right\| \xrightarrow{P}$

0 holds. By Assumptions A.1, 4 (i) and 5, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\boldsymbol{\theta}_1:\|\boldsymbol{\theta}_1-\boldsymbol{\theta}\|<\delta}\|\pi_0(T,\boldsymbol{X})m\{Y;g(T;\boldsymbol{\theta}_1)\}w(T;\boldsymbol{\theta}_1)-\pi_0(T,\boldsymbol{X})m\{Y;g(T;\boldsymbol{\theta})\}w(T;\boldsymbol{\theta})\|^2\right]\\
\leq&O(1)\cdot\mathbb{E}\left[\sup_{\boldsymbol{\theta}_1:\|\boldsymbol{\theta}_1-\boldsymbol{\theta}\|<\delta}\|m\{Y;g(T;\boldsymbol{\theta}_1)\}-m\{Y;g(T;\boldsymbol{\theta})\}\|^2\right]+O(1)\cdot\delta^2\\
\leq&O(1)\cdot\delta^2. \qquad \text{(B.3)}
\end{aligned}$$

With (B.3), Assumptions 2 (i), and Andrews (1994, Theorems 4 and 5), the class $\{\pi_0(T,\boldsymbol{X})m\{Y;g(T;\boldsymbol{\theta})\}w(T;\boldsymbol{\theta}):\boldsymbol{\theta}\in\Theta\}$ is stochastically equicontinuous. Then we have that (B.2) is of $o_p(1)$. Hence we have that $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|\xrightarrow{P}0$ under H_0.

We start to derive the asymptotic distribution of $\widehat{J}_N(t)$. Note that

$$\begin{aligned}
\widehat{U}_i=&U_i+\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\\
&+\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\\
&+\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right],
\end{aligned}$$

where $U_i=\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}$. Then, we have

$$\begin{aligned}
\widehat{J}_N(t)=&\frac{1}{\sqrt{N}}\sum_{i=1}^N\widehat{U}_i\mathscr{H}(T_i,t)=\frac{1}{\sqrt{N}}\sum_{i=1}^N U_i\mathscr{H}(T_i,t) &\text{(B.4)}\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^N\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\mathscr{H}(T_i,t) &\text{(B.5)}\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^N\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t) &\text{(B.6)}\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^N\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t). &\text{(B.7)}
\end{aligned}$$

The subsequent proof consists of the following key steps:

Step 1. Establishing the asymptotically equivalent representation for (B.5) in terms of *i.i.d.* summations;

Step 2. Establishing the asymptotically equivalent representation for (B.6) in terms of *i.i.d.* summations;

Step 3. Showing (B.7) is of $o_P(1)$.

Using Proposition 2, under $H_0 : \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}|T_i] = 0$, we have

$$
\begin{aligned}
(\text{B.5}) = &-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\pi_0(T_i, \boldsymbol{X}_i)\cdot\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\cdot\mathscr{H}(T_i, t)|T_i, \boldsymbol{X_i}] \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\cdot\mathscr{H}(T_i, t)|\boldsymbol{X}_i] + o_P(1) \\
= &-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\phi(T_i, \boldsymbol{X}_i; t) + o_P(1). \qquad (\text{B.8})
\end{aligned}
$$

We next find the expression for $\sqrt{N}\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\}$ by applying Pakes and Pollard (1989, Theorem 3.3). We begin to verify the Conditions (i)-(v) imposed in Pakes and Pollard (1989, Theorem 3.3).

- For Condition (i) of Pakes and Pollard (1989, Theorem 3.3). By Assumption 2,

$$\|M_N(\widehat{\boldsymbol{\theta}}, \widehat{\pi}_K)\| = \inf_{\boldsymbol{\theta}\in\Theta_\delta}\|M_N(\boldsymbol{\theta}, \widehat{\pi}_K)\| + o_P(N^{-1/2}),$$

where

$$M_N(\boldsymbol{\theta}, \pi) := \frac{1}{N}\sum_{i=1}^{N}\pi(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta}).$$

By Proposition 2, we have

$$
\begin{aligned}
&M_N(\boldsymbol{\theta}, \widehat{\pi}_K) \\
=&\frac{1}{N}\sum_{i=1}^{N}\Big[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta}) - \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta})|T_i, \boldsymbol{X}_i] \\
&\quad + \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta})|T_i] + \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta})|\boldsymbol{X}_i] \\
&\quad - \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta})]\Big] + o_P(N^{-1/2}) \\
=:&G_N(\boldsymbol{\theta}) + o_P(N^{-1/2}).
\end{aligned}
$$

where the definition of $G_N(\boldsymbol{\theta})$ is obvious and the equation holds uniformly in $\boldsymbol{\theta}$. Now we have

$$\|G_N(\widehat{\boldsymbol{\theta}})\| = \inf_{\boldsymbol{\theta}\in\Theta_\delta}\|G_N(\boldsymbol{\theta})\| + o_P(N^{-1/2}),$$

thus Condition (i) of Pakes and Pollard (1989, Theorem 3.3) holds.

- For Condition (ii) of Pakes and Pollard (1989, Theorem 3.3). Let

$$G(\boldsymbol{\theta}) := \mathbb{E}[G_N(\boldsymbol{\theta})] = \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta})],$$

and Assumption 4 (iii) ensures that the derivative $\nabla_{\boldsymbol{\theta}}G(\boldsymbol{\theta}^*)$ is full rank. Hence, Condition (ii) of Pakes and Pollard (1989, Theorem 3.3) holds.

- For Condition (iii) of Pakes and Pollard (1989, Theorem 3.3). Let

$$\nu_N(f) := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} [f(T_i, \boldsymbol{X}_i, Y_i) - \mathbb{E}\{f(T_i, \boldsymbol{X}_i, Y_i)\}]$$

be the empirical process indexed by $f(\cdot)$. Assumptions A.1, 4, Assumption 5 and the compactness of Θ imply the empirical processes

$$\Big\{\nu_N \left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{Y_i; g(T_i; \boldsymbol{\theta})\right\} w(T_i; \boldsymbol{\theta})\right] : \boldsymbol{\theta} \in \Theta\Big\},$$

$$\Big\{\nu_N \left(\mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) m \left\{Y_i; g(T_i; \boldsymbol{\theta})\right\} w(T_i; \boldsymbol{\theta}) | T_i, \boldsymbol{X}_i]\right) : \boldsymbol{\theta} \in \Theta\Big\},$$

$$\Big\{\nu_N \left(\mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) m \left\{Y_i; g(T_i; \boldsymbol{\theta})\right\} w(T_i; \boldsymbol{\theta}) | T_i]\right) : \boldsymbol{\theta} \in \Theta\Big\},$$

and

$$\Big\{\nu_N \left(\mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) m \left\{Y_i; g(T_i; \boldsymbol{\theta})\right\} w(T_i; \boldsymbol{\theta}) | \boldsymbol{X}_i]\right) : \boldsymbol{\theta} \in \Theta\Big\},$$

are stochastically equicontinuous (Andrews (1994, Theorems 4 and 5)). Note that

$$\begin{aligned}\sqrt{N}\{G_N(\boldsymbol{\theta}) - G(\boldsymbol{\theta})\} =& \nu_N(\pi_0(T_i, \boldsymbol{X}_i) m \left\{Y_i; g(T_i; \boldsymbol{\theta})\right\} w(T_i; \boldsymbol{\theta})) \\ & - \nu_N(\mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{Y_i; g(T_i; \boldsymbol{\theta})\right\} w(T_i; \boldsymbol{\theta}) | T_i, \boldsymbol{X}_i\right]) \\ & + \nu_N(\mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{Y_i; g(T_i; \boldsymbol{\theta})\right\} w(T_i; \boldsymbol{\theta}) | T_i\right]) \\ & + \nu_N(\mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{Y_i; g(T_i; \boldsymbol{\theta})\right\} w(T_i; \boldsymbol{\theta}) | \boldsymbol{X}_i\right])\end{aligned}$$

Then for every sequence $\{\delta_N\}$ of positive numbers that converges to zero,

$$\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \le \delta_N} \frac{\sqrt{N}\|G_N(\boldsymbol{\theta}) - G(\boldsymbol{\theta}) - G_N(\boldsymbol{\theta}^*)\|}{1 + \sqrt{N}\{\|G_N(\boldsymbol{\theta})\| + \|G(\boldsymbol{\theta})\|\}} = o_P(1).$$

Thus, Condition (iii) of Pakes and Pollard (1989, Theorem 3.3) holds.

- The Condition (iv) of Pakes and Pollard (1989, Theorem 3.3) is satisfied by noting that under $H_0 : \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\} w(T_i; \boldsymbol{\theta}^*) | T_i] = 0$ and

$$\begin{aligned}\sqrt{N} G_N(\boldsymbol{\theta}^*) =& \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Big[\pi_0(T_i, \boldsymbol{X}_i) m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\} w(T_i; \boldsymbol{\theta}^*) \\ & - \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\} w(T_i; \boldsymbol{\theta}^*) | T_i, \boldsymbol{X}_i] \\ & + \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\} w(T_i; \boldsymbol{\theta}^*) | \boldsymbol{X}_i]\Big]\end{aligned}$$

is a sum of i.i.d. random variables of mean zero.

- The Condition (v) of Pakes and Pollard (1989, Theorem 3.3), i.e. $\boldsymbol{\theta}^*$ is an interior point of Θ, is satisfied by Assumption 2 (i).

Therefore, all conditions of Pakes and Pollard (1989, Theorem 3.3) hold, and we get

$$
\begin{aligned}
&\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right\} \qquad \text{(B.9)}\\
=&\Bigg\{-\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)w(T_i;\boldsymbol{\theta}^*)^\top\right]\\
&\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot w(T_i;\boldsymbol{\theta}^*)\nabla_{\boldsymbol{\theta}}^\top g(T_i;\boldsymbol{\theta}^*)\right]\Bigg\}^{-1}\\
&\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)w(T_i;\boldsymbol{\theta}^*)^\top\right]\\
&\times\frac{1}{\sqrt{N}}\sum_{i=1}^N\Big\{\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}w(T_i;\boldsymbol{\theta}^*)\\
&\qquad-\pi_0(T_i,\boldsymbol{X}_i)w(T_i;\boldsymbol{\theta}^*)\cdot\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\\
&\qquad+\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)w(T_i;\boldsymbol{\theta}^*)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|\boldsymbol{X}_i]\Big\}+o_P(1).
\end{aligned}
$$

Consider the term (B.6). Note that

$$
\begin{aligned}
\text{(B.6)}=&\nu_N\left\{\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t)\right\}\\
&+\sqrt{N}\cdot\mathbb{E}\left\{\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t)\right\}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}.
\end{aligned}
$$

By Assumption 5, the compactness of Θ, and Andrews (1994, Theorems 4 and 5), then the empirical process

$$
\left\{\nu_N\left[\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t)\right]:\boldsymbol{\theta}\in\Theta\right\}
$$

is stochastically equicontinuous. With $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|\xrightarrow{p}0$ under H_0, we have

$$
\nu_N\left\{\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t)\right\}=o_P(1).
$$

Using the mean value theorem and $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|\xrightarrow{p}0$ under H_0, we have

$$
\begin{aligned}
&\sqrt{N}\cdot\mathbb{E}\left\{\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t)\right\}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}\\
=&\left\{\nabla_{\boldsymbol{\theta}}\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta})\right\}\mathscr{H}(T_i,t)\right]\Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}\right\}^\top\cdot\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right\}\\
=&\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)^\top\mathscr{H}(T_i,t)\right]\cdot\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right\}+o_P(1).
\end{aligned}
$$

By (B.9), we have

$$
\text{(B.6)}=-\frac{1}{\sqrt{N}}\sum_{i=1}^N\psi(T_i,\boldsymbol{X}_i,Y_i;t)+o_p(1), \qquad \text{(B.10)}
$$

where

$$
\begin{aligned}
\psi(T_i, \boldsymbol{X}_i, Y_i; t) :=& \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} | T_i, \boldsymbol{X}_i] \cdot \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*)^\top \mathscr{H}(T_i, t)\right] \\
& \times \left\{\mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} | T_i, \boldsymbol{X}_i] \cdot \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*) w(T_i; \boldsymbol{\theta}^*)^\top\right]\right. \\
& \quad \left. \cdot \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} | T_i, \boldsymbol{X}_i] \cdot w(T_i; \boldsymbol{\theta}^*) \nabla_{\boldsymbol{\theta}}^\top g(T_i; \boldsymbol{\theta}^*)\right]\right\}^{-1} \\
& \times \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} | T_i, \boldsymbol{X}_i] \cdot \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*) w(T_i; \boldsymbol{\theta}^*)^\top\right] \\
& \times \left\{\pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} w(T_i; \boldsymbol{\theta}^*)\right. \\
& \quad - \pi_0(T_i, \boldsymbol{X}_i) w(T_i; \boldsymbol{\theta}^*) \cdot \mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} | T_i, \boldsymbol{X}_i] \\
& \quad \left. + \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) w(T_i; \boldsymbol{\theta}^*) m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} | \boldsymbol{X}_i]\right\}.
\end{aligned}
$$

For the term (B.7), we have

$$
\begin{aligned}
|(\text{B.7})| =& \left|\frac{1}{\sqrt{N}} \sum_{i=1}^N \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\} \left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\right] \mathscr{H}(T_i, t)\right| \\
\leq& \sqrt{N} \cdot \sup_{(t, \boldsymbol{x}) \in \mathcal{T} \times \mathcal{X}} |\widehat{\pi}_K(t, \boldsymbol{x}) - \pi_0(t, \boldsymbol{x})| \\
& \cdot \frac{1}{N} \sum_{i=1}^N \left|m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} \mathscr{H}(T_i, t)\right| \\
=& \sqrt{N} \cdot O_P\left(\zeta(K) K^{-\alpha} + \zeta(K) \sqrt{\frac{K}{N}}\right) \\
& \cdot \left\{\mathbb{E}\left[\left|m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\right| \cdot |\mathscr{H}(T_i, t)|\right] + O_P\left(N^{-1/2}\right)\right\} \\
\leq& O_P\left(\zeta(K) K^{-\alpha} + \zeta(K) \sqrt{\frac{K}{N}}\right) \cdot \sqrt{N} \cdot \left\{O(1) \cdot \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| + O_P\left(N^{-1/2}\right)\right\} \\
=& o_P(1), && \text{(B.11)}
\end{aligned}
$$

where the second equality holds by Proposition 1 and the law of large numbers; the second inequality holds by Assumption 5; and the last equality holds by (B.9) and Assumption A.3.

Hence, combining (B.4), (B.8), (B.10), and (B.11), we have

$$
\begin{aligned}
\widehat{J}_N(t) =& \frac{1}{\sqrt{N}} \sum_{i=1}^N \eta(T_i, \boldsymbol{X}_i, Y_i; t) + o_P(1) \\
=& \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{U_i \mathscr{H}(T_i, t) - \phi(T_i, \boldsymbol{X}_i; t) - \psi(T_i, \boldsymbol{X}_i, Y_i; t)\right\} + o_P(1), \ \forall t \in \mathcal{T},
\end{aligned}
$$

where $\mathbb{E}\{\phi(T_i, \boldsymbol{X}_i; t)\} = 0$ and $\mathbb{E}\{\psi(T_i, \boldsymbol{X}_i, Y_i; t)\} = 0$. We know that

$$\mathbb{E}\left[\int_{\mathcal{T}}\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta(T_i, \boldsymbol{X}_i, Y_i; t)\right\}^2\right] = \int_{\mathcal{T}}\mathbb{E}[\{\eta(T_i, \boldsymbol{X}_i, Y_i; t)\}^2]dt < \infty,$$

that is, $N^{-1/2}\sum_{i=1}^{N}\eta(T_i, \boldsymbol{X}_i, Y_i; \cdot)$ is tight. Hence, by the functional central limit theorem for Hilbert-valued random arrays Li et al. (2003, Lemma 2.1), we have that under the null hypothesis H_0, $\widehat{J}_N(\cdot)$ weakly converges to $J_\infty(\cdot)$ in $L_2(\mathcal{T}, dt)$, where $J_\infty(\cdot)$ is a Gaussian process with zero mean and covariance function given by

$$\Sigma(t, t') = \mathbb{E}\left\{\eta(T_i, \boldsymbol{X}_i, Y_i; t)\eta(T_i, \boldsymbol{X}_i, Y_i; t')\right\}.$$

Hence, (i) and (ii) are proved.

(iii) Obviously, $h(J) := \int\{J(t)\}^2 dF_T(t)$ is a continuous function in $L_2(\mathcal{T}, dF_T)$. Given that $F_T(\cdot)$ is absolutely continuous with respect to the Lebesgue measure, $h(J)$ is also continuous in $L_2(\mathcal{T}, dt)$. Therefore, by Theorem 1 (i) and the continuous mapping theorem, we have that $h(\widehat{J}_N) = \int\{\widehat{J}_N(t)\}^2 dF_T(t)$ converges to $\int\{J_\infty(t)\}^2 dF_T(t)$ in distribution. By applying a similar argument to the proof of Theorem 2.2 (ii) of Li et al. (2003), we have $|\widehat{CM}_N - h(\widehat{J}_N)| = o_P(1)$. This completes the proof of Theorem 1 (iii). Part (iv) follows from Theorem 1 (i) and the continuous mapping theorem. □

C Proof of Theorem 2

Similar to Theorem 1, results (i) and (ii) can be established. We next prove $\Sigma_0(t, t) > \Sigma(t, t)$ for any fixed $t \in \mathcal{T}$. Let

$$\begin{aligned}
A_t :=& \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i; \boldsymbol{\theta}^*)^\top\mathscr{H}(T_i, t)\right] \\
&\times\left\{\mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i; \boldsymbol{\theta}^*)w(T_i; \boldsymbol{\theta}^*)^\top\right]\right. \\
&\qquad\left.\cdot\,\mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i]\cdot w(T_i; \boldsymbol{\theta}^*)\nabla_{\boldsymbol{\theta}}^\top g(T_i; \boldsymbol{\theta}^*)\right]\right\}^{-1} \\
&\times\mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i; \boldsymbol{\theta}^*)w(T_i; \boldsymbol{\theta}^*)^\top\right]
\end{aligned}$$

.

Then

$$\begin{aligned}
\psi(T_i, \boldsymbol{X}_i, Y_i; t) :=& A_t\cdot\left\{\pi_0(T_i, \boldsymbol{X}_i)m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}w(T_i; \boldsymbol{\theta}^*)\right. \\
&\quad - \pi_0(T_i, \boldsymbol{X}_i)w(T_i; \boldsymbol{\theta}^*)\cdot\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i] \\
&\quad\left. + \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)w(T_i; \boldsymbol{\theta}^*)m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|\boldsymbol{X}_i]\right\},
\end{aligned}$$

and

$$\begin{aligned}\phi(T_i,\boldsymbol{X}_i,Y_i;t) :=&\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}|T_i,\boldsymbol{X}_i]\cdot\mathscr{H}(T_i,t)\\&-\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\mathscr{H}(T_i,t)|\boldsymbol{X}_i].\end{aligned}$$

We have

$$\begin{aligned}\Sigma(t,t)&=\mathbb{E}\left[\{U_i\mathscr{H}(T_i,t)-\phi(T_i,\boldsymbol{X}_i;t)-\psi(T_i,\boldsymbol{X}_i,Y_i;t)\}^2\right]\\
=&\mathbb{E}\left[\left\{U_i\mathscr{H}(T_i,t)-A_t\cdot\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}w(T_i;\boldsymbol{\theta}^*)\right\}^2\right]\\
&+\mathbb{E}\left[\left\{\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}|T_i,\boldsymbol{X}_{\boldsymbol{i}}]\cdot(\mathscr{H}(T_i,t)-A_t\cdot w(T_i;\boldsymbol{\theta}^*))\right\}^2\right]\\
&+\mathbb{E}\left[\left\{\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\cdot(\mathscr{H}(T_i,t)-A_t\cdot w(T_i;\boldsymbol{\theta}^*))|\boldsymbol{X}_i]\right\}^2\right]\\
&-2\cdot\mathbb{E}\left[\left\{\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}|T_i,\boldsymbol{X}_{\boldsymbol{i}}]\cdot(\mathscr{H}(T_i,t)-A_t\cdot w(T_i;\boldsymbol{\theta}^*))\right\}^2\right]\\
&+2\cdot\mathbb{E}\left[\left\{\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\cdot(\mathscr{H}(T_i,t)-A_t\cdot w(T_i;\boldsymbol{\theta}^*))|\boldsymbol{X}_i]\right\}^2\right]\\
&-2\cdot\mathbb{E}\left[\left\{\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\cdot(\mathscr{H}(T_i,t)-A_t\cdot w(T_i;\boldsymbol{\theta}^*))|\boldsymbol{X}_i]\right\}^2\right]\\
=&\mathbb{E}\left[\left\{U_i\mathscr{H}(T_i,t)-A_t\cdot\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}w(T_i;\boldsymbol{\theta}^*)\right\}^2\right]\\
&+\mathbb{E}\left[\left\{\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\cdot(\mathscr{H}(T_i,t)-A_t\cdot w(T_i;\boldsymbol{\theta}^*))|\boldsymbol{X}_i]\right\}^2\right]\\
&-\mathbb{E}\left[\left\{\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}|T_i,\boldsymbol{X}_{\boldsymbol{i}}]\cdot(\mathscr{H}(T_i,t)-A_t\cdot w(T_i;\boldsymbol{\theta}^*))\right\}^2\right]\\
<&\mathbb{E}\left[\left\{U_i\mathscr{H}(T_i,t)-A_t\cdot\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}w(T_i;\boldsymbol{\theta}^*)\right\}^2\right]=\Sigma_0(t,t),\end{aligned}$$

where the second equality holds by using the tower property of the conditional expectation, the inequality holds by using Jensen's inequality.

D Proof of Theorem 5

Proof. We prove parts (i) and (ii). The proof is similar to that for Theorem 1. Let

$$g_N(t,\boldsymbol{\theta}) := g(t;\boldsymbol{\theta})+\frac{\delta(t)}{\sqrt{N}}\text{ and }U_{iN}=\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g_N(T_i;\boldsymbol{\theta}_N^*)\}.$$

Obviously, $g_N(t,\boldsymbol{\theta})\to g(t,\boldsymbol{\theta})$ and $U_{iN}\xrightarrow{a.s.}U_i$. Then

$$\widehat{U}_i = U_{iN}+\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}m\{Y_i;g_N(T_i;\boldsymbol{\theta}_N^*)\}$$

$$+ \pi_0(T_i, \boldsymbol{X}_i)\left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\right\}\right]$$
$$+ \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\}\left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\right\}\right].$$

Then, we have

$$\widehat{J}_N(t) = \frac{1}{\sqrt{N}}\sum_{i=1}^N \widehat{U}_i \mathscr{H}(T_i, t) = \frac{1}{\sqrt{N}}\sum_{i=1}^N U_{iN}\mathscr{H}(T_i, t) \tag{D.1}$$
$$+ \frac{1}{\sqrt{N}}\sum_{i=1}^N \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\} m\left\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\right\}\mathscr{H}(T_i, t) \tag{D.2}$$
$$+ \frac{1}{\sqrt{N}}\sum_{i=1}^N \pi_0(T_i, \boldsymbol{X}_i)\left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\right\}\right]\mathscr{H}(T_i, t) \tag{D.3}$$
$$+ \frac{1}{\sqrt{N}}\sum_{i=1}^N \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\}\left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\right\}\right]\mathscr{H}(T_i, t). \tag{D.4}$$

Obviously, by Chebyshev's inequality, we have

$$(\text{D.1}) = \frac{1}{\sqrt{N}}\sum_{i=1}^N U_i\mathscr{H}(T_i, t) + \frac{1}{\sqrt{N}}\sum_{i=1}^N (U_{iN} - U_i)\mathscr{H}(T_i, t) = \frac{1}{\sqrt{N}}\sum_{i=1}^N U_i\mathscr{H}(T_i, t) + o_P(1).$$

Using Proposition 2, under $H_L : \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\}|T_i] = 0$, we have

$$\begin{aligned}
(\text{D.2}) =& -\frac{1}{\sqrt{N}}\sum_{i=1}^N \pi_0(T_i, \boldsymbol{X}_i)\cdot\mathbb{E}[m\left\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\right\}\cdot\mathscr{H}(T_i, t)|T_i, \boldsymbol{X_i}] \\
&+ \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\left\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\right\}\cdot\mathscr{H}(T_i, t)|\boldsymbol{X}_i] + o_P(1) \\
=& -\frac{1}{\sqrt{N}}\sum_{i=1}^N \pi_0(T_i, \boldsymbol{X}_i)\cdot\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\cdot\mathscr{H}(T_i, t)|T_i, \boldsymbol{X_i}] \\
&+ \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\cdot\mathscr{H}(T_i, t)|\boldsymbol{X}_i] + o_P(1) \\
=& -\frac{1}{\sqrt{N}}\sum_{i=1}^N \phi(T_i, \boldsymbol{X}_i; t) + o_P(1),
\end{aligned}$$

where the second equality holds by using Chebyshev's inequality.

We consider the term (D.3). We first find the expression for $\sqrt{N}\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N^*\}$. Similar to (B.9) in the proof of Theorem 1, by applying Pakes and Pollard (1989, Theorem 3.3), we have

$$\sqrt{N}\left\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right\} \tag{D.5}$$

$$
\begin{aligned}
= - &\Bigg\{ \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i] \cdot \nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*) w(T_i;\boldsymbol{\theta}^*)^\top\right] \\
&\quad \cdot \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i] \cdot w(T_i;\boldsymbol{\theta}^*) \nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*)^\top\right] \Bigg\}^{-1} \\
&\times \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i] \cdot \nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*) w(T_i;\boldsymbol{\theta}^*)^\top\right] \\
&\times \frac{1}{\sqrt{N}} \sum_{i=1}^N \Bigg\{ \pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\} w(T_i;\boldsymbol{\theta}^*) \\
&\qquad\qquad - \pi_0(T_i, \boldsymbol{X}_i) w(T_i;\boldsymbol{\theta}^*) \cdot \mathbb{E}[m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i] \\
&\qquad\qquad + \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) w(T_i;\boldsymbol{\theta}^*) m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}|\boldsymbol{X}_i] \Bigg\} + o_P(1).
\end{aligned}
$$

We next find the expression for $\sqrt{N}\{\boldsymbol{\theta}^* - \boldsymbol{\theta}^*_N\}$. Note that under the local alternative $H_L : \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g_N(T_i;\boldsymbol{\theta}^*_N)\}|T_i] = 0$, using the mean value theorem, we have

$$
\begin{aligned}
0 =& \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\} w(T_i;\boldsymbol{\theta}^*)\right] \\
=& \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g_N(T_i;\boldsymbol{\theta}^*)\right\} w(T_i;\boldsymbol{\theta}^*)\right] \\
&- \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i; \widetilde{g}_N(T_i;\boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i\right] \cdot w(T_i;\boldsymbol{\theta}^*) \cdot \frac{\delta(T)}{\sqrt{N}}\right] \\
=& \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g_N(T_i;\boldsymbol{\theta}^*_N)\right\} w(T_i;\boldsymbol{\theta}^*)\right] \\
&+ \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i; g_N(T_i;\widetilde{\boldsymbol{\theta}})\right\}|T_i, \boldsymbol{X}_i\right] w(T_i;\boldsymbol{\theta}^*) \nabla_\theta^\top g(T_i;\widetilde{\boldsymbol{\theta}})\right] \cdot \{\boldsymbol{\theta}^* - \boldsymbol{\theta}^*_N\} \\
&- \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i; \widetilde{g}_N(T_i;\boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i\right] \cdot w(T_i;\boldsymbol{\theta}^*) \cdot \frac{\delta(T)}{\sqrt{N}}\right] \\
=& \left\{\mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i\right] w(T_i;\boldsymbol{\theta}^*) \nabla_\theta^\top g(T_i;\boldsymbol{\theta}^*)\right] + o_P(1)\right\} \{\boldsymbol{\theta}^* - \boldsymbol{\theta}^*_N\} \\
&- \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i\right] \cdot w(T_i;\boldsymbol{\theta}^*) \cdot \frac{\delta(T)}{\sqrt{N}}\right] + o_P\left(\frac{1}{\sqrt{N}}\right),
\end{aligned}
$$

where $\widetilde{\boldsymbol{\theta}}$ lies on the line joining from $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^*_N$, and $\widetilde{g}_N(T_i;\boldsymbol{\theta}) := g(T_i;\boldsymbol{\theta}) + \gamma \cdot \delta(T)/\sqrt{N}$ for some $\gamma \in (0, 1)$. Then

$$
\begin{aligned}
\sqrt{N}\{\boldsymbol{\theta}^* - \boldsymbol{\theta}^*_N\} =& \Bigg\{ \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i\right] \cdot \nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*) w(T_i;\boldsymbol{\theta}^*)^\top\right] \\
&\quad \cdot \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i\right] \cdot w(T_i;\boldsymbol{\theta}^*) \nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*)^\top\right] \Bigg\}^{-1} \\
&\times \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i\right] \cdot \nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*) w(T_i;\boldsymbol{\theta}^*)^\top\right] \\
&\times \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*_N))\right\}\Big|T_i, \boldsymbol{X}_i\right] \cdot \delta(T_i) \cdot w(T_i;\boldsymbol{\theta}^*)\right] + o_P(1).
\end{aligned}
$$

Hence, we get

$$
\begin{aligned}
&\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_N^*\right\}\\
=&\Bigg\{\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)w(T_i;\boldsymbol{\theta}^*)^\top\right]\\
&\quad\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot w(T_i;\boldsymbol{\theta}^*)\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)^\top\right]\Bigg\}^{-1}\\
&\times\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)w(T_i;\boldsymbol{\theta}^*)^\top\right]\\
&\times\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}_N^*))\right\}\middle|T_i,\boldsymbol{X}_i\right]\cdot\delta(T_i)\cdot w(T_i;\boldsymbol{\theta}^*)\right]\\
&-\Bigg\{\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)w(T_i;\boldsymbol{\theta}^*)^\top\right]\\
&\quad\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot w(T_i;\boldsymbol{\theta}^*)\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)^\top\right]\Bigg\}^{-1}\\
&\quad\times\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)w(T_i;\boldsymbol{\theta}^*)^\top\right]\\
&\quad\times\frac{1}{\sqrt{N}}\sum_{i=1}^N\Big\{\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}w(T_i;\boldsymbol{\theta}^*)\\
&\qquad\qquad-\pi_0(T_i,\boldsymbol{X}_i)w(T_i;\boldsymbol{\theta}^*)\cdot\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\\
&\qquad\qquad+\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)w(T_i;\boldsymbol{\theta}^*)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|\boldsymbol{X}_i]\Big\}+o_P(1).
\end{aligned}
$$

Then similar to (B.10), we have

$$
(\text{D.3})=-\frac{1}{\sqrt{N}}\sum_{i=1}^N\psi(T_i,\boldsymbol{X}_i,Y_i;t)+\mu(t)+o_p(1),
$$

where

$$
\begin{aligned}
\mu(t)=&\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)^\top\mathscr{H}(T_i,t)\right]\\
&\times\Bigg\{\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)w(T_i;\boldsymbol{\theta}^*)^\top\right]\\
&\quad\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot w(T_i;\boldsymbol{\theta}^*)\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)^\top\right]\Bigg\}^{-1}
\end{aligned}
$$

$$\times \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g} \mathbb{E}\left[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} | T_i, \boldsymbol{X}_i\right] \cdot \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*) w(T_i; \boldsymbol{\theta}^*)^\top\right]$$
$$\times \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g} \mathbb{E}\left[m\left\{Y_i; g(T_i; \boldsymbol{\theta}_N^*))\right\} \middle| T_i, \boldsymbol{X}_i\right] \cdot \delta(T_i) \cdot w(T_i; \boldsymbol{\theta}^*)\right].$$

Similar to (B.7), we have that (D.4) is of $o_P(1)$.

Hence, we have

$$\begin{aligned}
\widehat{J}_N(t) =& \frac{1}{\sqrt{N}} \sum_{i=1}^N \eta(T_i, \boldsymbol{X}_i, Y_i; t) + \mu(t) + o_P(1) \\
=& \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{U_i \mathscr{H}(T_i, t) - \phi(T_i, \boldsymbol{X}_i; t) - \psi(T_i, \boldsymbol{X}_i, Y_i; t)\right\} + \mu(t) + o_P(1),
\end{aligned}$$

where $\mathbb{E}\{\phi(T_i, \boldsymbol{X}_i; t)\} = 0$ and $\mathbb{E}\{\psi(T_i, \boldsymbol{X}_i, Y_i; t)\} = 0$. Therefore, under the null hypothesis H_0, $\widehat{J}_N(\cdot)$ weakly converges to $J_{\infty,\mu}(\cdot)$ in $L_2(\mathcal{T}, dt)$, where $J_{\infty,\mu}(\cdot)$ is a Gaussian process with mean function $\mu(t)$ and covariance function given by

$$\Sigma(t, t') = \mathbb{E}\left\{\eta(T_i, \boldsymbol{X}_i, Y_i; t)\eta(T_i, \boldsymbol{X}_i, Y_i; t')\right\}.$$

We prove part *(iii)*. Because

$$\frac{1}{\sqrt{N}} \widehat{J}_N(t) = \frac{1}{N} \sum_{i=1}^N \widehat{U}_i \mathscr{H}(T_i, t)$$
$$= \frac{1}{N} \sum_{i=1}^N U_i \mathscr{H}(T_i, t) \tag{D.6}$$
$$+ \frac{1}{N} \sum_{i=1}^N \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\} m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} \mathscr{H}(T_i, t) \tag{D.7}$$
$$+ \frac{1}{N} \sum_{i=1}^N \pi_0(T_i, \boldsymbol{X}_i) \left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\right] \mathscr{H}(T_i, t) \tag{D.8}$$
$$+ \frac{1}{N} \sum_{i=1}^N \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\} \left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\right] \mathscr{H}(T_i, t). \tag{D.9}$$

By applying a similar argument for (B.5)-(B.7), we have that (D.7)-(D.9) are of $o_P(1)$. Under H_1, the law of large numbers implies (D.6) $= \mu_1(t) + o_P(1)$. Hence, we conclude the proof.

□

E Asymptotic properties of $\widehat{J}_N(t;\widehat{\theta}_{opt})$ and $\widehat{CM}_N(\widehat{\theta}_{opt})$

Theorem 3. *Suppose that $m(y;g)$ is differentiable with respect to g, Assumptions 1-5 and Assumptions A.1-A.4 listed in Appendix A hold, then under H_0,*

$$(i)\quad \widehat{J}_N(t;\widehat{\boldsymbol{\theta}}_{opt}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta_{opt}(T_i,\boldsymbol{X}_i,Y_i;t) + o_P(1),$$

$$(ii)\quad \widehat{J}_N(\cdot;\widehat{\boldsymbol{\theta}}_{opt}) \text{ converges weakly to } J_{\infty,opt}(\cdot) \text{ in } L_2\{\mathcal{T},dF_T(t)\},$$

where $J_{\infty,opt}$ is a Gaussian process with zero mean and covariance function given by

$$\Sigma_{opt}(t,t') = \mathbb{E}\left\{\eta_{opt}(T_i,\boldsymbol{X}_i,Y_i;t)\eta_{opt}(T_i,\boldsymbol{X}_i,Y_i;t')\right\}.$$

Furthermore,

$$(iii)\quad \widehat{CM}_N(\widehat{\boldsymbol{\theta}}_{opt}) \text{ converges to } \int\{J_{\infty,opt}(t)\}^2 dF_T(t) \text{ in distribution.}$$

Proof. We first claim $\|\widehat{\boldsymbol{\theta}}_{opt}-\boldsymbol{\theta}^*\| \xrightarrow{P} 0$ under H_0. Since

- Θ is compact;
- by Proposition 1, $|N^{-1}\cdot\widehat{CM}_N(\boldsymbol{\theta}) - CM(\boldsymbol{\theta})| \xrightarrow{P} 0$ for every $\boldsymbol{\theta}\in\Theta$;
- $CM(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$;
- $|\widehat{U}_i(\boldsymbol{\theta})| = |\widehat{\pi}_K(T_i,\boldsymbol{X}_i)m(Y_i;g(T_i;\boldsymbol{\theta}))| \le O_p(1)\times\sup_{\boldsymbol{\theta}\in\Theta}|m(Y_i;g(T_i;\boldsymbol{\theta}))|$ and $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\Theta}|m(Y_i;g(T_i;\boldsymbol{\theta}))|]<\infty$;

then it follows from van der Vaart (1998, Theorem 5.7) that $\|\widehat{\boldsymbol{\theta}}_{opt}-\boldsymbol{\theta}^*\| \xrightarrow{P} 0$.

We then find the asymptotic expression for $\sqrt{N}\{\widehat{\boldsymbol{\theta}}_{opt}-\boldsymbol{\theta}^*\}$. By the first order condition, we get

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{J}_N(T_i;\widehat{\boldsymbol{\theta}}_{opt})\cdot\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widehat{\boldsymbol{\theta}}_{opt}) = 0$$

Using the mean value theorem, we get

$$0 = \frac{1}{N}\sum_{i=1}^{N}\widehat{J}_N(T_i;\boldsymbol{\theta}^*)\cdot\frac{\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\boldsymbol{\theta}^*)}{\sqrt{N}}$$
$$+\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})}{\sqrt{N}}\cdot\frac{\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})^{\top}}{\sqrt{N}}+\frac{\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})}{\sqrt{N}}\cdot\frac{\nabla^2_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})}{\sqrt{N}}\right\}\cdot\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}_{opt}-\boldsymbol{\theta}^*\right\},$$

where $\widetilde{\boldsymbol{\theta}}_{opt}$ lies on the joining from $\widehat{\boldsymbol{\theta}}_{opt}$ to $\boldsymbol{\theta}^*$. Using the fact that $\|\widehat{\boldsymbol{\theta}}_{opt} - \boldsymbol{\theta}^*\| \xrightarrow{p} 0$ and Proposition 1, under H_0, it is easy to obtain

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})}{\sqrt{N}}\cdot\frac{\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})^\top}{\sqrt{N}}+\frac{\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})}{\sqrt{N}}\cdot\frac{\nabla^2_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})}{\sqrt{N}}\right\}\\
&=\int_{\mathcal{T}}\mathbb{E}\left[\pi_0(T,\boldsymbol{X})\cdot\frac{\partial}{g}m\{Y;g(T;\boldsymbol{\theta}^*)\}\nabla_{\boldsymbol{\theta}}g(T;\boldsymbol{\theta}^*)\mathcal{H}(T;t)\right]\\
&\quad\times\mathbb{E}\left[\pi_0(T,\boldsymbol{X})\cdot\frac{\partial}{g}m\{Y;g(T;\boldsymbol{\theta}^*)\}\nabla_{\boldsymbol{\theta}}g(T;\boldsymbol{\theta}^*)^\top\mathcal{H}(T;t)\right]f_T(t)dt+o_P(1)\\
&=\int_{\mathcal{T}}B_tB_t^\top f_T(t)dt+o_P(1),
\end{aligned}
$$

where

$$
B_t := \mathbb{E}\left[\pi_0(T,\boldsymbol{X})\cdot\frac{\partial}{g}m\{Y;g(T;\boldsymbol{\theta}^*)\}\nabla_{\boldsymbol{\theta}}g(T;\boldsymbol{\theta}^*)\mathcal{H}(T;t)\right].
$$

For $\widehat{J}_N(t;\boldsymbol{\theta}^*)$, under $H_0 : \mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}|T=t]=0$, by using Proposition 2, we get

$$
\begin{aligned}
&\widehat{J}_N(t;\boldsymbol{\theta}^*)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\pi}_K(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\mathcal{H}(T_i;t)\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big\{\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\mathcal{H}(T_i;t)-\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}|T_i,\boldsymbol{X}_i]\cdot\mathcal{H}(T_i;t)\\
&\qquad+\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\mathcal{H}(T_i;t)|\boldsymbol{X}_i]\Big\}+o_P(1)\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_{opt}(T_i,\boldsymbol{X}_i,Y_i;t)+o_P(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\varphi_{opt}(T_i,\boldsymbol{X}_i,Y_i;t):=&\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\mathcal{H}(T_i;t)\\
&-\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}|T_i,\boldsymbol{X}_i]\cdot\mathcal{H}(T_i;t)\\
&+\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\mathcal{H}(T_i;t)|\boldsymbol{X}_i]
\end{aligned}
$$

Now, we have

$$
\begin{aligned}
&\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}_{opt}-\boldsymbol{\theta}^*\right\}\\
&=-\left\{\int_{\mathcal{T}}B_tB_t^\top f_T(t)dt\right\}^{-1}\left\{\frac{1}{N}\sum_{i=1}^{N}\widehat{J}_N(T_i;\boldsymbol{\theta}^*)\cdot\frac{\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\boldsymbol{\theta}^*)}{\sqrt{N}}\right\}\\
&=-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\int_{\mathcal{T}}B_tB_t^\top f_T(t)dt\right\}^{-1}\cdot\int_{\mathcal{T}}\varphi_{opt}(T_i,\boldsymbol{X}_i,Y_i;t)\cdot B_t\cdot f_T(t)dt.
\end{aligned}
$$

Table 1: Estimated sizes

$m(\cdot)$	Model	N	Logistic 1%	5%	10%	Cosine-Sine 1%	5%	10%	Indicator 1%	5%	10%
Average	**DGP0-L**	100	0.021	0.067	0.124	0.010	0.064	0.136	0.008	0.058	0.117
		200	0.015	0.056	0.111	0.014	0.062	0.135	0.012	0.056	0.107
		500	0.008	0.051	0.108	0.012	0.059	0.117	0.011	0.048	0.107
	DGP0-NL	100	0.025	0.085	0.153	0.014	0.070	0.134	0.007	0.064	0.119
		200	0.021	0.059	0.119	0.013	0.069	0.131	0.012	0.065	0.110
		500	0.012	0.058	0.110	0.011	0.052	0.105	0.011	0.057	0.111
Median	**DGP0-L**	100	0.035	0.107	0.182	0.016	0.065	0.132	0.031	0.106	0.162
		200	0.022	0.081	0.141	0.016	0.068	0.121	0.025	0.065	0.134
		500	0.017	0.064	0.121	0.011	0.066	0.125	0.013	0.052	0.110
	DGP0-NL	100	0.040	0.124	0.196	0.009	0.063	0.109	0.026	0.097	0.172
		200	0.025	0.078	0.133	0.010	0.073	0.127	0.021	0.074	0.140
		500	0.010	0.059	0.119	0.016	0.053	0.116	0.015	0.072	0.126

Let

$$\psi_{opt}(T_i, \boldsymbol{X}_i, Y_i; t) = \left\{ \int_{\mathcal{T}} B_t^\top f_T(t) dt \right\} \left\{ \int_{\mathcal{T}} B_t B_t^\top f_T(t) dt \right\}^{-1} \int_{\mathcal{T}} \varphi_{opt}(T_i, \boldsymbol{X}_i, Y_i; t) B_t f_T(t) dt.$$

Following a similar argument of establishing Theorem 1, we get

$$\begin{aligned} \widehat{J}_N(t; \widehat{\boldsymbol{\theta}}_{opt}) =& \frac{1}{\sqrt{N}} \sum_{i=1}^N \eta_{opt}(T_i, \boldsymbol{X}_i, Y_i; t) + o_P(1) \\ =& \frac{1}{\sqrt{N}} \sum_{i=1}^N \{ U_i \mathscr{H}(T_i, t) - \phi(T_i, \boldsymbol{X}_i; t) - \psi_{opt}(T_i, \boldsymbol{X}_i, Y_i; t) \} + o_P(1). \end{aligned}$$

The remaining results follow by using a similar argument of establishing Theorem 1. □

F Additional simulation results of KS-type statistic

We also performed the simulation studies described in section 6.2 of the paper using the KS-type statistic. The results are similar to those of the CM-type one.

Tables 1 and 2 summarize the empirical rejection probabilities computed at significance levels 1%, 5%, and 10% for each case, which respectively show the estimated sizes (DGP0-L and DGP0NL) and the estimated powers (DGP1-L and DGP1-NL) of our KS test method.

G Estimating and testing Tobit linear models

Let

$$Y(t) = \boldsymbol{\beta}^\top \boldsymbol{t} + \epsilon,$$

for some unknown parameter $\boldsymbol{\beta}$ in a compact set in $\mathbb{R}^p$, where $\boldsymbol{t} = (1, t, t^2, \ldots, t^{p-1})^\top$ for some positive integer p and ϵ is a normal random variable with mean 0 and unknown

Table 2: Estimated power

			Logistic			Cosine-Sine			Indicator		
$m(\cdot)$	Model	N	1%	5%	10%	1%	5%	10%	1%	5%	10%
Average	**DGP1-L**	100	0.587	0.807	0.885	0.488	0.714	0.825	0.505	0.748	0.850
		200	0.924	0.982	0.996	0.890	0.970	0.986	0.919	0.984	0.998
	DGP1-NL	100	0.483	0.693	0.800	0.444	0.698	0.803	0.382	0.600	0.732
		200	0.721	0.895	0.945	0.801	0.910	0.960	0.762	0.892	0.930
Median	**DGP1-L**	100	0.277	0.533	0.660	0.164	0.372	0.525	0.263	0.525	0.655
		200	0.606	0.818	0.907	0.505	0.7434	0.834	0.612	0.829	0.902
	DGP1-NL	100	0.209	0.399	0.523	0.167	0.363	0.495	0.161	0.365	0.487
		200	0.356	0.593	0.732	0.350	0.625	0.756	0.380	0.632	0.755

variance σ^2. A Tobit linear model assumes the potential outcome

$$Y^*(t) = \begin{cases} Y(t) & \text{if} \quad Y(t) > 0, \\ 0 & \text{if} \quad Y(t) \leq 0\,. \end{cases}$$

It can be shown that the log-likelihood function of $\boldsymbol{\beta}$ and σ given $Y^*(t)$ is

$$\ln f\{Y^*(t), t, \boldsymbol{\beta}, \sigma\} = \sum_{Y_i^*(t)=0} \ln\left[\Phi\left\{-\frac{\boldsymbol{\beta}^\top \boldsymbol{t}}{\sigma}\right\}\right] + \sum_{Y_i^*(t)>0} \ln\left[\sigma^{-1}\phi\left\{\frac{Y_i^*(t) - \boldsymbol{\beta}^\top \boldsymbol{t}}{\sigma}\right\}\right],$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are respectively the distribution function and density function of a standard normal random variable. Olsen (1978) proposed a reparametrization $\boldsymbol{\beta} = \boldsymbol{\delta}/\gamma$ and $\sigma^2 = \gamma^{-2}$, the resulting transformed log-likelihood of the parameter $\boldsymbol{\theta} = (\boldsymbol{\delta}, \gamma)$ is then

$$\ln f\{Y^*(t), t, \boldsymbol{\theta}\} = \sum_{Y_i^*(t)=0} \ln[\Phi\{-(\boldsymbol{\delta}^\top \boldsymbol{t})\}] + \sum_{Y_i^*(t)>0} \ln(\gamma) + \ln[\phi\{\gamma Y_i^*(t) - \boldsymbol{\delta}^\top \boldsymbol{t}\}],$$

which is globally concave in terms of $\boldsymbol{\theta}$.

Note that in this case, we can test the model by testing

$$H_0 : \exists \text{ some } \boldsymbol{\theta}^* \in \Theta, \text{ s.t. } \mathbb{E}[\nabla_{\boldsymbol{\theta}} \ln f\{Y^*(t), t, \boldsymbol{\theta}^*\}] = 0 \text{ for all } t \in \mathcal{T},$$

against the alternative hypothesis

$$H_1 : \nexists \text{ any } \boldsymbol{\theta} \in \Theta, \text{ s.t. } \mathbb{E}[\nabla_{\boldsymbol{\theta}} \ln f\{Y^*(t), t, \boldsymbol{\theta}\}] = 0 \text{ for all } t \in \mathcal{T},$$

where Θ is a compact set in $\mathbb{R}^{p+1}$.

This is a multi-dimensional moment condition. It is straightforward to extend our test method by taking $m\{Y^*(t); g(t; \boldsymbol{\theta})\} = \nabla_{\boldsymbol{\theta}} \ln f\{Y^*(t), t, \boldsymbol{\theta}\}$ and $w(T; \boldsymbol{\theta})$ in (3.6) to be 1. In particular, our Cramer-von Mises (CM)-type statistic and Kolmogorov-Smirnov(KS)-type statistic are extended to

$$CM_N^0 = \frac{1}{N}\sum_{i=1}^{N} \{J_N^0(T_i)\}^\top \{J_N^0(T_i)\} \quad \text{and} \quad KS_N^0 = \sup_{t \in \mathcal{T}} \|J_N^0(t)\|_\infty,$$

where $\|\cdot\|_\infty$ is the maximum norm of a vector and

$$J_N^0(t) = \frac{1}{N}\sum_{i=1}^{N}\pi_0(T_i,\boldsymbol{X}_i)\nabla_{\boldsymbol{\theta}}\ln f(Y_i,T_i,\boldsymbol{\theta})\mathscr{H}(T_i,t)\,.$$

We can estimate $J_N^0(t)$ by

$$\widehat{J}_N(t) = \frac{1}{N}\sum_{i=1}^{N}\widehat{\pi}_K(T_i,\boldsymbol{X}_i)\nabla_{\boldsymbol{\theta}}\ln f(Y_i,T_i,\widehat{\boldsymbol{\theta}}_{LL})\mathscr{H}(T_i,t)\,,$$

where $\widehat{\pi}_K$ is defined the same as that in section 3 and

$$\widehat{\boldsymbol{\theta}}_{LL} := \arg\min_{\boldsymbol{\theta}\in\Theta}\|M_N(\boldsymbol{\theta},\widehat{\pi}_K)\|\,,$$

where

$$M_N(\boldsymbol{\theta},\pi) := \frac{1}{N}\sum_{i=1}^{N}\pi(T_i,\boldsymbol{X}_i)\nabla_{\boldsymbol{\theta}}\ln f(Y_i,T_i,\boldsymbol{\theta})\,.$$

Theorem 1 can be applied here under Assumptions A.1 to 4 in section 4 and

Assumption G.1. *(i)* $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\Theta}\|\nabla_{\boldsymbol{\theta}}\ln f\{Y,T,\boldsymbol{\theta}\}\|^{2+\delta}\right] < \infty$ *for some* $\delta > 0$*; (ii) The function class* $\left\{\nabla_{\boldsymbol{\theta}}\ln f\{Y,T,\boldsymbol{\theta}\} : \boldsymbol{\theta}\in\Theta\right\}$ *satisfies:*

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}_1:\|\boldsymbol{\theta}_1-\boldsymbol{\theta}\|<\delta}\|\nabla_{\boldsymbol{\theta}}\ln f\{Y,T,\boldsymbol{\theta}_1\} - \nabla_{\boldsymbol{\theta}}\ln f\{Y,T,\boldsymbol{\theta}\}\|^2\right]^{1/2} \le C\cdot\delta$$

for any $\boldsymbol{\theta}\in\Theta$ *and any small* $\delta > 0$ *and for a finite positive constant* C*.*

With

$$\begin{aligned}\phi(T_i,\boldsymbol{X}_i;t) :=&\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathscr{H}(T_i,t)\cdot\mathbb{E}[\nabla_{\boldsymbol{\theta}}\ln f(Y_i,T_i,\boldsymbol{\theta})|T_i,\boldsymbol{X}_{\boldsymbol{i}}]\\ &-\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)\nabla_{\boldsymbol{\theta}}\ln f(Y_i,T_i,\boldsymbol{\theta})\cdot\mathscr{H}(T_i,t)|\boldsymbol{X}_i],\end{aligned}$$

and

$$\begin{aligned}\psi(T_i,\boldsymbol{X}_i,Y_i;t) :=&\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\nabla_{\boldsymbol{\theta\theta}}\ln f(Y_i,T_i,\boldsymbol{\theta})\mathscr{H}(T_i,t)\right]\\ &\times\left\{\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\nabla_{\boldsymbol{\theta\theta}}\ln f(Y_i,T_i,\boldsymbol{\theta})\right]\right\}^{-1}\\ &\times\Big\{\pi_0(T_i,\boldsymbol{X}_i)\nabla_{\boldsymbol{\theta}}\ln f(Y_i,T_i,\boldsymbol{\theta})\\ &\quad-\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[\nabla_{\boldsymbol{\theta}}\ln f(Y_i,T_i,\boldsymbol{\theta})|T_i,\boldsymbol{X}_i]\\ &\quad+\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)\nabla_{\boldsymbol{\theta}}\ln f(Y_i,T_i,\boldsymbol{\theta})|\boldsymbol{X}_i]\Big\},\end{aligned}$$

and

$$\eta(T_i,\boldsymbol{X}_i,Y_i;t) := \pi_0(T_i,\boldsymbol{X}_i)\cdot\nabla_{\boldsymbol{\theta}}\ln f(Y_i,T_i,\boldsymbol{\theta})\mathscr{H}(T_i,t) - \phi(T_i,\boldsymbol{X}_i;t) - \psi(T_i,\boldsymbol{X}_i,Y_i;t).$$

The approximation method of the null limiting distribution described in section 5 can be directly applied here.

References

Ai, C., O. Linton, K. Motegi, and Z. Zhang (2021). A unified framework for efficient estimation of general treatment models. *Quantitative Economics 12*(3), 779–816.

Ai, C., O. Linton, and Z. Zhang (2021). Estimation and inference for the counterfactual distribution and quantile functions in continuous treatment models. *Journal of Econometrics*.

Andrews, D. W. K. (1994). Empirical process methods in econometrics. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 37, pp. 2247–2294. Citeseer.

Li, Q., C. Hsiao, and J. Zinn (2003). Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *Journal of Econometrics 112*(2), 295–325.

Olsen, R. J. (1978). Note on the uniqueness of the maximum likelihood estimator for the tobit model. *Econometrica: Journal of the Econometric Society*, 1211–1215.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.