Specification tests for GPS using double projections: Online supplementary appendix

Pedro H. C. Sant'Anna* Xiaojun Song†

April 15, 2023

This online supplementary appendix contains auxiliary lemmas, proofs of the main theoretical results, and Monte Carlo simulations.

*Emory University and Causal Solutions. Email: pedrosantanna@causal-solutions.com

†Guanghua School of Management and Center for Statistical Science, Peking University. Email: sxj@gsm.pku.edu.cn. The author acknowledges the financial support from the National Natural Science Foundation of China (Grant No. 71973005).

Appendix A: Mathematical proofs

Proof of Lemma 2.1: For notational simplicity, let us focus on $J = 1$ (the binary treatment case) and omit the dependence of $e_i(t;\widehat{\theta}_n)$ and $\mathcal{P}_{n,t}1(\beta^\top X_i \le u)$ on t. The general case with $J \ge 2$ follows in a similar way.

First, we can rewrite CvM_n^{dpro} as

$$CvM_n^{dpro} = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{r=1}^n e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)\int_{\mathbb{S}^{d_x}} \mathcal{P}_n 1\left(\beta^\top X_i \le \beta^\top X_r\right)\mathcal{P}_n 1\left(\beta^\top X_j \le \beta^\top X_r\right)d\beta.$$

Recalling the expression of projection operator $\mathcal{P}_n 1(\beta^\top X_i \le u)$, by simple algebra, CvM_n^{dpro} is further equal to

$$\begin{aligned}
CvM_n^{dpro} =& \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{r=1}^n e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)\int_{\mathbb{S}^{d_x}} 1\left(\beta^\top X_i \le \beta^\top X_r\right)1\left(\beta^\top X_j \le \beta^\top X_r\right)d\beta \\
& - 2\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{r=1}^n\sum_{s=1}^n e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)g^\top(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)g(X_s,\widehat{\theta}_n) \\
& \times \int_{\mathbb{S}^{d_x}} 1\left(\beta^\top X_i \le \beta^\top X_r\right)1\left(\beta^\top X_s \le \beta^\top X_r\right)d\beta \\
& + \frac{1}{n^4}\sum_{i=1}^n\sum_{j=1}^n\sum_{r=1}^n\sum_{s=1}^n\sum_{t=1}^n e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)g^\top(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)g(X_s,\widehat{\theta}_n)) \\
& \times g^\top(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)g(X_s,\widehat{\theta}_n)\int_{\mathbb{S}^{d_x}} 1\left(\beta^\top X_s \le \beta^\top X_r\right)1\left(\beta^\top X_t \le \beta^\top X_r\right)d\beta \\
\equiv & B_{n1} - 2B_{n2} + B_{n3}.
\end{aligned}$$

As in Escanciano (2006), if denoting $A_{ijr} = \int_{\mathbb{S}^{d_x}} 1\left(\beta^\top X_i \le \beta^\top X_r\right)1\left(\beta^\top X_j \le \beta^\top X_r\right)d\beta$, we have that

$$B_{n1} = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{r=1}^n e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)A_{ijr}.$$

Similarly,

$$\begin{aligned}
B_{n2} =& \frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{r=1}^n\sum_{s=1}^n e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)g^\top(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)g(X_s,\widehat{\theta}_n)A_{isr} \\
=& \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\sum_{r=1}^n e_i(\widehat{\theta}_n)\left[g^\top(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)\frac{1}{n}\sum_{s=1}^n g(X_s,\widehat{\theta}_n)e_s(\widehat{\theta}_n)\right]A_{ijr},
\end{aligned}$$

where the second equality follows by letting $j = s$, and

$$\begin{aligned}
B_{n3} =& \frac{1}{n^4}\sum_{i=1}^n\sum_{j=1}^n\sum_{r=1}^n\sum_{s=1}^n\sum_{t=1}^n e_i(\widehat{\theta}_n)e_j(\widehat{\theta}_n)g^\top(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)g(X_s,\widehat{\theta}_n)) \\
& g^\top(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)g(X_s,\widehat{\theta}_n)A_{str}
\end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n}\left[g^\top(X_i,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)\frac{1}{n}\sum_{s=1}^{n}g(X_s,\widehat{\theta}_n)e_s(\widehat{\theta}_n)\right]\\
&\left[g^\top(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)\frac{1}{n}\sum_{t=1}^{n}g(X_t,\widehat{\theta}_n)e_t(\widehat{\theta}_n)\right]A_{ijr},
\end{aligned}
$$

where the second equality follows by letting $i = s$ and $j = t$.

As a result,

$$
\begin{aligned}
CvM_n^{dpro} =&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n}\left[e_i(\widehat{\theta}_n)-g^\top(X_i,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)\frac{1}{n}\sum_{s=1}^{n}g(X_s,\widehat{\theta}_n)e_s(\widehat{\theta}_n)\right]\\
&\left[e_j(\widehat{\theta}_n)-g^\top(X_j,\widehat{\theta}_n)\Delta_n^{-1}(\widehat{\theta}_n)\frac{1}{n}\sum_{t=1}^{n}g(X_t,\widehat{\theta}_n)e_t(\widehat{\theta}_n)\right]A_{ijr}\\
\equiv&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{r=1}^{n}e_i^{pro}(\widehat{\theta}_n)e_j^{pro}(\widehat{\theta}_n)A_{ijr}.
\end{aligned}
$$

This completes the proof of Lemma 2.1. □

The proof of Theorem 3.1 is similar to the proofs of Theorem 1 and Corollary 1 in Sant'Anna and Song (2019). We first need to introduce several auxiliary lemmas. Henceforth, $x = (\beta^\top, u)^\top$. The next lemma establishes the uniform asymptotic decomposition of the "once" projected empirical process

$$
R_{n,t}^{pro}(x;\widehat{\theta}_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i(t;\widehat{\theta}_n)1(\beta^\top X_i \le u), \quad x \in \Pi_{pro}.
$$

Lemma A.1 *Suppose Assumptions 3.1-3.3 hold. Then, for each $t \in \mathcal{T}$, we have that*

$$
\sup_{x\in\Pi_{pro}}\left|R_{n,t}^{pro}(x;\widehat{\theta}_n)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i(t;\theta_0)1(\beta^\top X_i\le u)+\sqrt{n}(\widehat{\theta}_n-\theta_0)^\top G_t(x;\theta_0)\right| = o_p(1),
$$

where $G_t(x;\theta) = \mathbb{E}[g_t(X,\theta)1(\beta^\top X \le u)]$ with $g_t(x,\theta) = \partial q_t(x,\theta)/\partial\theta$.

Proof of Lemma A.1: First note that, for each $t \in \mathcal{T}$, $R_{n,t}^{pro}(x;\widehat{\theta}_n)$ can be readily decomposed as

$$
\begin{aligned}
&R_{n,t}^{pro}(x;\widehat{\theta}_n)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i(t;\theta_0)1(\beta^\top X_i\le u)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(q_t(X_i,\widehat{\theta}_n)-q_t(X_i,\theta_0))1(\beta^\top X_i\le u).
\end{aligned}
$$

By the Mean Value Theorem (MVT) and Assumptions 3.2-3.3, the second term in the above expression is simply equal to

$$
\begin{aligned}
&\sqrt{n}(\widehat{\theta}_n-\theta_0)^\top\frac{1}{n}\sum_{i=1}^n\frac{\partial q_t(X_i,\widetilde{\theta}_n)}{\partial\theta}1(\beta^\top X_i\le u)\\
=&\sqrt{n}(\widehat{\theta}_n-\theta_0)^\top\mathbb{E}\left[g_t(X,\theta_0)1(\beta^\top X_i\le u)\right]+o_p(1),
\end{aligned}
$$

with $\widetilde{\theta}_n$ lying between $\widehat{\theta}_n$ and θ_0, where the latter equality follows by the uniform law of large numbers (ULLN) of Newey and McFadden (1994, Lemma 2.4). This finishes the proof of Lemma A.1. $\square$

To proceed, for each $t\in\mathcal{T}$, we introduce the following auxiliary quantity,

$$
A_{n,t}=\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(t;\widehat{\theta}_n)g_t(X_i,\widehat{\theta}_n),
$$

which admits the following decomposition.

Lemma A.2 *Suppose Assumptions 3.1-3.3 hold. Then, under the null hypothesis H_0 in (2.1), for each $t\in\mathcal{T}$, we have that*

$$
A_{n,t}=\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(t;\theta_0)g_t(X_i,\theta_0)-\Delta_t(\theta_0)\sqrt{n}(\widehat{\theta}_n-\theta_0)+o_p(1),
$$

where $\Delta_t(\theta)=\mathbb{E}\left[g_t(X,\theta)g_t^\top(X,\theta)\right]$.

Proof of Lemma A.2: We can rewrite $A_{n,t}$ as

$$
\begin{aligned}
A_{n,t}=&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(t;\theta_0)g_t(X_i,\theta_0)\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^n(e_i(t;\widehat{\theta}_n)-e_i(t;\theta_0))g_t(X_i,\theta_0)\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(t;\theta_0)(g_t(X_i,\widehat{\theta}_n)-g_t(X_i,\theta_0))\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^n(e_i(t;\widehat{\theta}_n)-e_i(t;\theta_0))(g_t(X_i,\widehat{\theta}_n)-g_t(X_i,\theta_0))\\
\equiv&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(t;\theta_0)g_t(X_i,\theta_0)+A_{n,t1}+A_{n,t2}+A_{n,t3}.
\end{aligned}
$$

We first show that $A_{n,t1}=-\sqrt{n}(\widehat{\theta}_n-\theta_0)^\top\Delta_t(\theta_0)+o_p(1)$. To this end, note that

$$
\begin{aligned}
A_{n,t1}=&-\frac{1}{\sqrt{n}}\sum_{i=1}^n(q_t(X_i,\widehat{\theta}_n)-q_t(X_i,\theta_0))g_t(X_i,\theta_0)\\
=&-\frac{1}{n}\sum_{i=1}^n g_t(X_i,\theta_0)\frac{\partial q_t(X_i,\widetilde{\theta}_n)}{\partial\theta^\top}\sqrt{n}(\widehat{\theta}_n-\theta_0)
\end{aligned}
$$

$$= - \mathbb{E}[g_t(X, \theta_0) g_t^\top(X, \theta_0)] \sqrt{n}(\widehat{\theta}_n - \theta_0) + o_p(1),$$

with $\widetilde{\theta}_n$ lying between $\widehat{\theta}_n$ and θ_0, where the second equality follows by the MVT, and the third equality follows from the ULLN of Newey and McFadden (1994, Lemma 2.4), and Assumptions 3.2 and 3.3.

It thus remains to show that both $A_{n,t2}$ and $A_{n,t3}$ are asymptotically negligible. To this end, note that

$$\begin{aligned}
A_{n,t2} =& \sqrt{n}(\widehat{\theta}_n - \theta_0)^\top \frac{1}{n} \sum_{i=1}^n e_i(t; \theta_0) \frac{\partial g_t(X_i, \widetilde{\theta}_n)}{\partial \theta} \\
=& \sqrt{n}(\widehat{\theta}_n - \theta_0)^\top \mathbb{E}\left[e(t; \theta_0) \frac{\partial g_t(X, \theta_0)}{\partial \theta} \right] + o_p(1) \\
=& o_p(1),
\end{aligned}$$

with $\widetilde{\theta}_n$ again lying between $\widehat{\theta}_n$ and θ_0, where the first equality follows by MVT, the second equality by ULLN of Newey and McFadden (1994, Lemma 2.4), and the last equality by Assumptions 3.2-3.3 as well as the law of iterated expectations (LIE) under the null hypothesis H_0.

For the last term $A_{n,t3}$, we have

$$\begin{aligned}
\sqrt{n} A_{n,t3} =& - \sum_{i=1}^n (q_t(X_i, \widehat{\theta}_n) - q_t(X_i, \theta_0))(g_t(X_i, \widehat{\theta}_n) - g_t(X_i, \theta_0)) \\
=& - \sqrt{n}(\widehat{\theta}_n - \theta_0)^\top \frac{1}{n} \sum_{i=1}^n \frac{\partial q_t(X_i, \widetilde{\theta}_n)}{\partial \theta} \frac{\partial g_t(X_i, \breve{\theta}_n)}{\partial \theta^\top} \sqrt{n}(\widehat{\theta}_n - \theta_0) \\
=& - \sqrt{n}(\widehat{\theta}_n - \theta_0)^\top \mathbb{E}\left[g_t(X, \theta_0) \frac{\partial g_t(X, \theta_0)}{\partial \theta^\top} \right] \sqrt{n}(\widehat{\theta}_n - \theta_0) + o_p(1) \\
=& O_p(1),
\end{aligned}$$

with $\widetilde{\theta}_n$ and $\breve{\theta}_n$ (potentially different) both lying between $\widehat{\theta}_n$ and θ_0, where the second equality follows by MVT, the third equality by ULLN of Newey and McFadden (1994, Lemma 2.4), and the last equality by Assumptions 3.2 and 3.3. Thus, $A_{n,t3} = O_p(n^{-1/2}) = o_p(1)$. This completes the proof of Lemma A.2. □

The next two lemmas establish the uniform convergence of $G_{n,t}(x; \widehat{\theta}_n)$ and $\Delta_{n,t}^{-1}(\widehat{\theta}_n)$ to $G_t(x; \theta_0)$ and $\Delta_t^{-1}(\theta_0)$ for each $t \in \mathcal{T}$, respectively.

Lemma A.3 *Suppose Assumptions 3.1-3.3 hold. Then, for each $t \in \mathcal{T}$, we have that*

$$\sup_{x \in \Pi_{pro}} \left| G_{n,t}(x; \widehat{\theta}_n) - G_t(x; \theta_0) \right| = o_p(1).$$

Proof of Lemma A.3: The proof follows directly from the ULLN of Newey and McFadden (1994, Lemma 2.4). □

Lemma A.4 *Suppose Assumptions 3.1-3.3 hold. Then, for each $t \in \mathcal{T}$, we have that*

$$\Delta_{n,t}^{-1}(\widehat{\theta}_n) = \Delta_t^{-1}(\theta_0) + o_p(1).$$

Proof of Lemma A.4: The proof follows from the ULLN of Newey and McFadden (1994, Lemma 2.4) and the continuous mapping theorem. □

With the assistance of Lemmas A.1-A.4, we are ready to proceed with the proofs of our main theorems.

Proof of Theorem 3.1: We first establish the uniform asymptotic representation of $R_{n,t}^{dpro}(x;\widehat{\theta}_n)$, which states that $R_{n,t}^{dpro}(x;\widehat{\theta}_n)$ is asymptotically invariant to $\widehat{\theta}_n$. Based on this representation, we can then readily prove the weak convergence of $R_{n,t}^{dpro}(x;\widehat{\theta}_n)$ to a centered Gaussian process with covariance structure given by $\mathbb{K}_t^{dpro}(x,x')$. Lastly, the limiting null distribution of CvM_n^{dpro} can be obtained by standard techniques.

By a straightforward decomposition, we have

$$\begin{aligned}
R_{n,t}^{dpro}(x;\widehat{\theta}_n) =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\widehat{\theta}_n)\left\{1\left(\beta^\top X_i \le u\right) - g_t(X_i,\widehat{\theta}_n)^\top \Delta_{n,t}^{-1}(\widehat{\theta}_n) G_{n,t}(x;\widehat{\theta}_n)\right\} \\
=& R_{n,t}^{pro}(x;\widehat{\theta}_n) - G_{n,t}(x;\widehat{\theta}_n)^\top \Delta_{n,t}^{-1}(\widehat{\theta}_n)\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\widehat{\theta}_n) g_t(X_i,\widehat{\theta}_n) \\
\equiv& R_{n,t}^{pro}(x;\widehat{\theta}_n) - G_{n,t}(x;\widehat{\theta}_n)^\top \Delta_{n,t}^{-1}(\widehat{\theta}_n) S_{n,t}.
\end{aligned}$$

Then, by Lemmas A.1-A.4,

$$\begin{aligned}
& R_{n,t}^{dpro}(x;\widehat{\theta}_n) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0) 1\left(\beta^\top X_i \le u\right) - G_t(x;\theta_0)^\top \sqrt{n}(\widehat{\theta}_n - \theta_0) \\
&- G_t(x;\theta_0)^\top \Delta_t^{-1}(\theta_0)\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0) g_t(X_i,\theta_0) - \Delta_t(\theta_0)\sqrt{n}(\widehat{\theta}_n - \theta_0)\right] + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)\left\{1\left(\beta^\top X_i \le u\right) - G_t(x;\theta_0)^\top \Delta_t^{-1}(\theta_0) g_t(X_i,\theta_0)\right\} + o_p(1) \\
&\equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)\mathcal{P}_t 1\left(\beta^\top X_i \le u\right) + o_p(1),
\end{aligned}$$

uniformly in $x \in \Pi_{pro}$.

The weak convergence of the infeasible double-projected empirical process

$$R_{n0,t}^{dpro}(x;\theta_0) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)\mathcal{P}_t 1\left(\beta^\top X_i \le u\right), \tag{A.1}$$

and consequently the weak convergence of $R_{n,t}^{dpro}(x;\widehat{\theta}_n)$ to the centered Gaussian process $R_{\infty,t}^{dpro}$ with covariance structure $\mathbb{K}_t^{dpro}(x,x')$ specified in (3.1) can be readily obtained by showing that the finite-dimensional distributions of $R^{dpro}_{n0,t}(x;\theta_0)$ converge to those of $R^{dpro}_{\infty,t}$ and the stochastic equicontinuity of $R^{dpro}_{n0,t}(x;\theta_0)$ by a straightforward application of the Donsker property of the class of linear indicator functions

$$\mathcal{F} = \left\{\nu \mapsto 1(\beta^\top \nu \leq u) : (\beta, u) \in \Pi_{pro}\right\}.$$

For the convergence in distribution of test statistic CvM^{dpro}_n, we will prove that

$$\sum_{t\in\mathcal{T}} a_n(t) \int_{\Pi_{pro}} \left(R^{dpro}_{n,t}(x;\widehat{\theta}_n)\right)^2 F_{n,\beta}(du)\, d\beta \xrightarrow{d} \sum_{t\in\mathcal{T}} a(t) \int_{\Pi_{pro}} \left(R^{dpro}_{\infty,t}(x)\right)^2 F_\beta(du)\, d\beta.$$

Given the assumption that $a_n(t) \xrightarrow{p} a(t)$ for each $t \in \mathcal{T}$, according to Slutsky's theorem, it suffices to show that, for each $t \in \mathcal{T}$,

$$\int_{\Pi_{pro}} \left(R^{dpro}_{n,t}(x;\widehat{\theta}_n)\right)^2 F_{n,\beta}(du)\, d\beta \xrightarrow{d} \int_{\Pi_{pro}} \left(R^{dpro}_{\infty,t}(x)\right)^2 F_\beta(du)\, d\beta.$$

First of all, note that the weak convergence of the double-projected process $R^{dpro}_{n,t}(x;\widehat{\theta}_n)$ and the Skorohod construction [see, e.g., Serfling, 1980] yield

$$\sup_{(\beta,u)\in\Pi_{pro}} \left|R^{dpro}_{n,t}(x;\widehat{\theta}_n) - R^{dpro}_{\infty,t}(x)\right| \xrightarrow{a.s.} 0. \tag{A.2}$$

Note that the empirical distribution function $F_{n,\beta}(u) \equiv n^{-1}\sum_{i=1}^n 1\left(\beta^\top X_i \leq u\right)$ estimates CDF $F_\beta(u) := \mathbb{P}\left(\beta^\top X \leq u\right)$ $a.s.$ uniformly for $(\beta, u) \in \Pi_{pro}$ by invoking the ULLN of Jennrich (1969) or the generalization by Wolfowitz (1954) of the Glivenko–Cantelli theorem. That is,

$$\sup_{(\beta,u)\in\Pi_{pro}} |F_{n,\beta}(u) - F_\beta(u)| \xrightarrow{a.s.} 0. \tag{A.3}$$

Now write

$$\begin{aligned}
&\left|\int_{\Pi_{pro}} \left(R^{dpro}_{n,t}(x;\widehat{\theta}_n)\right)^2 F_{n,\beta}(du)\, d\beta - \int_{\Pi_{pro}} \left(R^{dpro}_{\infty,t}(x)\right)^2 F_\beta(du)\, d\beta\right| \\
\leq &\left|\int_{\Pi_{pro}} \left[\left(R^{dpro}_{n,t}(x;\widehat{\theta}_n)\right)^2 - \left(R^{dpro}_{\infty,t}(x)\right)^2\right] F_{n,\beta}(du)\, d\beta\right| \\
&+ \left|\int_{\Pi_{pro}} \left(R^{dpro}_{\infty,t}(x)\right)^2 [F_{n,\beta}(du) - F_\beta(du)]\, d\beta\right|.
\end{aligned}$$

The first term of the right-hand side of the above inequality is $o(1)$ $a.s.$ due to (A.2). The trajectories of the limiting process $R^{dpro}_{\infty,t}(x)$ are bounded and continuous $a.s.$. Then, by applying the Helly–Bray Theorem (see p. 97 in Rao, 1965) to each of these trajectories and taking into account (A.3), we have that the second term of the right-hand side of the above inequality is also $o(1)$ $a.s.$. This completes the proof of Theorem 3.1. □

Proof of Theorem 3.2: Under Assumptions 3.1-3.3, uniformly in $x \in \Pi_{pro}$,

$$\begin{aligned}
&\sup_{x\in\Pi_{pro}}\left|\frac{1}{n}\sum_{i=1}^{n}\left\{e_i(t;\widehat{\theta}_n)\mathcal{P}_{n,t}1\left(\beta^\top X_i \le u\right) - \mathbb{E}\left[e(t;\theta^*)\mathcal{P}1\left(\beta^\top X \le u\right)\right]\right\}\right| \\
=&\sup_{x\in\Pi_{pro}}\left|\frac{1}{\sqrt{n}}R_{n,t}^{dpro}(x;\widehat{\theta}_n) - \mathbb{E}\left[\left(p_t\left(X\right) - q_t\left(X,\theta^*\right)\right)\mathcal{P}1\left(\beta^\top X \le u\right)\right]\right| \\
=&o_p(1)
\end{aligned}$$

by ULLN of Newey and McFadden (1994, Lemma 2.4) and similar arguments as proving Lemmas A.1, A.3 and A.4. □

Proof of Theorem 3.3: Note that under the sequence of local alternatives $H_{1,n}$ in (3.2), we have that, uniformly in $x \in \Pi_{pro}$:

$$\begin{aligned}
&R_{n,t}^{dpro}(x;\widehat{\theta}_n) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(e_i(t;\widehat{\theta}_n) - \frac{r_t(X_i)}{\sqrt{n}}\right)\mathcal{P}_{n,t}1\left(\beta^\top X_i \le u\right) \\
&+\frac{1}{n}\sum_{i=1}^{n} r_t(X_i)\mathcal{P}_{n,t}1\left(\beta^\top X_i \le u\right) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(e_i(t;\theta_0) - \frac{r_t(X_i)}{\sqrt{n}}\right)\mathcal{P}_t 1\left(\beta^\top X_i \le u\right) \\
&+\mathbb{E}\left[r_t(X)\mathcal{P}_t 1\left(\beta^\top X \le u\right)\right] + o_p(1) \\
\equiv&R_{n1,t}^{dpro}(x;\theta_0) + \delta_t(x) + o_p(1) \\
\Rightarrow&R_{\infty,t}^{dpro} + \delta_t,
\end{aligned}$$

where the second equality follows by similar arguments as proving Theorem 3.1 and by the ULLN.

Note that since $e_i(t;\theta_0) - n^{-1/2}r_t(X_i)$ forms a zero mean and an independent and identically distributed summand in the framework of local alternatives, we can establish the weak convergence of $R_{n1,t}^{dpro}(x;\theta_0)$ to a centered Gaussian process by checking the finite-dimensional distributions of $R_{n1,t}^{dpro}(x;\theta_0)$ and its stochastic equicontinuity, just as we have established for $R_{n0,t}^{dpro}(x;\theta_0)$ defined in (A.1). This yields that

$$R_{n1,t}^{dpro}(x;\theta_0) \Rightarrow R_{\infty,t}^{dpro}.$$

The last step thus follows and we finish the proof of Theorem 3.3. □

Proof of Theorem 4.1: As in Theorem 3.1, we have the following straightforward decomposition:

$$R_{n,t}^{dpro,*}(x;\widehat{\theta}_n)$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\widehat{\theta}_n)\left(1\left(\beta^\top X_i\le u\right) - g_t^\top(X_i,\widehat{\theta}_n)\Delta_{n,t}^{-1}(\widehat{\theta}_n)G_{n,t}(x;\widehat{\theta}_n)\right)V_i$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\widehat{\theta}_n)1\left(\beta^\top X_i\le u\right)V_i - G_{n,t}^\top(x;\widehat{\theta}_n)\Delta_{n,t}^{-1}(\widehat{\theta}_n)\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\widehat{\theta}_n)g_t(X_i,\widehat{\theta}_n)V_i$$
$$\equiv R_{n,t}^{pro,*}(x;\widehat{\theta}_n) - G_{n,t}^\top(x;\widehat{\theta}_n)\Delta_{n,t}^{-1}(\widehat{\theta}_n)A_{n,t}^*.$$

Conditional on the original sample $\{(X_i^\top,T_i)^\top\}_{i=1}^n$, it follows from the stochastic equicontinuity argument and the consistency of $\widehat{\theta}_n$ to θ_0 that, uniformly in $x\in\Pi_{pro}$,

$$R_{n,t}^{pro,*}(x;\widehat{\theta}_n)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)1\left(\beta^\top X_i\le u\right)V_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(e_i(t;\widehat{\theta}_n)-e_i(t;\theta_0)\right)1\left(\beta^\top X_i\le u\right)V_i$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)1\left(\beta^\top X_i\le u\right)V_i + o_p(1),$$

Similarly, by the MVT, the consistency of $\widehat{\theta}_n$ to θ_0 and the properties of the sequence of multipliers $\{V_i\}_{i=1}^n$, we can show that

$$A_{n,t}^*$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)g_t(X_i,\theta_0)V_i$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)\left(g_t(X_i,\widehat{\theta}_n)-g_t(X_i,\theta_0)\right)V_i$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(e_i(t;\widehat{\theta}_n)-e_i(t;\theta_0)\right)g_t(X_i,\theta_0)V_i$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(e_i(t;\widehat{\theta}_n)-e_i(t;\theta_0)\right)\left(g_t(X_i,\widehat{\theta}_n)-g_t(X_i,\theta_0)\right)V_i$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)g_t(X_i,\theta_0)V_i + o_p(1).$$

Thus, by Lemmas A.3 and A.4, uniformly in $x\in\Pi_{pro}$,

$$R_{n,t}^{dpro,*}(x;\widehat{\theta}_n)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)\left\{1\left(\beta^\top X_i\le u\right) - G_t(x;\theta_0)^\top\Delta_t^{-1}(\theta_0)g_t(X_i,\theta_0)\right\}V_i + o_p(1)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i(t;\theta_0)\mathcal{P}_t 1\left(\beta^\top X_i\le u\right)V_i + o_p(1)$$
$$\equiv R_{n0,t}^{dpro,*}(x;\theta_0) + o_p(1),$$

leading to the multiplier bootstrapped version of $R^{dpro}_{n0,t}(x;\theta_0)$ defined in (A.1).

Using the properties of $\{V_i\}_{i=1}^n$, the rest of the proof then follows readily from the conditional multiplier central limit theorem applied to the (infeasible) multiplier bootstrapped double-projected process $R^{dpro,*}_{n0,t}(x;\theta_0)$ regardless of whether the null hypothesis H_0 holds or not (see van der Vaart and Wellner, 1996, Theorem 2.9.6, p. 182), and the continuous mapping theorem. □

Proof of Proposition 5.1: We first define an intermediate doubly-projected empirical process

$$\widetilde{M}^{dpro}_{n,t}(x;\widehat{\theta}_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{im}(t;\widehat{\theta}_n)\left(1(\widetilde{X}_{i,\theta_0}^\top\beta \le u) - g_t(X_i,\widehat{\theta}_n)^\top\Delta_{n,t}^{-1}(\widehat{\theta}_n)G^{im}_{n,t}(x;\widehat{\theta}_n)\right).$$

Following similar arguments in the proofs of Lemmas A.1-A.4, under H_0^{im1} in (5.2), we can readily show that,

$$\begin{aligned}(i)\quad &\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{im}(t;\widehat{\theta}_n)1(\widetilde{X}_{i,\theta_0}^\top\beta \le u)\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{im}(t;\theta_0)1(\widetilde{X}_{i,\theta_0}^\top\beta \le u) - \sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top \mathbb{E}\left[g_t(X,\theta_0)1(\widetilde{X}_{\theta_0}^\top\beta \le u)\right] + o_p(1)\\ \equiv&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{im}(t;\theta_0)1(\widetilde{X}_{i,\theta_0}^\top\beta \le u) - \sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top G^{im}_t(x;\theta_0) + o_p(1)\end{aligned}$$

uniformly in x and for each $t \in \mathcal{T}$,

$$\begin{aligned}(ii)\quad &\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{im}(t;\widehat{\theta}_n)g_t(X_i,\widehat{\theta}_n)\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{im}(t;\theta_0)g_t(X_i,\theta_0) - \Delta_t(\theta_0)\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right) + o_p(1)\end{aligned}$$

for each $t \in \mathcal{T}$,

$$(iii)\quad G^{im}_{n,t}(x;\widehat{\theta}_n) = G^{im}_t(x;\theta_0) + o_p(1)$$

uniformly in x and for each $t \in \mathcal{T}$, and

$$(iv)\quad \Delta_{n,t}^{-1}(\widehat{\theta}_n) = \Delta_t^{-1}(\theta_0) + o_p(1)$$

for each $t \in \mathcal{T}$. As an immediate consequence of $(i)-(iv)$, uniformly in x,

$$\begin{aligned}&\widetilde{M}^{dpro}_{n,t}(x;\widehat{\theta}_n)\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{im}(t;\theta_0)1(\widetilde{X}_{i,\theta_0}^\top\beta \le u) - G^{im}_t(x;\theta_0)^\top\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)\end{aligned}$$

$$
\begin{aligned}
&- G_t^{im}(x;\theta_0)^\top \Delta_t^{-1}(\theta_0)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{im}(t;\theta_0) g_t(X_i,\theta_0) - \Delta_t(\theta_0)\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)\right) + o_p(1) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{im}(t;\theta_0)\left(1(\widetilde{X}_{i,\theta_0}^\top \beta \le u) - g_t(X_i,\theta_0)^\top \Delta_t^{-1}(\theta_0) G_t^{im}(x;\theta_0)\right) + o_p(1) \\
\equiv&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{im}(t;\theta_0)\mathcal{P}_t^{im} 1(\widetilde{X}_{i,\theta_0}^\top \beta \le u) + o_p(1).
\end{aligned}
$$

In light of the above result, it then suffices to show that $M_{n,t}^{dpro}(x;\widehat{\theta}_n)$ is asymptotically uniformly equivalent to $\widetilde{M}_{n,t}^{dpro}(x;\widehat{\theta}_n)$. To this end, note that uniformly in x,

$$
\begin{aligned}
&M_{n,t}^{dpro}(x;\widehat{\theta}_n) - \widetilde{M}_{n,t}^{dpro}(x;\widehat{\theta}_n) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{im}(t;\widehat{\theta}_n)\left(1(\widetilde{X}_{i,\widehat{\theta}_n}^\top \beta \le u) - 1(\widetilde{X}_{i,\theta_0}^\top \beta \le u)\right) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(e_i^{im}(t;\theta_0) - \left(q_t(\widetilde{X}_{i,\widehat{\theta}_n}) - q_t(\widetilde{X}_{i,\theta_0})\right)\right)\left(1(\widetilde{X}_{i,\widehat{\theta}_n}^\top \beta \le u) - 1(\widetilde{X}_{i,\theta_0}^\top \beta \le u)\right) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i(t)\left(1(\widetilde{X}_{i,\widehat{\theta}_n}^\top \beta \le u) - 1(\widetilde{X}_{i,\theta_0}^\top \beta \le u)\right) \\
&- \sqrt{n}(\widehat{\theta}_n - \theta_0)^\top \frac{1}{n}\sum_{i=1}^n g_t(X_i,\theta_0)\left(1(\widetilde{X}_{i,\widehat{\theta}_n}^\top \beta \le u) - 1(\widetilde{X}_{i,\theta_0}^\top \beta \le u)\right) + o_p(1) \\
\equiv&M_{n,t1}^{dpro}(x) - \sqrt{n}(\widehat{\theta}_n - \theta_0)^\top M_{n,t2}^{dpro}(x) + o_p(1),
\end{aligned}
$$

where $e_i^{im}(t;\theta_0) = \varepsilon_i(t)$ *a.s.* under H_0^{im1} in (5.2), and the second to last equality follows by the Taylor expansion of $q_t(\widetilde{X}_{i,\widehat{\theta}_n})$ around $q_t(\widetilde{X}_{i,\theta_0})$ and Assumptions 3.2 and 3.3. Since $\sqrt{n}(\widehat{\theta}_n - \theta_0) = O_p(1)$ by Assumption 3.3(ii), it then remains to show that both $M_{n,t1}^{dpro}(x)$ and $M_{n,t2}^{dpro}(x)$ are asymptotically uniformly negligible in x.

For the first term $M_{n,t1}^{dpro}(x)$, introduce an auxiliary process

$$
\alpha_{n,t}(x;\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i(t) 1(\widetilde{X}_{i,\theta}^\top \beta \le u).
$$

Since under H_0^{im1} for each $t \in \mathcal{T}$ the sequence of error terms $\{\varepsilon_1(t), \ldots, \varepsilon_n(t)\}$ is centered conditionally on $\{X_1, \ldots, X_n\}$, $\alpha_{n,t}(x;\theta)$ has i.i.d. centered summands. Then $M_{n,t1}^{dpro}(x)$ can be expressed as $\alpha_{n,t}(x;\widehat{\theta}_n) - \alpha_{n,t}(x;\theta_0)$. Clearly, the class of linear indicator functions is Donsker and $\alpha_{n,t}(\cdot,\cdot)$ is asymptotically equicontinuous. Since $\widehat{\theta}_n \to_p \theta_0$ under H_0^{im1} by Assumption 3.3(ii), $M_{n,t1}^{dpro}(x) = o_p(1)$ uniformly in x. The claim that $M_{n,t2}^{dpro}(x) = o_p(1)$ uniformly in x follows straightforwardly from the uniform convergence of

$$
\frac{1}{n}\sum_{i=1}^n g_t(X_i,\theta_0) 1(\widetilde{X}_{i,\theta}^\top \beta \le u)
$$

in x and θ together with the continuity of its limit.

The rest of the proof is similar to that of Theorem 3.1 and is thus omitted to avoid repetition. $\square$

Proof of (5.13)**:** First of all, recall $e_i^{sim}(t;\theta,\mu_t) \equiv 1(T_i = t) - \mu_t(X_i^\top \theta)$ and the following decomposition:

$$
\begin{aligned}
&S_{n,t}^{pro}(x;\widehat{\theta}_n,\widehat{\mu}_t) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(1(T_i=t)-\widehat{\mu}_t(X_i^\top\widehat{\theta}_n)\right)\widehat{f}(X_i^\top\widehat{\theta}_n)1(X_i^\top\beta\le u) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(1(T_i=t)-\mu_t(X_i^\top\widehat{\theta}_n)\right)\widehat{f}(X_i^\top\widehat{\theta}_n)1(X_i^\top\beta\le u) \\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\widehat{\mu}_t(X_i^\top\widehat{\theta}_n)-\mu_t(X_i^\top\widehat{\theta}_n)\right)\widehat{f}(X_i^\top\widehat{\theta}_n)1(X_i^\top\beta\le u) \\
\equiv&S_{n,t1}^{pro}(x;\widehat{\theta}_n,\mu_t)-S_{n,t2}^{pro}(x;\widehat{\theta}_n,\widehat{\mu}_t).
\end{aligned}
$$

We shall study the components $S_{n,t1}^{pro}(x;\widehat{\theta}_n,\mu_t)$ and $S_{n,t2}^{pro}(x;\widehat{\theta}_n,\widehat{\mu}_t)$ separately. In particular, we shall show that

$$
\begin{aligned}
&S_{n,t1}^{pro}(x;\widehat{\theta}_n,\mu_t) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{sim}(t;\theta_0,\mu_t)f(X_i^\top\theta_0)1(X_i^\top\beta\le u) \\
&-\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top \mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)X1(X^\top\beta\le u)\right]+o_p(1),
\end{aligned}
\tag{A.4}
$$

and

$$
\begin{aligned}
&S_{n,t2}^{pro}(x;\widehat{\theta}_n,\widehat{\mu}_t) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{sim}(t;\theta_0,\mu_t)f(X_i^\top\theta_0)\mathbb{E}[1(X_i^\top\beta\le u)|X_i^\top\theta_0] \\
&-\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top \mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)\mathbb{E}[X|X^\top\theta_0]1(X^\top\beta\le u)\right]+o_p(1).
\end{aligned}
\tag{A.5}
$$

Henceforth, for notational simplicity, we omit μ_t and $\widehat{\mu}_t$ in the subsequent stochastic processes. For example, we will use $S_{n,t1}^{pro}(x;\widehat{\theta}_n)$ instead of $S_{n,t1}^{pro}(x;\widehat{\theta}_n,\mu_t)$, and use $S_{n,t2}^{pro}(x;\widehat{\theta}_n)$ instead of $S_{n,t2}^{pro}(x;\widehat{\theta}_n,\widehat{\mu}_t)$.

Proof of (A.4)**:** It is clear that $S_{n,t1}^{pro}(x;\widehat{\theta}_n)$ has the following decomposition:

$$
\begin{aligned}
S_{n,t1}^{pro}(x;\widehat{\theta}_n) =&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{sim}(t;\widehat{\theta}_n)\widehat{f}(X_i^\top\widehat{\theta}_n)1(X_i^\top\beta\le u) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{sim}(t;\theta_0)\widehat{f}(X_i^\top\widehat{\theta}_n)1(X_i^\top\beta\le u)
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\mu_t(X_i^\top \widehat{\theta}_n) - \mu_t(X_i^\top \theta_0) \right) \widehat{f}(X_i^\top \widehat{\theta}_n) 1(X_i^\top \beta \le u) \\
\equiv & S_{n,t11}^{pro}(x; \widehat{\theta}_n) - S_{n,t12}^{pro}(x; \widehat{\theta}_n).
\end{aligned}
$$

In the following we shall study the asymptotic properties of $S_{n,t11}^{pro}(x; \widehat{\theta}_n)$ and $S_{n,t12}^{pro}(x; \widehat{\theta}_n)$ separately.

Part I: Note that $S_{n,t11}^{pro}(x; \widehat{\theta}_n)$ can be rewritten as

$$
S_{n,t11}^{pro}(x; \widehat{\theta}_n) = \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j \ne i}^{n} \frac{1}{h} K \left(\frac{(X_i - X_j)^\top \widehat{\theta}_n}{h} \right) e_i^{sim}(t; \theta_0) 1(X_i^\top \beta \le u),
$$

which is an interesting variant of U-process considered by Stute (1994), with its kernel varying with sample size n via $\widehat{\theta}_n$ and the bandwidth h. By the M-th order Taylor expansion of kernel function $K(v)$, we have

$$
\begin{aligned}
& S_{n,t11}^{pro}(x; \widehat{\theta}_n) \\
= & \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j \ne i}^{n} \frac{1}{h} K \left(\frac{(X_i - X_j)^\top \theta_0}{h} \right) e_i^{sim}(t; \theta_0) 1(X_i^\top \beta \le u) \\
& + \sum_{s=1}^{M-1} \frac{1}{s!} \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j \ne i}^{n} \frac{1}{h^{s+1}} K^{(s)} \left(\frac{(X_i - X_j)^\top \theta_0}{h} \right) \left[(X_i - X_j)^\top (\widehat{\theta}_n - \theta_0) \right]^s \\
& e_i^{sim}(t; \theta_0) 1(X_i^\top \beta \le u) \\
& + \frac{1}{M!} \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j \ne i}^{n} \frac{1}{h^{M+1}} K^{(M)} \left(\frac{(X_i - X_j)^\top \widetilde{\theta}_n}{h} \right) \left[(X_i - X_j)^\top (\widehat{\theta}_n - \theta_0) \right]^M \\
& e_i^{sim}(t; \theta_0) 1(X_i^\top \beta \le u) \\
\equiv & S_{n,t110}^{pro}(x; \theta_0) + \sum_{s=1}^{M-1} \frac{1}{s!} S_{n,t11s}^{pro}(x; \widehat{\theta}_n) + \frac{1}{M!} S_{n,t11M}^{pro}(x; \widehat{\theta}_n),
\end{aligned}
$$

where $K^{(s)}(v) \equiv d^s K(v)/dv^s$ denotes the s-th derivative of $K(v)$ for $s = 1, \ldots, M$ with $M \ge 5$, which is assumed to be uniformly bounded from above. Here and henceforth, $\widetilde{\theta}_n$ lies between θ_0 and $\widehat{\theta}_n$ and satisfies $||\widetilde{\theta}_n - \theta_0|| \le ||\widehat{\theta}_n - \theta_0||$.

First, we analyze $S_{n,t110}^{pro}(x; \theta_0)$. Hereafter, $\chi_i = (X_i^\top, T_i)^\top$. We denote by

$$
\begin{aligned}
\Phi_{t110,x}(\chi_i, \chi_j) = & \frac{1}{2} \left[\frac{1}{h} K \left(\frac{(X_i - X_j)^\top \theta_0}{h} \right) e_i^{sim}(t; \theta_0) 1(X_i^\top \beta \le u) \right. \\
& \left. + \frac{1}{h} K \left(\frac{(X_i - X_j)^\top \theta_0}{h} \right) e_j^{sim}(t; \theta_0) 1(X_j^\top \beta \le u) \right].
\end{aligned}
$$

Note that we can express $S^{pro}_{n,t110}(x;\theta_0)$ as a second order U-process of the following form

$$S^{pro}_{n,t110}(x;\theta_0) = \frac{2}{\sqrt{n}(n-1)} \sum_{1\le i<j\le n}^{n} \Phi_{t110,x}(\chi_i,\chi_j).$$

Obviously, under the null hypothesis H_0^{sim}, by LIE,

$$\int \Phi_{t110,x}(\chi_i,\chi_j)dP(\chi_i)dP(\chi_j) = 0,$$

where $P(\chi_i)$ denotes the distribution of χ_i. Then by the standard Hoeffding decomposition of U-process, we have

$$S^{pro}_{n,t110}(x;\theta_0) = S^{pro(1)}_{n,t110}(x;\theta_0) + S^{pro(2)}_{n,t110}(x;\theta_0),$$

where

$$S^{pro(1)}_{n,t110}(x;\theta_0) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \int \Phi_{t110,x}(\chi_i,\chi_j)dP(\chi_j),$$

$$\begin{aligned} S^{pro(2)}_{n,t110}(x;\theta_0) =& \frac{2}{\sqrt{n}(n-1)} \sum_{1\le i<j\le n}^{n} \left[\Phi_{t110,x}(\chi_i,\chi_j) - \int \Phi_{t110,x}(\chi_i,\chi_j)dP(\chi_i) \right. \\ & \left. - \int \Phi_{t110,x}(\chi_i,\chi_j)dP(\chi_j)\right]. \end{aligned}$$

Straightforward calculations yield that

$$\int \Phi_{t110,x}(\chi_i,\chi_j)dP(\chi_j) = \frac{1}{2} e_i^{sim}(t;\theta_0) f(X_i^\top\theta_0) 1(X_i^\top\beta \le u) + O_p(h^k),$$

where we have used $\int K(v)dv = 1$, $\int v^j K(v)dv = 0$ for $j = 1,\ldots,k-1$ and $\int v^k K(v)dv \neq 0$ in Assumption 5.2. On the other hand, the remainder term $S^{pro(2)}_{n,t110}(x;\theta_0)$ is a degenerate U-process of second order. We note that $\max_{1\le i\le n} |e_i^{sim}(t;\theta_0)| \le 2$. Then by applying Proposition 4 of Delgado and González Manteiga (2001), there exists an absolute constant C, which does not depend on h, such that

$$\begin{aligned} \mathbb{E}\left[\sup_x \left|S^{pro(2)}_{n,t110}(x;\theta_0)\right|^2\right] \le& C\frac{1}{n}\mathbb{E}\left[\Phi_{t110,x}(\chi_i,\chi_j)^2\right] \\ \le& C\frac{1}{nh^2}\mathbb{E}\left[K^2\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)\right] \\ =& O\left(\frac{1}{nh}\right). \end{aligned}$$

This implies that $\sup_x |S^{pro(2)}_{n,t110}(x;\theta_0)| = O_p((nh)^{-1/2}) = o_p(1)$ under $nh^3 \to \infty$ in Assumption 6. Thus, uniformly in x,

$$S^{pro}_{n,t110}(x;\theta_0) = S^{2pro(1)}_{n,t110}(x;\theta_0) + o_p(1)$$

$$
\begin{aligned}
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i^{sim}(t;\theta_0)f(X_i^\top\theta_0)1(X_i^\top\beta\le u)+O_p(\sqrt{n}h^k)+o_p(1)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i^{sim}(t;\theta_0)f(X_i^\top\theta_0)1(X_i^\top\beta\le u)+o_p(1),
\end{aligned}
$$

where the last step follows by $nh^{2k}\to 0$ in Assumption 5.3.

Next, we prove that $S_{n,t11s}^{pro}(x;\widehat{\theta}_n)=o_p(1)$ uniformly in x for $s=1,\ldots,M$. Note that

$$
\begin{aligned}
&S_{n,t111}^{pro}(x;\widehat{\theta}_n)\\
=&(\widehat{\theta}_n-\theta_0)^\top\frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^{n}\sum_{j\ne i}^{n}\frac{1}{h^2}K^{(1)}\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)(X_i-X_j)e_i^{sim}(t;\theta_0)1(X_i^\top\beta\le u)\\
\equiv&(\widehat{\theta}_n-\theta_0)^\top S_{n,t1111}^{pro}(x;\theta_0).
\end{aligned}
$$

We can rewrite $S_{n,t1111}^{pro}(x;\theta_0)$ as the following second order U-process

$$
S_{n,t1111}^{pro}(x;\theta_0)=\frac{2}{\sqrt{n}(n-1)}\sum_{1\le i<j\le n}^{n}\Phi_{t1111,x}(\chi_i,\chi_j),
$$

where

$$
\begin{aligned}
\Phi_{t1111,x}(\chi_i,\chi_j)=&\frac{1}{2}\left[\frac{1}{h^2}K^{(1)}\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)(X_i-X_j)e_i^{sim}(t;\theta_0)1(X_i^\top\beta\le u)\right.\\
&\left.-\frac{1}{h^2}K^{(1)}\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)(X_j-X_i)e_j^{sim}(t;\theta_0)1(X_j^\top\beta\le u)\right].
\end{aligned}
$$

Note that we have used the antisymmetry property $K^{(1)}(-v)=-K^{(1)}(v)$ in the above expression of kernel function $\Phi_{t1111,x}(\chi_i,\chi_j)$. Clearly, under the null H_0^{sim},

$$
\int\Phi_{t1111,x}(\chi_i,\chi_j)dP(\chi_i)dP(\chi_j)=0.
$$

As before, we can decompose $S_{n,t1111}^{2pro}(x;\theta_0)$ by

$$
S_{n,t1111}^{pro}(x;\theta_0)=S_{n,t1111}^{pro(1)}(x;\theta_0)+S_{n,t1111}^{pro(2)}(x;\theta_0),
$$

where

$$
S_{n,t1111}^{pro(1)}(x;\theta_0)=\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\int\Phi_{t1111,x}(\chi_i,\chi_j)dP(\chi_j),
$$

$$
\begin{aligned}
S_{n,t1111}^{pro(2)}(x;\theta_0)=&\frac{2}{\sqrt{n}(n-1)}\sum_{1\le i<j\le n}^{n}\left[\Phi_{t1111,x}(\chi_i,\chi_j)-\int\Phi_{t1111,x}(\chi_i,\chi_j)dP(\chi_i)\right.\\
&\left.-\int\Phi_{t1111,x}(\chi_i,\chi_j)dP(\chi_j)\right].
\end{aligned}
$$

Recall that $\varphi(v) = \mathbb{E}[X|X^\top\theta_0 = v]$, which is assumed to be k-times continuously differentiable with respect to v. Then we obtain that

$$\begin{aligned}
&\int \Phi_{t1111,x}(\chi_i,\chi_j)dP(\chi_j)\\
=&\frac{1}{2}e_i^{sim}(t;\theta_0)\left[\left(X_i-\varphi(X_i^\top\theta_0)\right)f'(X_i^\top\theta_0)-\varphi'(X_i^\top\theta_0)f(X_i^\top\theta_0)\right]1(X_i^\top\beta\le u)+O_p(h^k),
\end{aligned}$$

where we have used the properties that $\int K^{(1)}(v)dv = 0$, $\int vK^{(1)}(v)dv = -1$ and $\int v^jK^{(1)}(v)dv = 0$ for $j = 2,\ldots,k$. In addition, $S^{pro(2)}_{n,t1111}(x;\theta_0)$ is a degenerate U-process of second order. By applying Proposition 4 of Delgado and González Manteiga (2001), there exists a constant C, which does not depend on h, such that

$$\begin{aligned}
\mathbb{E}\left[\sup_x\left|S^{pro(2)}_{n,t1111}(x;\theta_0)\right|^2\right]\le&C\frac{1}{n}\mathbb{E}\left[\Phi_{t1111,x}(\chi_i,\chi_j)^2\right]\\
\le&C\frac{1}{nh^4}\mathbb{E}\left[K^{(1)}\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)^2\right]\\
=&O\left(\frac{1}{nh^3}\right),
\end{aligned}$$

implying that $\sup_x|S^{pro(2)}_{n,t1111}(x;\theta_0)| = O_p((nh^3)^{-1/2}) = o_p(1)$ under $nh^3 \to \infty$. Consequently, under $nh^{2k} \to 0$, uniformly in x,

$$\begin{aligned}
&S^{pro}_{n,t1111}(x;\theta_0)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i^{sim}(t;\theta_0)\left[\left(X_i-\varphi(X_i^\top\theta_0)\right)f'(X_i^\top\theta_0)-\varphi'(X_i^\top\theta_0)f(X_i^\top\theta_0)\right]1(X_i^\top\beta\le u)+o_p(1)\\
=&O_p(1),
\end{aligned}$$

where the second equality follows by the weak convergence of the single i.i.d. summation in the first equality. As a result, we have, uniformly in x,

$$S^{pro}_{n,t111}(x;\widehat{\theta}_n) = O_p(n^{-1/2})O_p(1) = o_p(1),$$

due to the assumption $||\widehat{\theta}-\theta_0|| = O_p(n^{-1/2})$.

For $S^{pro}_{n,t112}(x;\widehat{\theta}_n)$, note that it can be expresses as

$$\begin{aligned}
&S^{pro}_{n,t112}(x;\widehat{\theta}_n)\\
=&\frac{1}{2h^2}(\widehat{\theta}_n-\theta_0)^\top\frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^n\sum_{j\ne i}^n\frac{1}{h}K^{(2)}\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)(X_i-X_j)(X_i-X_j)^\top\\
&e_i^{sim}(t;\theta_0)1(X_i^\top\beta\le u)(\widehat{\theta}_n-\theta_0)\\
\equiv&\frac{1}{2h^2}(\widehat{\theta}_n-\theta_0)^\top S^{pro}_{n,t1121}(x;\theta_0)(\widehat{\theta}_n-\theta_0).
\end{aligned}$$

As before, we can rewrite $S^{pro}_{n,t1121}(x;\theta_0)$ as the following second order U-process

$$S^{pro}_{n,t1121}(x;\theta_0) = \frac{2}{\sqrt{n}(n-1)} \sum_{1\le i<j\le n}^{n} \Phi_{t1121,x}(\chi_i,\chi_j),$$

where, by noting that $K^{(2)}(-v) = K^{(2)}(v)$,

$$\begin{aligned}\Phi_{t1121,x}(\chi_i,\chi_j) =& \frac{1}{2}\left[\frac{1}{h}K^{(2)}\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)(X_i-X_j)e_i^{sim}(t;\theta_0)1(X_i^\top\beta\le u)\right.\\ &\left.+\frac{1}{h}K^{(2)}\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)(X_j-X_i)e_j^{sim}(t;\theta_0)1(X_j^\top\beta\le u)\right].\end{aligned}$$

Under the null hypothesis H_0^{sim2},

$$\int \Phi_{t1121,x}(\chi_i,\chi_j)dP(\chi_i)dP(\chi_j) = 0.$$

We decompose $S^{pro}_{n,t1121}(x;\theta_0)$ by

$$S^{pro}_{n,t1121}(x;\theta_0) = S^{pro(1)}_{n,t1121}(x;\theta_0) + S^{pro(2)}_{n,t1121}(x;\theta_0),$$

where

$$S^{pro(1)}_{n,t1121}(x;\theta_0) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\int \Phi_{t1121,x}(\chi_i,\chi_j)dP(\chi_j),$$

$$\begin{aligned}S^{pro(2)}_{n,t1121}(x;\theta_0) =& \frac{2}{\sqrt{n}(n-1)}\sum_{1\le i<j\le n}^{n}\left[\Phi_{t1121,x}(\chi_i,\chi_j) - \int \Phi_{t1121,x}(\chi_i,\chi_j)dP(\chi_i)\right.\\ &\left.-\int \Phi_{t1121,x}(\chi_i,\chi_j)dP(\chi_j)\right].\end{aligned}$$

It can be shown that

$$\int \Phi_{t1121,x}(\chi_i,\chi_j)dP(\chi_j) = O_p(h^k),$$

where we have used the properties that $\int K^{(2)}(v)dv = 0$ and $\int v^j K^{(2)}(v)dv = 0$ for $j = 1,\ldots,k$. Again, $S^{pro(2)}_{n,t1121}(x;\theta_0)$ is a degenerate U-process of second order, and by applying Proposition 4 of Delgado and González Manteiga (2001), there exists a constant C, which does not depend on h, such that

$$\begin{aligned}\mathbb{E}\left[\sup_x\left|S^{pro(2)}_{n,t1121}(x;\theta_0)\right|^2\right] \le& C\frac{1}{n}\mathbb{E}\left[\Phi_{t1121,x}(\chi_i,\chi_j)^2\right]\\ \le& C\frac{1}{nh^2}\mathbb{E}\left[K^{(2)}\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)^2\right]\\ =& O\left(\frac{1}{nh}\right).\end{aligned}$$

This implies that $\sup_x |S^{pro(2)}_{n,t1121}(x;\theta_0)| = O_p((nh)^{-1/2}) = o_p(1)$ under $nh^3 \to \infty$. Therefore, uniformly in x,

$$S^{pro}_{n,t1121}(x;\theta_0) = O_p(\sqrt{n}h^k) + o_p(1) = o_p(1),$$

under $nh^{2k} \to 0$. Consequently, uniformly in x,

$$S^{pro}_{n,t112}(x;\widehat{\theta}_n) = O_p((nh^2)^{-1})o_p(1) = o_p(1).$$

Then following similar arguments as showing $S^{pro}_{n,t112}(x;\widehat{\theta}_n) = o_p(1)$ uniformly in x, we can prove that $S^{pro}_{n,t11s}(x;\widehat{\theta}_n) = O_p((nh^2)^{-s/2})o_p(1) = o_p(1)$ uniformly in x for $s = 3,\ldots,M-1$.

Finally, noting that $\max_{1\le i\le n} |e^{sim}_i(t;\theta_0)| \le 2$, it is clear to see

$$\begin{aligned}
&\sup_x \left|S^{pro}_{n,t11M}(x;\widehat{\theta}_n)\right| \\
\le& \frac{2}{\sqrt{n}(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{1}{h^{M+1}} \left|K^{(M)}\left(\frac{(X_i-X_j)^\top \widetilde{\theta}_n}{h}\right)\right| \|X_i - X_j\|^M \left\|\widehat{\theta}_n - \theta_0\right\|^M \\
=& \frac{\sqrt{n}}{h^{M+1}} O_p(1) O_p(n^{-M/2}) = O_p((n^{M-1}h^{2(M+1)})^{-1/2}) = o_p(1),
\end{aligned}$$

under the assumptions that $\sup_v |K^{(M)}(v)| < \infty$, $\mathbb{E}||X||^M < \infty$ and $n^{M-1}h^{2(M+1)} \to \infty$, which is implied by $nh^3 \to \infty$ when $M \ge 5$.

Combining the above results obtained for $S^{pro}_{n,t110}(x;\theta_0)$ and $S^{pro}_{n,t11s}(x;\widehat{\theta}_n)$ for $s = 1,\ldots,M$, we have, uniformly in x,

$$S^{pro}_{n,t11}(x;\widehat{\theta}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n e^{sim}_i(t;\theta_0) f(X_i^\top\theta_0) 1(X_i^\top\beta \le u) + o_p(1).$$

Part II: Note that $S^{pro}_{n,t12}(x;\widehat{\theta}_n)$ can be rewritten as

$$\begin{aligned}
&S^{pro}_{n,t12}(x;\widehat{\theta}_n) \\
=& \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{1}{h} K\left(\frac{(X_i-X_j)^\top \widehat{\theta}_n}{h}\right) \left(\mu_t(X_i^\top\widehat{\theta}_n) - \mu_t(X_i^\top\theta_0)\right) 1(X_i^\top\beta \le u) \\
=& \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{1}{h} K\left(\frac{(X_i-X_j)^\top \theta_0}{h}\right) \left(\mu_t(X_i^\top\widehat{\theta}_n) - \mu_t(X_i^\top\theta_0)\right) 1(X_i^\top\beta \le u) \\
&+ \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{1}{h} \left[K\left(\frac{(X_i-X_j)^\top \widehat{\theta}_n}{h}\right) - K\left(\frac{(X_i-X_j)^\top \theta_0}{h}\right)\right] \\
&\left(\mu_t(X_i^\top\widehat{\theta}_n) - \mu_t(X_i^\top\theta_0)\right) 1(X_i^\top\beta \le u) \\
\equiv& S^{pro}_{n,t121}(x;\widehat{\theta}_n) + S^{pro}_{n,t122}(x;\widehat{\theta}_n).
\end{aligned}$$

First of all, by Assumption 5.1 and the second order Taylor expansion of $\mu_t(v)$,

$$
\begin{aligned}
&S^{pro}_{n,t121}(x;\widehat{\theta}_n)\\
=&\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}^n \frac{1}{h}K\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)\mu_t'(X_i^\top\theta_0)X_i1(X_i^\top\beta\le u)\\
&+\frac{1}{2}\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}^n \frac{1}{h}K\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)\mu_t''(X_i^\top\widetilde{\theta}_n)X_iX_i^\top\\
&1(X_i^\top\beta\le u)\left(\widehat{\theta}_n-\theta_0\right)\\
\equiv&\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top S^{pro}_{n,t1211}(x;\theta_0)+\frac{1}{2}\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top S^{pro}_{n,t1212}(x;\widehat{\theta}_n)\left(\widehat{\theta}_n-\theta_0\right).
\end{aligned}
$$

Note that $S^{pro}_{n,t1211}(x;\theta_0)$ can be written as the following second order U-process

$$
S^{pro}_{n,t1211}(x;\theta_0)=\frac{2}{n(n-1)}\sum_{1\le i<j\le n}^n \Phi_{t1211,x}(X_i,X_j),
$$

where

$$
\begin{aligned}
\Phi_{t1211,x}(X_i,X_j)=&\frac{1}{2}\left[\frac{1}{h}K\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)\mu_t'(X_i^\top\theta_0)X_i1(X_i^\top\beta\le u)\right.\\
&\left.+\frac{1}{h}K\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)\mu_t'(X_j^\top\theta_0)X_j1(X_j^\top\beta\le u)\right].
\end{aligned}
$$

Henceforth, $P(X_i)$ denotes the distribution of X_i. With a slight abuse of notation μ_t and its derivatives, we denote by

$$
\mu_1\equiv\int\Phi_{t1211,x}(X_i,X_j)dP(X_i)dP(X_j),
$$

which is not necessarily zero. We can then decompose $S^{pro}_{n,t1211}(x;\theta_0)$ by

$$
S^{pro}_{n,t1211}(x;\theta_0)=\mu_1+S^{pro(1)}_{n,t1211}(x;\theta_0)+S^{pro(2)}_{n,t1211}(x;\theta_0),
$$

where

$$
S^{pro(1)}_{n,t1211}(x;\theta_0)=\frac{2}{n}\sum_{i=1}^n\left[\int\Phi_{t1211,x}(X_i,X_j)dP(X_j)-\mu_1\right],
$$

$$
\begin{aligned}
S^{pro(2)}_{n,t1211}(x;\theta_0)=&\frac{2}{n(n-1)}\sum_{1\le i<j\le n}^n\left[\Phi_{t1211,x}(X_i,X_j)-\int\Phi_{t1211,x}(X_i,X_j)dP(X_i)\right.\\
&\left.-\int\Phi_{t1211,x}(X_i,X_j)dP(X_j)-\mu_1\right].
\end{aligned}
$$

Let $\eta_x(v)=\mathbb{E}[X1(X^\top\beta\le u)|X^\top\theta_0=v]$, which is assumed to be k-times continuously differentiable with respect to v. Straightforward calculations then yield that

$$
\begin{aligned}
&\int \Phi_{t1211,x}(X_i, X_j)dP(X_j) \\
=&\frac{1}{2}\mu_t'(X_i^\top\theta_0)f(X_i^\top\theta_0)\left(X_i 1(X_i^\top\beta \le u) + \eta_x(X_i^\top\theta_0)\right) + O_p(h^k),
\end{aligned}
$$

where we have used the facts that $\int K(v)dv = 1$, $\int v^j K(v)dv = 0$ for $j = 1, \dots, k-1$ and $\int v^k K(v)dv \neq 0$ in Assumption 5.2, and

$$
\mu_1 = \mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)X1(X^\top\beta \le u)\right] + O(h^k).
$$

As before, $S_{n,t1211}^{pro(2)}(x;\theta_0)$ is a degenerate U-process of second order. Proposition 4 of Delgado and González Manteiga (2001) then indicates that there exists a constant C, which does not depend on h, such that

$$
\begin{aligned}
\mathbb{E}\left[\sup_x \left\|\sqrt{n}S_{n,t1211}^{pro(2)}(x;\theta_0)\right\|^2\right] \le& C\frac{1}{n}\mathbb{E}\left[\|\Phi_{t1211,x}(X_i,X_j)\|^2\right] \\
\le& C\frac{1}{nh^2}\mathbb{E}\left[K^2\left(\frac{(X_i - X_j)^\top\theta_0}{h}\right)\right] \\
=& O\left(\frac{1}{nh}\right),
\end{aligned}
$$

by the assumptions of $\sup_v |\mu_t'(v)| < \infty$ and $\mathbb{E}||X||^2 < \infty$. This immediately yields that $S_{n,t1211}^{pro(2)}(x;\theta_0) = n^{-1/2}O_p((nh)^{-1/2}) = o_p(n^{-1/2})$ uniformly in x. Consequently, by the assumption of $nh^{2k} \to 0$, uniformly in x,

$$
\begin{aligned}
&S_{n,t1211}^{pro}(x;\theta_0) \\
=&\mu_1 + \frac{1}{n}\sum_{i=1}^n \left[\mu_t'(X_i^\top\theta_0)f(X_i^\top\theta_0)\left(X_i 1(X_i^\top\beta \le u) + \eta_x(X_i^\top\theta_0)\right) - 2\mu_1\right] + o_p\left(\frac{1}{\sqrt{n}}\right) \\
=&\mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)X1(X^\top\beta \le u)\right] \\
&+\frac{1}{n}\sum_{i=1}^n \left\{\mu_t'(X_i^\top\theta_0)f(X_i^\top\theta_0)\left(X_i 1(X_i^\top\beta \le u) + \eta_x(X_i^\top\theta_0)\right)\right. \\
&\left.-2\mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)X1(X^\top\beta \le u)\right]\right\} + o_p\left(\frac{1}{\sqrt{n}}\right) \\
=&\mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)X1(X^\top\beta \le u)\right] + o_p(1),
\end{aligned}
$$

where the last equality follows by an application of ULLN to the single i.i.d. summation in the second equality.

On the other hand,

$$
\sup_x \left\|S_{n,t1212}^{pro}(x;\widehat{\theta}_n)\right\|
$$

$$
\begin{aligned}
\leq & \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{h}\left|K\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)\right|\left|\mu_t''(X_i^\top\widetilde{\theta}_n)\right|\|X_i\|^2\\
= & O_p(h^{-1})
\end{aligned}
$$

under the assumptions that $\sup_v |K(v)| < \infty$, $\sup_v |\mu_t''(v)| < \infty$ and $\mathbb{E}||X||^2 < \infty$.

Therefore, uniformly in x,

$$
\begin{aligned}
S_{n,t121}^{pro}(x;\widehat{\theta}_n) = & \sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top\left(\mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)X1(X^\top\beta\leq u)\right]+o_p(1)\right)\\
& +O_p(1)O_p(h^{-1})O_p(n^{-1/2})\\
= & \sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top\mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)X1(X^\top\beta\leq u)\right]+o_p(1),
\end{aligned}
$$

under the assumptions that $||\widehat{\theta}_n-\theta_0|| = O_p(n^{-1/2})$ and $nh^3 \to \infty$.

For the analysis of $S_{n,t122}^{pro}(x;\widehat{\theta}_n)$, by the second order Taylor expansion of $\mu_t(v)$,

$$
\begin{aligned}
& S_{n,t122}^{pro}(x;\widehat{\theta}_n)\\
= & \sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{h}\left[K\left(\frac{(X_i-X_j)^\top\widehat{\theta}_n}{h}\right)-K\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)\right]\\
& \mu_t'(X_i^\top\theta_0)X_i1(X_i^\top\beta\leq u)\\
& +\frac{1}{2}\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{h}\left[K\left(\frac{(X_i-X_j)^\top\widehat{\theta}_n}{h}\right)-K\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)\right]\\
& \mu_t''(X_i^\top\widetilde{\theta}_n)X_iX_i^\top1(X_i^\top\beta\leq u)\left(\widehat{\theta}_n-\theta_0\right)\\
\equiv & \sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top S_{n,t1221}^{pro}(x;\widehat{\theta}_n)+\frac{1}{2}\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top S_{n,t1222}^{pro}(x;\widehat{\theta}_n)\left(\widehat{\theta}_n-\theta_0\right).
\end{aligned}
$$

It is straightforward to see that

$$
\begin{aligned}
& \sup_x\left\|S_{n,t1222}^{pro}(x;\widehat{\theta}_n)\right\|\\
\leq & \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{h}\left[\left|K\left(\frac{(X_i-X_j)^\top\widehat{\theta}_n}{h}\right)\right|+\left|K\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)\right|\right]\left|\mu_t''(X_i^\top\widetilde{\theta}_n)\right|\|X_i\|^2\\
= & O_p(h^{-1})
\end{aligned}
$$

under the assumptions that $\sup_v |K(v)| < \infty$, $\sup_v |\mu_t''(v)| < \infty$ and $\mathbb{E}||X||^2 < \infty$. Thus, uniformly in x, $\sqrt{n}(\widehat{\theta}_n-\theta_0)^\top S_{n,t1222}^{pro}(x;\widehat{\theta}_n)(\widehat{\theta}_n-\theta_0) = O_p((nh^2)^{-1/2}) = o_p(1)$ under the assumption of $nh^3 \to \infty$.

For the analysis of $S_{n,t1221}^{pro}(x;\widehat{\theta}_n)$, by the M-th order Taylor expansion of $K(\cdot)$, it can be decomposed as

$$
S_{n,t1221}^{pro}(x;\widehat{\theta}_n)
$$

$$
\begin{aligned}
&= \sum_{s=1}^{M-1} \frac{1}{s!} \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \frac{1}{h^{s+1}} K^{(s)} \left(\frac{(X_i - X_j)^\top \theta_0}{h} \right) \left[(X_i - X_j)^\top (\widehat{\theta}_n - \theta_0) \right]^s \\
&\quad \mu_t'(X_i^\top \theta_0) X_i 1(X_i^\top \beta \le u) \\
&\quad + \frac{1}{M!} \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \frac{1}{h^{M+1}} K^{(M)} \left(\frac{(X_i - X_j)^\top \widetilde{\theta}_n}{h} \right) \left[(X_i - X_j)^\top (\widehat{\theta}_n - \theta_0) \right]^M \\
&\quad \mu_t'(X_i^\top \theta_0) X_i 1(X_i^\top \beta \le u) \\
&\equiv \sum_{s=1}^{M-1} \frac{1}{s!} S^{pro}_{n,t1221s}(x; \widehat{\theta}_n) + \frac{1}{M!} S^{pro}_{n,t1221M}(x; \widehat{\theta}_n).
\end{aligned}
$$

Following similar arguments as proving $S^{pro}_{n,t112}(x; \widehat{\theta}_n) = o_p(1)$ and $S^{pro}_{n,t11M}(x; \widehat{\theta}_n) = o_p(1)$ uniformly in x in **Part I**, we can show that $S^{pro}_{n,t1221s}(x; \widehat{\theta}_n) = O_p((nh^2)^{-s/2}) = o_p(1)$ uniformly in x for $s = 1, \dots, M$. As a result, $S^{pro}_{n,t1221}(x; \widehat{\theta}_n) = o_p(1)$ uniformly in x, implying that $S^{pro}_{n,t122}(x; \widehat{\theta}_n) = O_p(1) o_p(1) + o_p(1) = o_p(1)$ uniformly in x. Combining the results of $S^{pro}_{n,t121}(x; \widehat{\theta}_n)$ and $S^{pro}_{n,t122}(x; \widehat{\theta}_n)$ yields that, uniformly in x,

$$
S^{pro}_{n,t12}(x; \widehat{\theta}_n) = \sqrt{n} \left(\widehat{\theta}_n - \theta_0 \right)^\top \mathbb{E} \left[\mu_t'(X^\top \theta_0) f(X^\top \theta_0) X 1(X^\top \beta \le u) \right] + o_p(1).
$$

Finally, combining the result of $S^{pro}_{n,t11}(x; \widehat{\theta}_n)$ in **Part I** and that of $S^{pro}_{n,t12}(x; \widehat{\theta}_n)$ in **Part II** yields the proof of (A.4). $\square$

Proof of (A.5): First note that $S^{pro}_{n,t2}(x; \widehat{\theta}_n)$ can be expressed as

$$
\begin{aligned}
&S^{pro}_{n,t2}(x; \widehat{\theta}_n) \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\widehat{\mu}_t(X_i^\top \widehat{\theta}_n) - \mu_t(X_i^\top \widehat{\theta}_n) \right) \widehat{f}(X_i^\top \widehat{\theta}_n) 1(X_i^\top \beta \le u) \\
=& \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \frac{1}{h} K \left(\frac{(X_i - X_j)^\top \widehat{\theta}_n}{h} \right) \left(1(T_j = t) - \mu_t(X_i^\top \widehat{\theta}_n) \right) 1(X_i^\top \beta \le u),
\end{aligned}
$$

which can be further decomposed as

$$
\begin{aligned}
&S^{pro}_{n,t2}(x; \widehat{\theta}_n) \\
=& \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \frac{1}{h} K \left(\frac{(X_i - X_j)^\top \theta_0}{h} \right) \left(1(T_j = t) - \mu_t(X_i^\top \theta_0) \right) 1(X_i^\top \beta \le u) \\
&- \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \frac{1}{h} K \left(\frac{(X_i - X_j)^\top \theta_0}{h} \right) \left(\mu_t(X_i^\top \widehat{\theta}_n - \mu_t(X_i^\top \theta_0) \right) 1(X_i^\top \beta \le u) \\
&+ \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \frac{1}{h} \left[K \left(\frac{(X_i - X_j)^\top \widehat{\theta}_n}{h} \right) - K \left(\frac{(X_i - X_j)^\top \theta_0}{h} \right) \right] \\
&\left(1(T_j = t) - \mu_t(X_i^\top \theta_0) \right) 1(X_i^\top \beta \le u)
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{1}{h} \left[K\left(\frac{(X_i - X_j)^\top \widehat{\theta}_n}{h} \right) - K\left(\frac{(X_i - X_j)^\top \theta_0}{h} \right) \right] \\
&\left(\mu_t(X_i^\top \widehat{\theta}_n) - \mu_t(X_i^\top \theta_0) \right) 1(X_i^\top \beta \leq u) \\
\equiv & S^{pro}_{n,t21}(x;\theta_0) - S^{pro}_{n,t22}(x;\widehat{\theta}_n) + S^{pro}_{n,t23}(x;\widehat{\theta}_n) - S^{pro}_{n,t24}(x;\widehat{\theta}_n).
\end{aligned}
$$

In the following we shall investigate the asymptotic properties of $S^{pro}_{n,t21}(x;\theta_0)$ and $S^{pro}_{n,t2j}(x;\widehat{\theta}_n)$ for $j = 2, 3, 4$.

Part I: Note that $S^{pro}_{n,t21}(x;\theta_0)$ can be rewritten as a second order U-process of the following form

$$
S^{pro}_{n,t21}(x;\theta_0) = \frac{2}{\sqrt{n}(n-1)} \sum_{1 \leq i < j \leq n}^{n} \Phi_{t21,x}(\chi_i, \chi_j),
$$

where

$$
\begin{aligned}
\Phi_{t21,x}(\chi_i, \chi_j) =& \frac{1}{2} \left[\frac{1}{h} K\left(\frac{(X_i - X_j)^\top \theta_0}{h} \right) \left(1(T_j = t) - \mu_t(X_i^\top \theta_0) \right) 1(X_i^\top \beta \leq u) \right. \\
&\left. + \frac{1}{h} K\left(\frac{(X_i - X_j)^\top \theta_0}{h} \right) \left(1(T_i = t) - \mu_t(X_j^\top \theta_0) \right) 1(X_j^\top \beta \leq u) \right].
\end{aligned}
$$

We denote by

$$
\mu_2 \equiv \int \Phi_{t21,x}(\chi_i, \chi_j) dP(\chi_i) dP(\chi_j),
$$

which is not exactly zero but converges to zero as $n \to \infty$ under the null H_0^{sim}. As before, we can decompose $S^{pro}_{n,t21}(x;\theta_0)$ by

$$
S^{pro}_{n,t21}(x;\theta_0) = \mu_2 + S^{pro(1)}_{n,t21}(x;\theta_0) + S^{pro(2)}_{n,t21}(x;\theta_0),
$$

where

$$
S^{pro(1)}_{n,t21}(x;\theta_0) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \left[\int \Phi_{t21,x}(\chi_i, \chi_j) dP(\chi_j) - \mu_2 \right],
$$

$$
\begin{aligned}
S^{pro(2)}_{n,t21}(x;\theta_0) =& \frac{2}{\sqrt{n}(n-1)} \sum_{1 \leq i < j \leq n}^{n} \left[\Phi_{t21,x}(\chi_i, \chi_j) - \int \Phi_{t21,x}(\chi_i, \chi_j) dP(\chi_i) \right. \\
&\left. - \int \Phi_{t21,x}(\chi_i, \chi_j) dP(\chi_j) - \mu_2 \right].
\end{aligned}
$$

Henceforth, we denote by $\phi_x(v) = \mathbb{E}[1(X^\top \beta \leq u) | X^\top \theta_0 = v]$, which is assumed to be k-times continuously differentiable with respect to v. Straightforward calculations yield that

$$
\int \Phi_{t21,x}(\chi_i, \chi_j) dP(\chi_j) = \frac{1}{2} e_i^{sim}(t;\theta_0) f(X_i^\top \theta_0) \phi_x(X_i^\top \theta_0) + O_p(h^k),
$$

where we have used the facts that $\int K(v)dv = 1$, $\int v^j K(v)dv = 0$ for $j = 1, \ldots, k-1$ and $\int v^k K(v)dv \neq 0$. In addition, we find that $\mu_2 = O(h^k)$ uniformly in x. Note that $S^{pro(2)}_{n,t21}(x;\theta_0)$ is also a degenerate U-process of second order. By applying Proposition 4 of Delgado and González Manteiga (2001), there exists a constant C, which does not depend on h, such that

$$\begin{aligned}\mathbb{E}\left[\sup_x \left|S^{pro(2)}_{n,t21}(x;\theta_0)\right|^2\right] \leq & C\frac{1}{n}\mathbb{E}\left[\Phi_{t21,x}(\chi_i,\chi_j)^2\right] \\ \leq & C\frac{1}{nh^2}\mathbb{E}\left[K^2\left(\frac{(X_i - X_j)^\top\theta_0}{h}\right)\right] \\ = & O\left(\frac{1}{nh}\right).\end{aligned}$$

implying that $\sup_x |S^{pro(2)}_{n,t21}(x;\theta_0)| = O_p((nh)^{-1/2}) = o_p(1)$ under $nh^3 \to \infty$. Thus, under $nh^{2k} \to 0$, we have, uniformly in x,

$$\begin{aligned}S^{pro}_{n,t21}(x;\theta_0) = & \mu_2 + \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[e^{sim}_i(t;\theta_0)f(X_i^\top\theta_0)\phi_x(X_i^\top\theta_0) - 2\mu_2\right] + O_p(\sqrt{n}h^k) + o_p(1) \\ = & \frac{1}{\sqrt{n}}\sum_{i=1}^n e^{sim}_i(t;\theta_0)f(X_i^\top\theta_0)\phi_x(X_i^\top\theta_0) + o_p(1).\end{aligned}$$

Part II: Note that $S^{pro}_{n,t22}(x;\widehat{\theta}_n)$ is simply identical to $S^{pro}_{n,t121}(x;\widehat{\theta}_n)$ in the proof of (A.4). We thus have, uniformly in x,

$$S^{pro}_{n,t121}(x;\widehat{\theta}_n) = \sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)^\top \mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)X1(X^\top\beta \leq u)\right] + o_p(1).$$

Part III: By the M-th order Taylor expansion of $K(v)$, $S^{pro}_{n,t23}(x;\widehat{\theta}_n)$ can be decomposed as

$$\begin{aligned}& S^{pro}_{n,t23}(x;\widehat{\theta}_n) \\ = & \sum_{s=1}^{M-1}\frac{1}{s!}\frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^n\sum_{j\neq i}^n \frac{1}{h^{s+1}}K^{(s)}\left(\frac{(X_i - X_j)^\top\theta_0}{h}\right)\left[(X_i - X_j)^\top(\widehat{\theta}_n - \theta_0)\right]^s \\ & \left(1(T_j = t) - \mu_t(X_i^\top\theta_0)\right)1(X_i^\top\beta \leq u) \\ & + \frac{1}{M!}\frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^n\sum_{j\neq i}^n \frac{1}{h^{M+1}}K^{(M)}\left(\frac{(X_i - X_j)^\top\widetilde{\theta}_n}{h}\right)\left[(X_i - X_j)^\top(\widehat{\theta}_n - \theta_0)\right]^M \\ & \left(1(T_j = t) - \mu_t(X_i^\top\theta_0)\right)1(X_i^\top\beta \leq u) \\ \equiv & \sum_{s=1}^{M-1}\frac{1}{s!}S^{pro}_{n,t23s}(x;\widehat{\theta}_n) + \frac{1}{M!}S^{pro}_{n,t23M}(x;\widehat{\theta}_n).\end{aligned}$$

First, note that

$$
\begin{aligned}
& S^{pro}_{n,t231}(x;\widehat{\theta}_n) \\
=& \sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}^n \frac{1}{h^2}K^{(1)}\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)(X_i-X_j) \\
& \left(1(T_j=t)-\mu_t(X_i^\top\theta_0)\right)1(X_i^\top\beta\leq u) \\
\equiv& \sqrt{n}\left(\widehat{\theta}_n-\theta_0\right)^\top S^{pro}_{n,t2311}(x;\theta_0).
\end{aligned}
$$

As before, we can rewrite $S^{pro}_{n,t2311}(x;\theta_0)$ as the following second order U-process

$$
S^{pro}_{n,t2311}(x;\theta_0)=\frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}^n \Phi_{t2311,x}(\chi_i,\chi_j),
$$

where, by the antisymmetry property $K^{(1)}(-v)=-K^{(1)}(v)$,

$$
\begin{aligned}
& \Phi_{t2311,x}(\chi_i,\chi_j) \\
=& \frac{1}{2}\left[\frac{1}{h^2}K^{(1)}\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)(X_i-X_j)\left(1(T_j=t)-\mu_t(X_i^\top\theta_0)\right)1(X_i^\top\beta\leq u)\right. \\
& \left.-\frac{1}{h^2}K^{(1)}\left(\frac{(X_i-X_j)^\top\theta_0}{h}\right)(X_j-X_i)\left(1(T_i=t)-\mu_t(X_j^\top\theta_0)\right)1(X_j^\top\beta\leq u)\right].
\end{aligned}
$$

We denote by

$$
\mu_3=\int \Phi_{t2311,x}(\chi_i,\chi_j)dP(\chi_i)dP(\chi_j).
$$

Again, we can decompose $S^{2pro}_{n,t2311}(x;\theta_0)$ by

$$
S^{pro}_{n,t2311}(x;\theta_0)=\mu_3+S^{pro(1)}_{n,t2311}(x;\theta_0)+S^{pro(2)}_{n,t2311}(x;\theta_0),
$$

where

$$
S^{pro(1)}_{n,t2311}(x;\theta_0)=\frac{2}{n}\sum_{i=1}^n\left[\int \Phi_{t2311,x}(\chi_i,\chi_j)dP(\chi_j)-\mu_3\right],
$$

$$
\begin{aligned}
S^{pro(2)}_{n,t2311}(x;\theta_0)=& \frac{2}{n(n-1)}\sum_{1\leq i<j\leq n}^n\left[\Phi_{t2311,x}(\chi_i,\chi_j)-\int \Phi_{t2311,x}(\chi_i,\chi_j)dP(\chi_i)\right. \\
& \left.-\int \Phi_{t2311,x}(\chi_i,\chi_j)dP(\chi_j)-\mu_3\right].
\end{aligned}
$$

Standard calculations yield that

$$
\begin{aligned}
& \int \Phi_{t2311,x}(\chi_i,\chi_j)dP(\chi_j) \\
=& \mu_t'(X_i^\top\theta_0)f(X_i^\top\theta_0)\left(\mathbb{E}[X_i 1(X_i^\top\beta\leq u)|X_i^\top\theta_0]-X_i\mathbb{E}[1(X_i^\top\beta\leq u)|X_i^\top\theta_0]\right)+O_p(h^k),
\end{aligned}
$$

where we have used $\int K^{(1)}(v)dv = 0$, $\int vK^{(1)}(v)dv = -1$ and $\int v^j K^{(1)}(v)dv = 0$ for $j = 2, \ldots, k$. In addition, we find that, uniformly in x,

$$\begin{aligned}\mu_3 =& \mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)X1(X^\top\beta \le u)\right] \\ &- \mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)\mathbb{E}[X|X^\top\theta_0]1(X^\top\beta \le u)\right] + O(h^k).\end{aligned}$$

Since $S^{pro(2)}_{n,t2311}(x;\theta_0)$ is a degenerate U-process of second order, by applying Proposition 4 of Delgado and González Manteiga (2001), there exists a constant C, which does not depend on h, such that

$$\begin{aligned}\mathbb{E}\left[\sup_x \left|\sqrt{n}S^{pro(2)}_{n,t2311}(x;\theta_0)\right|^2\right] \le& C\frac{1}{n}\mathbb{E}\left[\Phi_{t2311,x}(\chi_i,\chi_j)^2\right] \\ \le& C\frac{1}{nh^4}\mathbb{E}\left[K^{(1)}\left(\frac{(X_i - X_j)^\top\theta_0}{h}\right)^2\right] \\ =& O\left(\frac{1}{nh^3}\right),\end{aligned}$$

which implies that $\sup_x |S^{pro(2)}_{n,t1111}(x;\theta_0)| = O_p((n^2h^3)^{-1/2}) = o_p(1)$ under the assumption of $nh^3 \to \infty$. Thus, under $h \to 0$, uniformly in x,

$$\begin{aligned}& S^{pro}_{n,t2311}(x;\theta_0) \\ =& \mu_3 + \frac{2}{n}\sum_{i=1}^n \left\{\mu_t'(X_i^\top\theta_0)f(X_i^\top\theta_0)\left(\mathbb{E}[X_i 1(X_i^\top\beta \le u)|X_i^\top\theta_0] - X_i\mathbb{E}[1(X_i^\top\beta \le u)|X_i^\top\theta_0]\right)\right. \\ & \left. -\mu_3\right\} + O_p(h^k) + o_p(1) \\ =& \mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)X1(X^\top\beta \le u)\right] - \mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)\mathbb{E}[X|X^\top\theta_0]1(X^\top\beta \le u)\right] \\ & + o_p(1),\end{aligned}$$

where the second equality follows by an application of ULLN to the single i.i.d. summation in the first equality. Consequently, uniformly in x,

$$\begin{aligned}& S^{pro}_{n,t231}(x;\widehat{\theta}_n) \\ =& \sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)^\top \mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)X1(X^\top\beta \le u)\right] \\ & - \sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)^\top \mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)\mathbb{E}[X|X^\top\theta_0]1(X^\top\beta \le u)\right] + o_p(1).\end{aligned}$$

Following similar arguments, we can prove that $S^{pro}_{n,t23s}(x;\widehat{\theta}_n) = o_p((nh^2)^{-s/2}) = o_p(1)$ uniformly in x for $s = 2, \ldots, M-1$. Lastly,

$$\begin{aligned}& \sup_x \left|S^{pro}_{n,t23M}(x;\widehat{\theta}_n)\right| \\ \le& \frac{2}{\sqrt{n}(n-1)}\sum_{i=1}^n\sum_{j\ne i}^n \frac{1}{h^{M+1}}\left|K^{(M)}\left(\frac{(X_i - X_j)^\top\widetilde{\theta}_n}{h}\right)\right| \|X_i - X_j\|^M \left\|\widehat{\theta}_n - \theta_0\right\|^M\end{aligned}$$

$$=\frac{\sqrt{n}}{h^{M+1}}O_p(1)O_p(n^{-M/2}) = O_p((n^{M-1}h^{2(M+1)})^{-1/2}) = o_p(1),$$

under the assumptions that $\sup_v |K^{(M)}(v)| < \infty$, $\mathbb{E}||X||^M < \infty$ and $n^{M-1}h^{2(M+1)} \to \infty$, which is implied by $nh^3 \to \infty$ when $M \geq 5$.

Therefore, collecting all the previous results, we have, uniformly in x,

$$\begin{aligned}S^{pro}_{n,t23}(x;\widehat{\theta}_n) =& S^{pro}_{n,t231}(x;\widehat{\theta}_n) + o_p(1)\\ =&\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)^\top \mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)X1(X^\top\beta \leq u)\right]\\ &- \sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)^\top \mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)\mathbb{E}[X|X^\top\theta_0]1(X^\top\beta \leq u)\right] + o_p(1).\end{aligned}$$

Part IV: Note that $S^{pro}_{n,t24}(x;\widehat{\theta}_n)$ is simply identical to $S^{pro}_{n,t122}(x;\widehat{\theta}_n)$ in the proof of (A.4). Thus, we have $S^{pro}_{n,t24}(x;\widehat{\theta}_n) = o_p(1)$ uniformly in x.

Finally, combining the results obtained for $S^{pro}_{n,t21}(x;\theta_0)$ and $S^{pro}_{n,t2j}(x;\widehat{\theta}_n)$ for $j = 2, 3, 4$ yields that, uniformly in x,

$$\begin{aligned}S^{pro}_{n,t2}(x;\widehat{\theta}_n) =& S^{pro}_{n,t21}(x;\theta_0) - S^{pro}_{n,t22}(x;\widehat{\theta}_n) + S^{pro}_{n,t23}(x;\widehat{\theta}_n) + o_p(1)\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} e_i^{sim}(t;\theta_0)f(X_i^\top\theta_0)\phi_x(X_i^\top\theta_0)\\ &- \sqrt{n}\left(\widehat{\theta}_n - \theta_0\right)^\top \mathbb{E}\left[\mu_t'(X^\top\theta_0)f(X^\top\theta_0)\mathbb{E}[X|X^\top\theta_0]1(X^\top\beta \leq u)\right] + o_p(1),\end{aligned}$$

which completes the proof of (A.5). □

Appendix B: Monte Carlo simulations

In this section, we conduct a series of Monte Carlo experiments to study the finite sample properties of the double projection-based tests in the context of observational treatment effect studies. We consider three different setups with (i) a binary treatment, (ii) multinomial, unordered treatments (henceforth multinomial treatments), and (iii) multinomial, ordered treatments (henceforth ordered treatments).

Given that the simulation results in Sant'Anna and Song (2019) indicate that their test dominates several others in terms of size and power in the binary treatment setup, we only compare our proposed CvM^{dpro}_n statistic given in (2.11) to their CvM statistic. For the ordered and multinomial treatment setups, we consider extensions of the Sant'Anna and Song (2019)'s projection-based tests that are able to accommodate multi-valued treatment variables.

Critical values for the test statistic CvM^{dpro}_n are obtained using the multiplier bootstrap procedure described in Section 4, whereas for the single projection-based tests, we use the multiplier bootstrap procedure described in Sant'Anna and Song (2019).

We consider sample sizes n equal to 200, 400, and 800. For each design, we consider $1,000$ Monte Carlo experiments. The multipliers $\{V_i, i = 1, \ldots, n\}$ used in the bootstrap implementations are independently generated as V with $\mathbb{P}(V = 1 - \kappa) = \kappa/\sqrt{5}$ and $\mathbb{P}(V = \kappa) = 1 - \kappa/\sqrt{5}$, where $\kappa = \left(\sqrt{5} + 1\right)/2$, as proposed by Mammen (1993). The bootstrapped critical values are approximated using $B = 999$ bootstrap replications. Moreover, we report simulation results from the test statistic $CvM_n^{im,dpro}$ given in (5.7), which is used to test a parametric (multiple) index propensity score model against a semiparametric alternative. In all simulations, we use the constant weight $a_n(t) \equiv 1$ for all $t \in \mathcal{T}$ in CvM_n^{dpro} and $CvM_n^{im,dpro}$ as well as their bootstrap counterparts $CvM_n^{dpro,*}$ and $CvM_n^{im,dpro,*}$.

Simulation 1: Binary treatment

We first consider the binary treatment case with $J = 1$. Consider the following data generating processes (DGPs), which are similar to Sant'Anna and Song (2019):

$$DGP1.\ T^* = -\frac{\sum_{j=1}^{10} X_j}{6} - \varepsilon;$$

$$DGP2.\ T^* = -1 - \frac{\sum_{j=1}^{10} X_j}{10} + \frac{X_1 X_2}{2} - \varepsilon;$$

$$DGP3.\ T^* = -1 - \frac{\sum_{j=1}^{10} X_j}{10} + \frac{X_1 \sum_{k=2}^{5} X_k}{4} - \varepsilon;$$

$$DGP4.\ T^* = -1.5 - \frac{\sum_{j=1}^{10} X_j}{6} + \frac{\sum_{k=1}^{10} X_k^2}{10} - \varepsilon;$$

$$DGP5.\ T^* = \frac{-0.1 + 0.1 \sum_{j=1}^{5} X_j}{\exp\left(-0.2 \sum_{k=1}^{10} X_j\right)} - \varepsilon.$$

For each of these five DGPs, $T = 1\{T^* > 0\}, \varepsilon \perp\!\!\!\perp X$, with $X = (1, X_1, X_2, \ldots, X_{10})^\top$ where $X_1 = Z_1$, $X_2 = (Z_1 + Z_2)/\sqrt{2}$, $X_k = Z_k$, $k = 3, \ldots, 10$, and $\{Z_k\}_{k=1}^{10}$ and ε are independent standard normal random variables. For each of these DGPs we consider the following potential outcomes:

$$Y(1) = 2m_1(X) + u(1) \quad \text{and} \quad Y(0) = m_1(X) + u(0),$$

where $m_1(X) = 1 + \sum_{j=1}^{10} X_j$, $u(1)$ and $u(0)$ are independent normal random variables with mean zero and variance one. The observed outcome is $Y = TY(1) + (1 - T) Y(0)$, and the true average treatment effect (ATE) is 1. Although these outcome equations are not necessary to assess the size and power properties of the tests, they can be used to assess the utility of our proposed tests to distinguish between " good" and "bad" estimates of the ATE.

For $DGP1 - DGP5$, the null hypothesis H_0 considered is

$$H_0 : \exists \theta_0 = (\delta_0, \delta_1, \ldots, \delta_{10})^\top \in \Theta : \mathbb{E}[T|X] = \Phi\left(X^\top \theta_0\right) \ a.s., \tag{B.1}$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution. Note that under the null hypothesis in (B.1), the propensity score model satisfies a parametric single-index restriction with $q(x,\theta) = \Phi(x^\top \theta)$ and the corresponding score function $g(x,\theta) = \partial q(x,\theta)/\partial\theta = \phi(x^\top\theta)x$, with $\phi(\cdot)$ the PDF of the standard normal distribution. We estimate θ_0 using the probit maximum likelihood, i.e.,

$$\widehat{\theta}_n = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \left[T_i \ln\left(\Phi\left(X_i^\top\theta\right)\right) + (1-T_i)\ln\left(1-\Phi\left(X_i^\top\theta\right)\right)\right].$$

Clearly, $DGP1$ falls under H_0, whereas $DGP2 - DGP5$ fall under H_1, i.e., the negation of (B.1). Note that the treatment status T follows a heteroskedastic probit model in $DGP5$.

We compare the finite sample performance of our proposed omnibus test statistic CvM_n^{dpro} in (2.11) based on double projections with the directional test statistic $CvM_n^{im,dpro}$ in (5.7), as well as Sant'Anna and Song (2019)'s CvM test statistic

$$CvM_n^{ss} = \frac{1}{n}\sum_{i=1}^{n}\left(R_n^{ss}\left(q\left(X_i,\widehat{\theta}_n\right)\right)\right)^2, \tag{B.2}$$

where $q\left(X_i,\widehat{\theta}_n\right) = \Phi\left(X_i^\top\widehat{\theta}_n\right)$ and $R_n^{ss}(u) \equiv n^{-1/2}\sum_{j=1}^{n} e_j(\widehat{\theta}_n)\mathcal{P}_n 1\left(q(X_j,\widehat{\theta}_n) \leq u\right)$, with $e(\theta) = T - q(X,\theta)$,

$$\mathcal{P}_n 1(q(X,\theta)\leq u) = 1(q(X,\theta)\leq u) - g(X,\theta)^\top \Delta_n^{-1}(\theta)\, G_n(u,\theta),$$

$G_n(u,\theta) = n^{-1}\sum_{i=1}^{n} g(X_i,\theta)1(q(X_i,\theta)\leq u)$, and $\Delta_n(\theta) = n^{-1}\sum_{i=1}^{n} g(X_i,\theta)g^\top(X_i,\theta)$.

The simulation results are presented in Table B.1. We report empirical rejection frequencies at the 5% significance level. Results for 10% and 1% significance levels are similar and are available upon request. We also report the bias, root mean squared error (RMSE), and coverage of the 95% confidence interval of the following *stabilized* inverse probability weighted estimator:

$$ATE_n = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{w_{i,1}}{\bar{w}_{n,1}} - \frac{w_{i,0}}{\bar{w}_{n,0}}\right)Y_i, \tag{B.3}$$

where $w_{i,1} = T_i/q\left(X_i,\widehat{\theta}_n\right)$ and $w_{i,1} = (1-T_i)/\left(1-q\left(X_i,\widehat{\theta}_n\right)\right)$ are the inverse probability weights (Hájek, 1971), and $\bar{w}_{n,t}$ is the sample mean of $w_{i,t}$ for $t=0,1$. The 95% confidence interval is estimated via the percentile bootstrap with 499 draws.

We first analyze the size of our proposed test. From the results of $DGP1$, we find that the actual finite sample sizes of all considered tests are close to their nominal size, even when the sample size is as small as 200. In addition, when the propensity score is correctly specified, the bias of the ATE_n estimator in (B.3) is small, and the coverage probability is close to its nominal value even when $n = 200$.

Table B.1: Monte Carlo results under designs $DGP1$-$DGP5$: Binary Treatment

DGP	n	CvM_n^{dpro}	$CvM_n^{im,dpro}$	CvM_n^{ss}	Bias	RMSE	COV
1	200	0.060	0.059	0.057	-0.085	0.623	0.962
1	400	0.053	0.057	0.056	-0.007	0.504	0.948
1	800	0.053	0.055	0.054	-0.006	0.381	0.936
2	200	0.648	0.147	0.154	0.582	1.301	0.986
2	400	0.990	0.237	0.264	0.613	0.958	0.894
2	800	1.000	0.458	0.495	0.589	0.709	0.671
3	200	0.356	0.175	0.183	1.106	1.768	0.976
3	400	0.885	0.444	0.465	1.359	1.764	0.705
3	800	0.999	0.770	0.799	1.330	1.530	0.256
4	200	0.368	0.148	0.151	0.603	1.619	0.975
4	400	0.856	0.267	0.304	0.963	1.772	0.947
4	800	1.000	0.509	0.576	1.031	1.424	0.708
5	200	0.123	0.098	0.100	-0.106	0.358	0.974
5	400	0.265	0.191	0.192	-0.105	0.244	0.947
5	800	0.590	0.497	0.502	-0.118	0.192	0.892

Note: Simulations based on $1,000$ Monte Carlo experiments. "CvM_n^{dpro}" stands for our proposed double-projected Cramér-von Mises test in (2.11). "$CvM_n^{im,dpro}$" stands for the directional Cramér-von Mises test in (5.7). "CvM_n^{ss}" stands for Sant'Anna and Song (2019)'s test defined in (B.2). Finally, "Bias", "RMSE" and "COV" stand for the average simulated bias, average simulated root mean squared error, and 95% coverage probability for the ATE estimator ATE_n as defined in (B.3). The 95% coverage probability is based on the percentile bootstrap with 499 draws. See the main text for further details.

Note that when the propensity score is misspecified in $DGP2$-$DGP5$, the ATE estimator (B.3) can be severely biased and its 95% confidence interval is too liberal, i.e., it can severely undercover the true ATE. Thus, tests with a higher power to detect such model misspecifications can prevent one from making misleading conclusions about the effectiveness of a given policy. Our proposed CvM_n^{dpro} test perform admirably well in such a task. Perhaps, what is more important to emphasize in terms of power is that in all alternative hypotheses and sample sizes analyzed, our omnibus CvM_n^{dpro} test has substantially higher power than our directional test, $CvM_n^{im,dpro}$, and Sant'Anna and Song (2019)'s CvM_n^{ss} test. For instance, for $DGP2$ with $n = 200$, our proposed omnibus test is four times more powerful than Sant'Anna and Song (2019)'s test, which is already more powerful than other tests available in the literature, including covariate balancing tests and traditional specification test based on kernel method; see Section 4.2 of Sant'Anna and Song (2019) for additional details. In addition, the power properties of $CvM_n^{im,dpro}$ and CvM_n^{ss} are similar. This, however, is not surprising given that the one-dimensional estimated conditioning variable $\Phi(X_i^\top \widehat{\theta}_n)$ in CvM_n^{ss} is only a strictly increasing transformation of the one-dimensional estimated linear-index $X_i^\top \widehat{\theta}_n$ (i.e., $\widetilde{X}_{i,\widehat{\theta}_n}$) in $CvM_n^{im,dpro}$.

Simulation 2: Multinomial treatments

In this section, we consider unordered, multinomial treatments. Our DGPs are similar to Yang et al. (2016). The covariates X_1, X_2, X_3 are generated from a multivariate normal distribution with mean zero, variances of $(2, 1, 1)$ and covariances of $(1, -1, -0.5)$; X_4 follows a uniform distribution from -3 to 3; X_5 follows a chi-squared distribution with one degree of freedom; and X_6 follows a Bernoulli distribution with $p = 0.5$. Let $X = (X_1, \ldots, X_6)^\top$. We consider three treatment groups, $T = \{0, 1, 2\}$, whose assignment mechanism follows the multinomial logistic model

$$(T_0, T_1, T_2) \,|X \sim \mathit{Multinomial}\ (p_0 (X), p_1 (X), p_2 (X)),$$

where T_t is the treatment indicator, i.e., $T_t = 1 (T = t)$, and for $t = 0, 1, 2$,

$$p_t (X) = \frac{\exp (\phi_t (X))}{\sum_{s=0}^2 \exp (\phi_s (X))}.$$

In what follows, we take $\phi_0 (X) = 0$ and vary $\phi_1 (X)$ and $\phi_2 (X)$ as follows:

$$DGP6.\ \phi_1 (X) = -1 + 0.4 \sum_{j=1}^{6} X_j,\ \phi_2 (X) = -1 + 0.2 \sum_{j=1}^{6} X_j;$$

$$DGP7.\ \phi_1 (X) = -0.2 \sum_{j=1}^{6} X_j + X_1 X_6, \phi_2 (X) = -0.1 \sum_{j=1}^{6} X_j + X_1 X_4;$$

$$DGP8.\ \phi_1 (X) = 0.3 \sum_{j=1}^{6} X_j,\ \phi_2 (X) = -0.5 + 0.1 \sum_{j=1}^{6} X_j^2;$$

$$DGP9.\ \phi_1 (X) = -0.1 + \frac{\sum_{j=1}^6 X_j}{5} + \frac{X_6 \sum_{l=1}^3 X_l}{2},\ \phi_2 (X) = -0.3 \sum_{j=1}^{6} X_j - \frac{X_6 (X_4 + X_5)}{2};$$

$$DGP10.\ \phi_1 (X) = \sin \left(\sum_{j=1}^{6} X_j \right) + \sum_{l=1}^{3} X_l,\ \phi_2 (X) = 2 \sin \left(\sum_{j=1}^{6} X_j \right) + \frac{1}{2} \sum_{l=1}^{3} X_l.$$

For each of these DGPs, we consider the potential outcomes

$$Y(0) = 1 + X^\top \beta_0 + u(0), \quad Y(1) = 20 + X^\top \beta_1 + u(1), \quad \text{and} \quad Y(2) = 6 + X^\top \beta_2 + u(2),$$

where $u(0)$, $u(1)$ and $u(2)$ are independent normal random variables with mean zero and variance 1, $\beta_0 = (-6, -6, -6, 6, 6, 6)^\top$, $\beta_1 = -\beta_0$, and $\beta_2 = 4$. The observed outcome is $Y = 1(T=0)Y(0) + 1(T=1)Y(1) + 1(T=2)Y(2)$, and the true $ATE_{1,0} = 1$, $ATE_{2,0} = 2$, and $ATE_{2,1} = 1$.

Let $\alpha = (\alpha_1, \alpha_2)^\top$ and $\delta = \left(\delta_1^\top, \delta_2^\top\right)^\top$. For $DGP6 - DGP10$, the H_0 considered is

$$H_0 : \exists \theta_0 = (\alpha^\top, \delta^\top)^\top \in \Theta : \mathbb{P}(T = t|X) = \frac{\exp\left(\alpha_t + X^\top \delta_t\right)}{\sum_{s=0}^2 \exp\left(\alpha_s + X^\top \delta_s\right)} \ a.s., \text{ for } t = 1, 2, \tag{B.4}$$

where, with some abuse of notation, we set $\alpha_0 = \delta_0 = 0$. Under (B.4), the GPS satisfies parametric single-index restrictions with the link function being the multinomial logistic CDF. We estimate θ_0 using the multinomial logit likelihood, i.e.,

$$\widehat{\theta}_n = \arg\max_{\theta \in \Theta} \sum_{i=1}^n \sum_{t=0}^2 \left[T_{i,t} \cdot \ln\left(\frac{\exp\left(\alpha_t + X_i^\top \delta_t\right)}{\sum_{s=0}^2 \exp\left(\alpha_s + X_i^\top \delta_s\right)} \right) \right].$$

Clearly, $DGP6$ falls under H_0, whereas $DGP7 - DGP10$ fall under H_1, i.e., the negation of (B.4). Note that either interactive, quadratic, or periodic functions of components in X enter $DGP7 - DGP10$.

In the multinomial setup, our proposed test statistic CvM_n^{dpro} is given by (2.11), where, for $t = 1, 2$, $a_n(t) = 1$, the residual $e_i\left(t; \widehat{\theta}_n\right) = 1(T_i = t) - q_t\left(X_i, \widehat{\theta}_n\right)$,

$$q_t\left(X_i, \widehat{\theta}_n\right) = \Lambda_t^m(X_i, \theta) = \frac{\exp\left(\alpha_t + X_i^\top \delta_t\right)}{\sum_{s=0}^2 \exp\left(\alpha_s + X_i^\top \delta_s\right)},$$

and the score function

$$g_t(X_i, \widehat{\theta}_n) = q_t(X_i, \widehat{\theta}_n)(1 - q_t(X_i, \widehat{\theta}_n))(1 \ \ X_i)^\top.$$

Although Sant'Anna and Song (2019) only considered specification tests for binary treatments, we note that their tests can also be extended to test (B.4). We consider two different extensions of Sant'Anna and Song (2019)'s specification tests. The first variant of Sant'Anna and Song (2019)'s test statistic is given by

$$CvM_n^{ss1,m} = CvM_{n,1}^{ss1,m} + CvM_{n,2}^{ss1,m}, \tag{B.5}$$

where, for $t = 1, 2$, $CvM_{n,t}^{ss1,m} = n^{-1} \sum_{i=1}^n \left(R_{n,t}^{ss1,m}\left(q_1\left(X_i, \widehat{\theta}_n\right), q_2\left(X_i, \widehat{\theta}_n\right)\right)\right)^2$, with

$$R_{n,t}^{ss1,m}(u_1, u_2) \equiv \frac{1}{\sqrt{n}} \sum_{j=1}^n e_j(t; \widehat{\theta}_n) \mathcal{P}_{n,t}\left(1\left(\Lambda_1^m\left(X_j, \widehat{\theta}_n\right) \le u_1\right) 1\left(\Lambda_2^m\left(X_j, \widehat{\theta}_n\right) \le u_2\right)\right),$$

and

$$\begin{aligned} &\mathcal{P}_{n,t}\left(1\left(\Lambda_1^m(X, \theta) \le u_1\right) 1\left(\Lambda_2^m(X, \theta) \le u_2\right)\right) \\ =&1\left(\Lambda_1^m(X, \theta) \le u_1\right) 1\left(\Lambda_2^m(X, \theta) \le u_2\right) - g_t X, \theta)^\top \Delta_{n,t}^{-1}(\theta) G_{n,t}^m(u_1, u_2; \theta), \end{aligned}$$

$g_t(X,\theta)$ and $\Delta_{n,t}^{-1}(\theta)$ being defined as before, and

$$G_{n,t}^{m}(u_1,u_2;\theta)=\frac{1}{n}\sum_{i=1}^{n}g_t(X_i,\theta)1\left(\Lambda_1^m(X,\theta)\le u_1\right)1\left(\Lambda_2^m(X,\theta)\le u_2\right).$$

The second variant of Sant'Anna and Song (2019) test statistic is given by

$$CvM_n^{ss2,m}=CvM_{n,1}^{ss2,m}+CvM_{n,2}^{ss2,m}, \tag{B.6}$$

where, for $t=1,2$, $CvM_{n,t}^{ss2,m}=n^{-1}\sum_{i=1}^{n}\left(R_{n,t}^{ss2,m}\left(\Lambda_t^m\left(X_i,\widehat{\theta}_n\right)\right)\right)^2$, and

$$R_{n,t}^{ss2,m}(u)\equiv\frac{1}{\sqrt{n}}\sum_{j=1}^{n}e_j(t;\widehat{\theta}_n)\mathcal{P}_{n,t}1\left(\Lambda_t^m\left(X_j,\widehat{\theta}_n\right)\le u\right),$$

with

$$\mathcal{P}_{n,t}1\left(\Lambda_t^m(X,\theta)\le u\right)=1\left(\Lambda_t^m(X,\theta)\le u\right)-g_t(X,\theta)^\top\Delta_{n,t}^{-1}(\theta)\,G_{n,t}(u,\theta),$$

and $g_t(X,\theta)$, $G_{n,t}(u,\theta)$ and $\Delta_{n,t}(\theta)$ defined analogously to the binary treatment setup.

It is important to emphasize that (B.5) and (B.6) are test statistics for *implications* of (B.4), and not for (B.4) itself. More precisely, (B.5) is a test statistic for the null hypothesis

$$H_0':\exists\theta_0=(\alpha^\top,\delta^\top)^\top\in\Theta:\mathbb{E}\left[e^m(t;\theta_0)|\Lambda_1^m(X,\theta_0),\Lambda_2^m(X,\theta_0)\right]=0\ \ a.s.\text{ for }t=0,1.$$

whereas (B.6) is a test statistic based on the null hypothesis

$$H_0'':\exists\theta_0=(\alpha^\top,\delta^\top)^\top\in\Theta:\mathbb{E}\left[e^m(t;\theta_0)|\Lambda_t^m(X,\theta_0)\right]=0\ a.s.\text{ for }t=0,1,$$

Importantly and in sharp contrast to CvM_n^{dpro}, both $CvM_n^{ss1,m}$ and $CvM_n^{ss2,m}$ are not consistent against general nonparametric alternatives H_1 in (2.2). Similarly, our directional test statistic $CvM_n^{im,dpro}$ for H_0^{im1} is consistent against the semiparametric alternatives H_1^{im1} in (5.3), but not necessarily consistent against general nonparametric alternatives H_1^{im}, the negation of H_0^{im} in (5.1).

The simulation results are presented in Table B.2. We report empirical rejection frequencies at the 5% significance level. We also report the bias, RMSE, and coverage of the 95% confidence interval of the following *stabilized* inverse probability weighted estimators based on the identification result in (2.4):

$$ATE_{n,j,\ell}=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{w_{i,j}}{\bar{w}_{n,j}}-\frac{w_{i,\ell}}{\bar{w}_{n,\ell}}\right)Y_i, \tag{B.7}$$

where

$$w_{i,0} = \frac{1\{T_i = 0\}}{q_0\left(X_i, \widehat{\theta}_n\right)}, \quad w_{i,1} = \frac{1\{T_i = 1\}}{q_1\left(X_i, \widehat{\theta}_n\right)} \quad \text{and} \quad w_{i,2} = \frac{1\{T_i = 2\}}{q_2\left(X_i, \widehat{\theta}_n\right)}$$

are the inverse probability weights, and $\bar{w}_{n,t}$ is the sample mean of $w_{i,t}$ for $t = 0, 1, 2$. The 95% confidence interval is estimated via the percentile bootstrap with 499 draws.

Table B.2: Monte Carlo results under designs $DGP6$-$DGP10$: Multinomial Treatment

DGP	n	CvM_n^{dpro}	$CvM_n^{im,dpro}$	$CvM_n^{ss1,m}$	$CvM_n^{ss2,m}$	$Bias_{1,0}$	$Bias_{2,0}$	$RMSE_{1,0}$	$RMSE_{2,0}$	$COV_{1,0}$	$COV_{2,0}$
6	200	0.054	0.064	0.051	0.048	0.286	0.075	3.493	2.458	0.955	0.970
6	400	0.057	0.049	0.044	0.063	0.033	0.064	2.509	1.744	0.943	0.953
6	800	0.060	0.053	0.044	0.059	0.045	0.063	1.722	1.165	0.935	0.954
7	200	0.993	0.254	0.219	0.331	0.126	2.074	3.409	2.772	0.961	0.838
7	400	1.000	0.499	0.442	0.536	-0.166	1.974	2.363	2.278	0.962	0.637
7	800	1.000	0.770	0.718	0.728	-0.005	2.037	1.536	2.184	0.962	0.271
8	200	0.445	0.122	0.074	0.132	1.122	1.172	3.113	2.303	0.926	0.914
8	400	0.835	0.222	0.087	0.240	0.859	1.038	2.141	1.649	0.917	0.865
8	800	0.992	0.475	0.140	0.473	0.906	1.045	1.615	1.369	0.886	0.784
9	200	0.099	0.084	0.057	0.079	0.746	-0.829	3.232	3.815	0.945	0.888
9	400	0.149	0.123	0.076	0.116	0.558	-0.132	2.185	3.524	0.938	0.863
9	800	0.258	0.174	0.061	0.137	0.494	-0.167	1.608	3.042	0.927	0.834
10	200	0.133	0.076	0.090	0.075	-0.528	-1.369	4.043	2.803	0.954	0.872
10	400	0.302	0.102	0.157	0.137	-0.548	-1.108	3.406	2.347	0.939	0.823
10	800	0.747	0.176	0.332	0.281	-0.843	-1.285	2.424	1.818	0.921	0.689

Note: Simulations based on 1,000 Monte Carlo experiments. "CvM_n^{dpro}" stands for our proposed Cramér-von Mises tests (2.11). "$CvM_n^{im,dpro}$" stands for the directional Cramér-von Mises test in (5.7). "$CvM_n^{ss1,m}$" and "$CvM_n^{ss2,m}$" stand for extensions of Sant'Anna and Song (2019)'s test defined in (B.5) and (B.6), respectively. Finally, "$Bias_{k,s}$", "$RMSE_{k,s}$" and "$COV_{k,s}$" stands for the average simulated bias, average simulated root mean squared error, and 95% coverage probability for the $ATE_{k,s}$ estimator $ATE_{n,k,s}$ as defined in (B.7). The 95% coverage probability is based on the percentile bootstrap with 499 draws. See the main text for further details.

As before, we first discuss the size properties of the tests. From the results of $DGP6$, we find that all considered tests have good size properties and the IPW estimators for the average treatment effects have little to no bias, their RMSEs decrease with sample size, and their coverage probabilities are very close to the nominal level. Among the considered test statistics, our directional test $CvM_n^{im,dpro}$ is the only one with some size distortions when sample size $n = 200$, but such distortions disappear as n increases.

In terms of power, note that, under $DGP7 - DGP10$, our proposed double-projection omnibus test CvM_n^{dpro} outperforms $CvM_n^{im,dpro}$, $CvM_n^{ss1,m}$ and $CvM_n^{ss2,m}$ in all considered scenarios by a significant margin. We note that $CvM_n^{ss2,m}$ tends to outperform $CvM_n^{ss1,m}$ in all DGPs, except $DGP10$. We also note that the power performance of $CvM_n^{im,dpro}$ and $CvM_n^{ss2,m}$ is similar. Finally, it is evident from Table B.2 that GPS misspecification can indeed lead to misleading conclusions about the treatment effect effectiveness.

Simulation 3: ordered treatments

In this section, we consider ordered, multinomial treatments with $J = 2$. Specifically, we consider three treatment groups, $T = \{0, 1, 2\}$, whose assignment mechanisms are given by the following conditional distributions:

$$\mathbb{P}\left(T \leq t|X\right) = \Lambda\left(\frac{\pi}{\sqrt{3}}\frac{\alpha_t - \phi\left(X\right)}{\gamma\left(X\right)}\right),\ t = 0, 1,$$

where $\Lambda\left(\cdot\right)$ is the logistic CDF, i.e., $\Lambda\left(a\right) = \exp\left(a\right)/\left(1+\exp\left(a\right)\right)$, $\alpha_1 > \alpha_0$, and of course, $\mathbb{P}\left(T \leq 2|X\right) = 1$ *a.s.* by construction.[1] To formulate our DGPs, we vary $\phi\left(X\right)$, $\gamma\left(X\right)$ and α_t as following:

$$DGP11.\ \phi\left(X\right) = -\frac{\sum_{j=1}^{10} X_j}{8},\ \gamma\left(X\right) = 1,\ \alpha_0 = -1,\ \alpha_1 = 0.5.$$

$$DGP12.\ \phi\left(X\right) = \frac{\sum_{j=1}^{10} X_j}{10} - X_1 X_2,\ \gamma\left(X\right) = 1,\ \alpha_0 = -1.2,\ \alpha_1 = 0;$$

$$DGP13.\ \phi\left(X\right) = \frac{-\sum_{j=1}^{10} X_j}{10} + \frac{X_1\sum_{k=2}^{5} X_k}{2},\ \gamma\left(X\right) = 1,\ \alpha_0 = 0,\ \alpha_1 = 1.5;$$

$$DGP14.\ \phi\left(X\right) = -\frac{\sum_{j=1}^{10} X_j}{6} + \frac{\sum_{k=1}^{10} X_k^2}{10},\ \gamma\left(X\right) = 1,\ \alpha_0 = 0,\ \alpha_1 = 1.5;$$

$$DGP15.\ \phi\left(X\right) = 0.1\sum_{j=1}^{5} X_j,\ \gamma\left(X\right) = \exp\left(-0.2\sum_{k=1}^{10} X_j\right),\ \alpha_0 = -0.5,\ \alpha_1 = 1.$$

For each of these five DGPs, $X = \left(X_1, X_2, \ldots, X_{10}\right)^\top$ with $\{X_j\}_{j=1}^{10}$ defined as in Section . For each of these DGPs we consider the following potential outcomes:

$$Y\left(0\right) = 1 + X^\top\beta_0 + u\left(0\right),\quad Y\left(1\right) = 2 + X^\top\beta_1 + u\left(1\right),\ \text{and}\quad Y\left(2\right) = 3 + X^\top\beta_2 + u\left(2\right),$$

where $u\left(0\right)$, $u\left(1\right)$ and $u\left(2\right)$ are independent normal random variables with mean zero and variance 1, $\beta_0 = \left(-4, -4, -4, -4, -4, 4, 4, 4, 4, 4\right)^\top$, $\beta_1 = -\beta_0$, and $\beta_2 = 3$. The observed outcome is $Y = 1\left\{D = 0\right\}Y\left(0\right) + 1\left\{D = 1\right\}Y\left(1\right) + 1\left\{D = 2\right\}Y\left(2\right)$, and the true $ATE_{1,0} = 1$, $ATE_{2,0} = 2$, and $ATE_{2,1} = 1$.

For $DGP6 - DGP10$, the H_0 considered is

$$H_0 : \exists\theta_0 = \left(\alpha_0, \alpha_1, \delta_0^\top\right)^\top \in \Theta : P\left(T \leq t|X\right) = \Lambda\left(\alpha_t - X^\top\delta_0\right)\ a.s.\ \text{for}\ t = 0, 1. \tag{B.8}$$

We estimate θ_0 using the ordered logit (proportional odds) likelihood, i.e.,

$$\widehat{\theta}_n = \arg\max_{\left(\alpha_0, \alpha_1, \delta^\top\right)^\top \in \Theta}\sum_{i=1}^{n}\sum_{t=0}^{2}\left[1\left\{T_i = t\right\}\ln\left(\Lambda\left(\alpha_t - X_i^\top\delta\right) - \Lambda\left(\alpha_{t-1} - X_i^\top\delta\right)\right)\right],$$

where, with some abuse of notation, $\alpha_{-1} \equiv -\infty$ and $\alpha_2 \equiv +\infty$. Clearly, $DGP11$ falls under H_0, whereas $DGP12 - DGP15$ fall under H_1, i.e., the negation of (B.8). Note

1 The scaling constant $\pi/\sqrt{3}$ is to make the GPS coefficients comparable to those used in Section , where we adopt a Gaussian link function instead of a logistic link function as we do here.

that DGPs in the ordered treatment setup resemble those in the binary setup described in Section .

In the ordered setup, our proposed test statistic CvM_n^{dpro} is given by (2.11) with $e_i\left(t;\widehat{\theta}_n\right) = 1\left(T_i \leq t\right) - \Lambda\left(\alpha_{n,t} - X_i^\top \delta_n\right)$, and the score function defined accordingly. That is, we use (2.11) with weight $a_n(t) \equiv 1$ for every t and n. Although Sant'Anna and Song (2019) only considered specification tests for binary treatments, we note that their tests can be extended to test (B.8) by using the following test statistic:

$$CvM_n^{ss,o} = CvM_{n,0}^{ss,o} + CvM_{n,1}^{ss,o}, \tag{B.9}$$

such that, for $t = 0, 1$,

$$CvM_{n,t}^{ss,o} = \frac{1}{n}\sum_{i=1}^{n}\left(R_{n,t}^{ss,o}\left(q_t\left(X_i, \widehat{\theta}_n\right)\right)\right)^2,$$

$$q_t\left(X_i, \widehat{\theta}_n\right) = \Lambda\left(\alpha_{n,t} - X_i^\top \delta_n\right),\ e_j(t;\widehat{\theta}_n) = 1\left(T_j \leq t\right) - q_t\left(X_j, \widehat{\theta}_n\right),$$

$$R_{n,t}^{ss,o}(u) \equiv \frac{1}{\sqrt{n}}\sum_{j=1}^{n} e_j(t;\widehat{\theta}_n)\mathcal{P}_{n,t}1\left(q_t\left(X_j, \widehat{\theta}_n\right) \leq u\right),$$

with projection-based weights given by

$$\mathcal{P}_{n,t}1\left(q_t\left(X,\theta\right) \leq u\right) = 1\left(q_t\left(X,\theta\right) \leq u\right) - g_t^\top(X,\theta)\Delta_{n,t}^{-1}\left(\theta\right) G_{n,t}\left(u,\theta\right), \tag{B.10}$$

and $g_t(x,\theta)$, $G_{n,t}(u,\theta)$ and $\Delta_{n,t}\left(\theta\right)$ defined analogously to the binary treatment setup. It is important to emphasize that (B.9) is a test statistic for an *implication* of (B.8), i.e., (B.9) is a test statistic for the following null hypothesis:

$$H_0' : \exists \theta_0 = \left(\alpha_0, \alpha_1, \delta_0^\top\right)^\top \in \Theta : \mathbb{E}\left[e(t;\theta_0)|\Lambda\left(\alpha_t - X^\top \delta_0\right)\right] = 0\ a.s. \text{ for } t = 0, 1,$$

where $e(t;\theta_0) = 1(T \leq t) - \Lambda\left(\alpha_t - X^\top \delta_0\right)$. Thus, although $CvM_{n,t}^{ss,o}$ also avoids the "curse of dimensionality", we cannot ensure its consistency against general nonparametric alternatives H_1, only against H_1', the negation of H_0'.[2] Similarly, the statistic $CvM_n^{im,dpro}$, as a directional-type test, only tests an *implication* of (B.8). The statistic $CvM_{n,t}^{dpro}$ thus is always preferred to.

The simulation results are presented in Table B.3. As it is evident from Table B.3,

---

2 Given the ordered nature of the treatment assignment, we know that the sigma algebra generated by $\left(\Lambda\left(\alpha_0 - X^\top \delta_0\right), \Lambda\left(\alpha_1 - X^\top \delta_0\right)\right)$ is equivalent to the sigma algebra generated by $\Lambda\left(\alpha_t - X^\top \delta_0\right)$, $t = 0, 1$, implying that we can rewrite H_0' as

$$H_0' : \exists \theta_0 = (\alpha_0, \alpha_1, \delta_0^\top)^\top \in \Theta : \mathbb{E}\left[e(t;\theta_0)|\Lambda\left(\alpha_t - X^\top \delta_0\right), \Lambda\left(\alpha_1 - X^\top \delta_0\right)\right] = 0\ a.s. \text{ for } t = 0, 1.$$

Thus, in the ordered treatment setup, augmenting the conditioning set to $\Lambda\left(\alpha_0 - X^\top \delta_0\right), \Lambda\left(\alpha_1 - X^\top \delta_0\right)$ does not affect the power properties of $CvM_n^{ss,o}$.

our proposed double-projected omnibus test, CvM_n^{dpro}, our directional test $CvM_n^{im,dpro}$, and the extension of Sant'Anna and Song (2019)'s test, $CvM_n^{ss,o}$, have finite sample size close to their nominal size. Furthermore, when the GPS is correctly specified, all IPW estimators for the average treatment effects have little to no bias, their RMSE reduces with sample size, and their coverage probability is very close to its nominal level.

Table B.3: Monte Carlo results under designs $DGP6$-$DGP10$: Ordered Treatment

DGP	n	CvM_n^{dpro}	$CvM_n^{im,dpro}$	$CvM_n^{ss,o}$	$Bias_{1,0}$	$Bias_{2,0}$	$RMSE_{1,0}$	$RMSE_{2,0}$	$COV_{1,0}$	$COV_{2,0}$
11	200	0.057	0.065	0.054	-0.090	-0.100	2.362	2.313	0.957	0.960
11	400	0.045	0.047	0.047	-0.038	-0.045	1.682	1.586	0.951	0.951
11	800	0.040	0.058	0.061	-0.025	-0.051	1.172	1.052	0.949	0.950
12	200	0.968	0.130	0.121	0.951	0.821	2.508	1.695	0.935	0.932
12	400	1.000	0.153	0.150	0.794	0.782	1.803	1.280	0.932	0.884
12	800	1.000	0.251	0.248	0.855	0.796	1.417	1.046	0.893	0.812
13	200	0.926	0.244	0.211	1.232	1.604	2.376	2.289	0.905	0.892
13	400	1.000	0.472	0.450	1.297	1.605	1.943	1.929	0.844	0.746
13	800	1.000	0.732	0.713	1.277	1.637	1.629	1.798	0.758	0.432
14	200	0.443	0.168	0.176	0.663	0.768	2.578	2.530	0.946	0.946
14	400	0.907	0.367	0.395	0.596	0.808	1.744	1.817	0.942	0.917
14	800	1.000	0.669	0.693	0.622	0.877	1.313	1.403	0.923	0.885
15	200	0.065	0.085	0.072	-0.162	-4.100	2.203	4.570	0.949	0.497
15	400	0.179	0.167	0.139	-0.113	-4.086	1.471	4.279	0.956	0.134
15	800	0.488	0.398	0.331	-0.146	-4.089	1.013	4.183	0.960	0.006

Note: Simulations based on 1,000 Monte Carlo experiments. "CvM_n^{dpro}" stands for our proposed Cramér-von Mises tests (2.11). "$CvM_n^{im,dpro}$" stands for the directional Cramér-von Mises test in (5.7). "$CvM_n^{ss,o}$" stands for the extension of Sant'Anna and Song (2019)'s test defined in (B.9). Finally, "$Bias_{k,s}$", "$RMSE_{k,s}$" and "$COV_{k,s}$" stand for the average simulated bias, average simulated root mean squared error, and 95% coverage probability for the $ATE_{k,s}$ estimator $ATE_{n,k,s}$ as defined in (B.7). The 95% coverage probability is based on the percentile bootstrap with 499 draws. See the main text for further details.

Under model misspecifications as in $DGP12-DGP15$, the IPW estimators (B.7) are biased for the true average treatment effects and, in general, such biases do not reduce with sample size. In addition, inference procedures for the treatment effects can be unreliable. Thus, detecting GPS misspecifications can prevent misleading inference about the causal effect of interest. As in the binary setup, our proposed test CvM_n^{dpro} performs remarkably well and strictly dominates $CvM_n^{im,dpro}$ and $CvM_n^{ss,o}$ in all considered DGPs. For instance, for $DGP12$ with $n = 200$, CvM_n^{dpro} is eight times more powerful than $CvM_n^{im,dpro}$ or $CvM_n^{ss,o}$; for $DGP13$ with $n = 200$, CvM_n^{dpro} rejects (B.8) more than four times more often than $CvM_n^{im,dpro}$ or $CvM_n^{ss,o}$. As mentioned before, part of this gain of power can be credited to the fact that $CvM_n^{ss,o}$ and $CvM_n^{im,dpro}$ may have trivial power against some directions, and this can have important practical consequences. On the other hand, CvM_n^{dpro} is consistent against all nonparametric (fixed) alternatives, highlighting its potential attractiveness.

Overall, our simulation results highlight that the proposed double-projection-type omnibus tests perform favorably compared to other alternative testing procedures, which include directional tests based on multiple-indexes and tests constructed using one-dimensional GPS $q_t(X, \theta)$. Importantly, the simulations illustrate that our proposed tests are suitable for setups with many covariates and that the gains in power, compared to other alternatives, can be substantial. Given these attractive features, we believe our tests can be useful in practice.

References

Delgado, M. A., and González Manteiga, W. (2001), "Significance testing in nonparametric regression based on the bootstrap," *The Annals of Statistics*, 29(5), 1469–1507.

Escanciano, J. C. (2006), "A consistent diagnostic test for regression models using projections," *Econometric Theory*, 22, 1030–1051.

Hájek, J. (1971), "Discussion of 'An essay on the logical foundations of survey sampling, Part I', by D. Basu," in *Foundations of Statistical Inference*, eds. V. P. Godambe, and D. A. Sprott, Toronto: Holt, Rinehart, and Winston.

Jennrich, R. I. (1969), "Asymptotic properties of non-linear least squares estimators," *The Annals of Mathematical Statistics*, 40(2), 633–643.

Mammen, E. (1993), "Bootstrap and wild bootstrap for high dimensional linear models," *The Annals of Statistics*, 21(1), 255–285.

Newey, W. K., and McFadden, D. (1994), "Large sample estimation and hypothesis testing," in *Handbook of Econometrics*, Vol. 4, Amsterdam: North-Holland: Elsevier, chapter 36, pp. 2111–2245.

Rao, C. R. (1965), *Linear Statistical Inference and its Applications*, New York: Wiley.

Sant'Anna, P. H., and Song, X. (2019), "Specification tests for the propensity score," *Journal of Econometrics*, 210(2), 379–404.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

Stute, W. (1994), "U-statistic processes: a martingale approach," *The Annals of Probability*, pp. 1725–1744.

van der Vaart, A. W., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer.

Wolfowitz, J. (1954), "Generalization of the Theorem of Glivenko-Cantelli," *Annals of Statistics*, 25(1), 131–138.

Yang, S., Imbens, G. W., Cui, Z., Faries, D. E., and Kadziola, Z. (2016), "Propensity score matching and subclassification in observational studies with multi-level treatments," *Biometrics*, 72(4), 1055–1065.