Supplementary material for

"Nonparametric Quantile Regression for Homogeneity Pursuit in Panel Data Models"

Xiaoyu Zhang, Di Wang, Heng Lian and Guodong Li

University of Hong Kong, Shanghai Jiao Tong University and City University of Hong Kong

August 22, 2022

Abstract

This supplementary material gives the technical proofs and additional simulation results. Specifically, Appendix A presents the asymptotic properties of the oracle estimator by providing the proofs of Theorem 1 and some auxiliary lemmas. The oracle property of the SCAD-penalized estimator and the consistency of the SIC are proved in Appendix B. Appendix C gives more information on the data generating process in Section 5.2 and presents additional simulation results.

A Proofs of Theorem 1 and auxiliary lemmas

In this appendix, we present the proof of Theorem 1 and relegate some auxiliary lemmas to the end of this appendix.

We start with some notations. Throughout Appendix A, since we focus on the oracle estimator with a fixed quantile level τ, we omit τ in all notations and simplify $\widehat{\mu}_i^{\text{o}}$ and $\widehat{\boldsymbol{\theta}}^{\text{o}}$ to $\widehat{\mu}$ and $\widehat{\boldsymbol{\theta}}$. let $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \dots, \widehat{\mu}_n)^\top$.

For any $i = 1, 2, \dots, n$, denote $\widetilde{\mathbf{\Pi}}(x_{it}) = (1, \mathbf{\Pi}(x_{it}^\top)^\top)$, $\boldsymbol{\vartheta}_i = (\mu_i, \boldsymbol{\theta}^\top)^\top$ and $\boldsymbol{\vartheta}_{0i} = (\mu_{0i}, \boldsymbol{\theta}_0^\top)^\top$. Let $M_i(\boldsymbol{\vartheta}_i) := T^{-1}\sum_{t=1}^T \rho_\tau(y_{it} - \widetilde{\mathbf{\Pi}}(x_{it})^\top \boldsymbol{\vartheta}_i)$, $\Delta_i^{(1)}(\boldsymbol{\vartheta}_i) := M_i(\boldsymbol{\vartheta}_i) - M_i(\boldsymbol{\vartheta}_{0i})$ and $\Delta_i^{(2)}(\boldsymbol{\vartheta}_i) := T^{-1}\sum_{t=1}^T [(\mu_i - \mu_{0i}) + \mathbf{\Pi}(x_{it})^\top(\boldsymbol{\theta} - \boldsymbol{\theta}_0)](\tau - I\{e_{it} \leq 0\})$. Define $f_i(0) = \mathbb{E}[f_i(0|x_{it})]$, $\boldsymbol{\gamma}_i := f_i(0)^{-1}\mathbb{E}[f_i(0|x_{it})\mathbf{\Pi}(x_{it})]$, and $\mathbf{\Gamma} := n^{-1}\sum_{i=1}^n \mathbb{E}[f_i(0|x_{it})\mathbf{\Pi}(x_{it})(\mathbf{\Pi}(x_{it}) -$

$\boldsymbol{\gamma}_i)^\top]$. Define the score vectors of the quantile regression problem $(nT)^{-1}\sum_{i=1}^n\sum_{t=1}^T \rho_\tau(y_{it}-\mu_i-\boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta})$ as

$$
\begin{aligned}
\mathbb{H}_i^{(1)}(\mu_i,\boldsymbol{\theta}) :=& \frac{1}{T}\sum_{t=1}^T\{\tau - I(y_{it}\le \mu_i+\boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta})\}\\
H_i^{(1)}(\mu_i,\boldsymbol{\theta}) :=& \mathbb{E}[\mathbb{H}_i^{(1)}(\mu_i,\boldsymbol{\theta})]\\
=& \mathbb{E}[\tau - F_i(\mu_i-\mu_{i0}+\boldsymbol{\Pi}(x_{it})(\boldsymbol{\theta}-\boldsymbol{\theta}_0)-R_{it}|x_{it})]\\
\mathbb{H}^{(2)}(\boldsymbol{\mu},\boldsymbol{\theta}) :=& \frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T\{\tau - I(y_{it}\le \mu_i+\boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta})\}\boldsymbol{\Pi}(x_{it})\\
H^{(2)}(\boldsymbol{\mu},\boldsymbol{\theta}) :=& \mathbb{E}[\mathbb{H}^{(2)}(\boldsymbol{\mu},\boldsymbol{\theta})]\\
=& \frac{1}{n}\sum_{i=1}^n \mathbb{E}[\{\tau - F_i(\mu_i-\mu_{i0}+\boldsymbol{\Pi}(x_{it})(\boldsymbol{\theta}-\boldsymbol{\theta}_0)-R_{it}|x_{it})\}\boldsymbol{\Pi}(x_{it})]
\end{aligned}
$$

Proof of Theorem 1. Throughout the proof, to ease the notations, we focus on the case with $K=1$. Since the subgroup structure is completely known when we define the oracle estimator, the results can be directly extended to the general $K>1$ groups. The proof consists of three steps. In the first step, we show the consistency of the oracle estimator $(\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\theta}})$. In the second step, we show the convergence rates of $\max_{1\le i\le n}|\widehat{\mu}_i-\mu_{0i}|$ and $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2$. Finally, we prove the asymptotic normality of $\widehat{\boldsymbol{\theta}}$ and the conditional variance of the estimated smooth function.

Step 1. Consistency of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\theta}}$

We first prove the consistency of $\widehat{\boldsymbol{\theta}}$. For $i=1,\dots,n$, note that

$$
\begin{aligned}
&M_i(\boldsymbol{\vartheta}_i)-M_i(\boldsymbol{\vartheta}_{0i}) = \Delta_i^{(1)}(\boldsymbol{\vartheta}_i)\\
=&\underbrace{\Delta_i^{(1)}(\boldsymbol{\vartheta}_i)+\Delta_i^{(2)}(\boldsymbol{\vartheta}_i)-\mathbb{E}[\Delta_i^{(1)}(\boldsymbol{\vartheta}_i)|\{x_{it}\}]}_{T_{1i}}-\underbrace{\Delta_i^{(2)}(\boldsymbol{\vartheta}_i)}_{T_{2i}}+\underbrace{\mathbb{E}[\Delta_i^{(1)}(\boldsymbol{\vartheta}_i)|\{x_{it}\}]}_{T_{3i}}.
\end{aligned}
$$

Let $\xi_1(n,T)=\sqrt{H\log(n)^2/T}+H^{-d}$. Suppose that $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2\ge L\xi_1(n,T)$ for some constant $L>0$. Then $\widehat{\boldsymbol{\vartheta}}_i=(\widehat{\mu}_i,\widehat{\boldsymbol{\theta}}^\top)^\top$ satisfies $\|\widehat{\boldsymbol{\vartheta}}_i-\boldsymbol{\vartheta}_{0i}\|_2\ge L\xi_1(n,T)$, for all $1\le i\le n$. By Lemmas 2, 4 and 5, we have $\max_{1\le i\le n}T_{1i}=o_p(\xi_1^2(n,T))$, $T_{2i}=L\cdot O_p(\xi_1^2(n,T))$, and $T_{3i}\ge CL^2\xi_1^2(n,T)$, respectively. Hence, for some sufficiently large L, $M_i(\boldsymbol{\vartheta}_i)>M_i(\boldsymbol{\vartheta}_{0i})$ for all $1\le i\le n$. Hence, with probability approaching one, $\sum_{i=1}^n M_i(\boldsymbol{\vartheta}_i)>\sum_{i=1}^n M_i(\boldsymbol{\vartheta}_{0i})$, which however contradicts the definition of $\widehat{\mu}_i$ and $\widehat{\boldsymbol{\theta}}$. Therefore, we conclude that $\widehat{\boldsymbol{\theta}}=\boldsymbol{\theta}_0+O_p(\xi_1(n,T))=\boldsymbol{\theta}_0+o_p(1)$.

Next, we prove the consistency of $\widehat{\mu}_i$, for $i = 1, \ldots, n$. Note that each $\widehat{\mu}_i$ is the minimizer of $M_i((\mu, \widehat{\boldsymbol{\theta}}^\top)^\top)$. Note that

$$
\begin{aligned}
& M_i((\mu_i, \widehat{\boldsymbol{\theta}}^\top)^\top) - M_i((\mu_{0i}, \widehat{\boldsymbol{\theta}}^\top)^\top) = \Delta_i^{(1)}((\mu_i, \widehat{\boldsymbol{\theta}}^\top)^\top) - \Delta_i^{(1)}((\mu_{0i}, \widehat{\boldsymbol{\theta}}^\top)^\top) \\
=& \left[\Delta_i^{(1)}((\mu_i, \widehat{\boldsymbol{\theta}}^\top)^\top) + \Delta_i^{(2)}((\mu_i, \widehat{\boldsymbol{\theta}}^\top)^\top) - \mathbb{E}[\Delta_i^{(1)}((\mu_i, \boldsymbol{\theta}^\top)^\top)|\{x_{it}\}]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}\right] \\
-& \left[\Delta_i^{(1)}((\mu_{0i}, \widehat{\boldsymbol{\theta}}^\top)^\top) + \Delta_i^{(2)}((\mu_{0i}, \widehat{\boldsymbol{\theta}}^\top)^\top) - \mathbb{E}[\Delta_i^{(1)}((\mu_{0i}, \boldsymbol{\theta}^\top)^\top)|\{x_{it}\}]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}\right] \\
+& \left[\mathbb{E}\Delta_i^{(1)}[((\mu_i, \boldsymbol{\theta}^\top)^\top)|\{x_{it}\}]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} - \mathbb{E}[\Delta_i^{(1)}((\mu_i, \boldsymbol{\theta}_0^\top)^\top)|\{x_{it}\}]\right] \\
-& \left[\mathbb{E}[\Delta_i^{(1)}((\mu_{0i}, \boldsymbol{\theta}^\top)^\top)|\{x_{it}\}]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} - \mathbb{E}[\Delta_i^{(1)}((\mu_{0i}, \boldsymbol{\theta}_0^\top)^\top)|\{x_{it}\}]\right] \\
+& \underbrace{\mathbb{E}[\Delta_i^{(1)}((\mu_i, \boldsymbol{\theta}_0^\top)^\top)|\{x_{it}\}] - \Delta_i^{(2)}((\mu_i, \widehat{\boldsymbol{\theta}}^\top)^\top) - \Delta_i^{(2)}((\mu_{0i}, \widehat{\boldsymbol{\theta}}^\top)^\top)}_{T_{4i}}.
\end{aligned}
$$

As $H^2 \log(n)^2/T \to 0$, we consider a positive sequence $\xi_2(n, T)$ such that $\sqrt{H}\xi_1(n, T) = o(\xi_2^2(n, T))$, suppose that $|\widehat{\mu}_i - \mu_{0i}| = L\xi_2(n, T)$. Then, by Lemmas 4 and 5, we have that $\mathbb{E}[\Delta_i^{(1)}((\mu_i, \boldsymbol{\theta}_0^\top)^\top)|\{x_{it}\}] \geq CL^2\xi_2^2(n, T)$, $\Delta_i^{(2)}((\mu_i, \widehat{\boldsymbol{\theta}}^\top)^\top) = O_p(\xi_2^2(n, T))$, and $\Delta_i^{(2)}((\mu_{0i}, \widehat{\boldsymbol{\theta}}^\top)^\top) = O_p(\xi_2^2(n, T))$. For some sufficiently large constant L, we have that $T_{8i} = CL\xi_2^2(n, T)$. Hence

$$
\begin{aligned}
& \mathbb{P}\left(\max_{1\leq i\leq n} |\widehat{\mu}_i - \mu_{0i}| > C\xi_2(n, T)\right) \\
\leq & \mathbb{P}\left(M_i((\mu_i, \widehat{\boldsymbol{\theta}}^\top)^\top) < M_i((\mu_{0i}, \widehat{\boldsymbol{\theta}}^\top)^\top),\ \exists 1 \leq i \leq n,\ \exists |\mu_i - \mu_{0i}| > C\xi_2(n, T)\right) \\
\leq & \mathbb{P}\left(\max_{1\leq i\leq n} \sup_{|\mu-\mu_{0i}|\leq L\xi_2(n,T)} \left|\Delta_i^{(1)}((\mu, \widehat{\boldsymbol{\theta}}^\top)^\top) + \Delta_i^{(2)}((\mu, \widehat{\boldsymbol{\theta}}^\top)^\top) - \mathbb{E}[\Delta_i^{(1)}((\mu, \boldsymbol{\theta}^\top)^\top)|\{x_{it}\}]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}\right| > \frac{T_{8i}}{4}\right) \\
+ & \mathbb{P}\left(\max_{1\leq i\leq n} \sup_{|\mu-\mu_{0i}|\leq L\xi_2(n,T)} \left|\mathbb{E}[\Delta_i^{(1)}((\mu_{0i}, \boldsymbol{\theta}^\top)^\top)|\{x_{it}\}]|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} - \mathbb{E}[\Delta_i^{(1)}((\mu_{0i}, \boldsymbol{\theta}_0^\top)^\top)|\{x_{it}\}]\right| > \frac{T_{8i}}{4}\right) \\
:= & \mathbb{P}(A_1) + \mathbb{P}(A_2).
\end{aligned}
$$

By Lemma 2, we have $\mathbb{P}(A_1) \to 0$, as $T \to \infty$. In addition, since

$$
|\Delta_i^{(1)}((\mu_{0i}, \boldsymbol{\theta}^\top)^\top) - \Delta_i^{(1)}((\mu_{0i}, \boldsymbol{\theta}_0^\top)^\top)| \leq 2\|\widetilde{\boldsymbol{\Pi}}(x_{it})\|_2 \cdot \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2,
$$

it is obtained that $\mathbb{P}(A_2) \to 0$ provided that $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 = O_p(\xi_1(n, T))$, $\|\widetilde{\boldsymbol{\Pi}}(x_{it})\|_2 \leq \sqrt{H}$, and $H^2 \log(n)^2/T \to 0$. Therefore, we prove the consistency of $\widehat{\mu}_1, \ldots, \widehat{\mu}_n, \widehat{\boldsymbol{\theta}}$ under the conditions in Theorem 1.

Step 2. Rate of $\max_{1\le i\le n}|\widehat{\mu}_i-\mu_{0i}|$ *and* $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2$

As $\widehat{\mu}_1,\ldots,\widehat{\mu}_n,\widehat{\boldsymbol{\theta}}$ are consistent, by Lemma 6, we have the following asymptotic representations

$$\begin{aligned}&\widehat{\mu}_i-\mu_{0i}+o_p(|\widehat{\mu}_i-\mu_{0i}|)\\ =&-\boldsymbol{\gamma}_i^\top(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+f_i(0)^{-1}\left\{\mathbb{H}_i^{(1)}(\widehat{\mu}_i,\widehat{\boldsymbol{\theta}})-\mathbb{H}_i^{(1)}(\mu_{0i},\boldsymbol{\theta}_0)-H_i^{(1)}(\widehat{\mu}_i,\widehat{\boldsymbol{\theta}})\right\}\\ &+f_i(0)^{-1}\mathbb{H}_i^{(1)}(\mu_{0i},\boldsymbol{\theta}_0)+O_p(T^{-1}\vee H^{-d}\vee\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2^2),\end{aligned}$$

for all $i=1,\ldots,n$, and

$$\begin{aligned}&\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0+o_p(\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2)\\ =&\boldsymbol{\Gamma}^{-1}\underbrace{\left[-\frac{1}{n}\sum_{i=1}^n\mathbb{H}_i^{(1)}(\mu_{0i},\boldsymbol{\theta}_0)\boldsymbol{\gamma}_i+\mathbb{H}^{(2)}(\boldsymbol{\mu}_0,\boldsymbol{\theta}_0)\right]}_{T_{5i}}\\ &-\boldsymbol{\Gamma}^{-1}\underbrace{\left[\frac{1}{n}\sum_{i=1}^n\left\{\mathbb{H}_i^{(1)}(\widehat{\mu}_i,\widehat{\boldsymbol{\theta}})-\mathbb{H}_i^{(1)}(\mu_{0i},\boldsymbol{\theta}_0)-H_i^{(1)}(\widehat{\mu}_i,\widehat{\boldsymbol{\theta}})\right\}\boldsymbol{\gamma}_i\right]}_{T_{6i}}\\ &+\boldsymbol{\Gamma}^{-1}\underbrace{\left[\mathbb{H}^{(2)}(\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\theta}})-\mathbb{H}^{(2)}(\boldsymbol{\mu}_0,\boldsymbol{\theta}_0)-H^{(2)}(\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\theta}})\right]}_{T_{7i}}\\ &+O_p\left(T^{-1}H^{1/2}\vee H^{-d}\vee\max_{1\le i\le n}|\widehat{\mu}_i-\mu_{0i}|^2\right).\end{aligned}$$

As $\boldsymbol{\gamma}_i\le\sqrt{H}$, $\|T_{5i}\|_2=O_p(\sqrt{H/(nT)})$. Because of the consistency of $(\widehat{\boldsymbol{\mu}},\widehat{\boldsymbol{\theta}})$, by taking $\delta=H^{-1/2}n^{-1/2}T^{-1/3}$ in Lemma 7, $\|T_{6i}\|_2$ and $\|T_{7i}\|_2$ are both $o_p(\sqrt{H/(nT)})$, which implies that

$$\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2=O_p\left(\max_{1\le i\le n}|\widehat{\mu}_i-\mu_{0i}|^2\right)+O_p(\sqrt{H/(nT)}\vee T^{-1}H^{1/2}\vee H^{-d})$$

and

$$\begin{aligned}&\max_{1\le i\le d}|\widehat{\mu}_i-\mu_{0i}|\\ \le&C\left\{\max_{1\le i\le n}|\mathbb{H}_i^{(1)}(\mu_{0i},\boldsymbol{\theta}_0)|+\max_{1\le i\le n}|\mathbb{H}_i^{(1)}(\widehat{\mu}_i,\widehat{\boldsymbol{\theta}})-\mathbb{H}_i^{(1)}(\mu_{0i},\boldsymbol{\theta}_0)-H_i^{(1)}(\widehat{\mu}_i,\widehat{\boldsymbol{\theta}})|\right\}\\ &+O_p(\sqrt{H/(nT)}\vee T^{-1}H^{1/2}\vee H^{-d}).\end{aligned}$$

By taking the union upper bound and Lemma 1,

$$\begin{aligned}&\mathbb{P}\left[\max_{1\le i\le n}|\mathbb{H}_i^{(1)}(\mu_{0i},\boldsymbol{\theta}_0)|\ge C\sqrt{\frac{\log(n)}{T}}\right]\\ \le&\sum_{i=1}^n\mathbb{P}\left[\max_{1\le i\le n}|\mathbb{H}_i^{(1)}(\mu_{0i},\boldsymbol{\theta}_0)|\ge C\sqrt{\frac{\log(n)}{T}}\right]\le 2\exp(-C\log(n)),\end{aligned}$$

which implies that $\max_{1\le i\le n} |\mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0)| = O_p(\sqrt{\log(n)/T})$. Additionally, because of consistency of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\theta}}$, by Lemma 7, for any $\epsilon > 0$,

$$\max_{1\le i\le n} \mathbb{P}\left[|\mathbb{H}_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}}) - \mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0) - H_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}})| > \epsilon\sqrt{\log(n)/T}\right] = o(n^{-1}).$$

Therefore, we have $\max_{1\le i\le n} |\widehat{\mu}_i - \mu_{0i}| = O_p(\sqrt{\log(n)/T} + H^{-d})$ and $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 = O_p(\sqrt{H/(nT)} + (H\log(n)/T)^{3/4} + H^{-d})$. If $Hn^2\log(n)^3/T \to 0$, then $(H\log(n)/T)^{3/4} = o(\sqrt{H/(nT)})$ and $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 = O_p(\sqrt{H/(nT)} + H^{-d})$.

Step 3. Asymptotic normality of $\widehat{\boldsymbol{\theta}}$ and estimated function

Note that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 = O_p(\sqrt{H/(nT)} + H^{-d})$ and

$$\begin{aligned}
&\sum_{t=1}^{T}\sum_{i=1}^{n}\mathbb{E}[\rho_\tau(y_{it} - \widehat{\mu}_i - \boldsymbol{\Pi}(x_{it})^\top\widehat{\boldsymbol{\theta}})|x_{it}] - \sum_{t=1}^{T}\sum_{i=1}^{n}\mathbb{E}[\rho_\tau(y_{it} - \widehat{\mu}_i - \boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}_0)|x_{it}]\\
=&\sum_{t=1}^{T}\sum_{i=1}^{n}\int_{\boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}_0 - m_{it} + \widehat{\mu}_i - \mu_{0i}}^{\boldsymbol{\Pi}(x_{it})^\top\widehat{\boldsymbol{\theta}} - m_{it} + \widehat{\mu}_i - \mu_{0i}} F_i(z|x_{it}) - F_i(0|x_{it})dz\\
=&\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{n} f_k(0|x_{it})[(\boldsymbol{\Pi}(x_{it})^\top(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0))^2 + 2\boldsymbol{\Pi}(x_{it})^\top(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\widetilde{R}_{it}]\\
&+O_p\left(nT[\sqrt{H}(\sqrt{H/(nT)} + H^{-d})]^3\right),
\end{aligned} \tag{A.1}$$

where $\widetilde{R}_{it} = \boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}_0 - m_i(x_{it}) + \widetilde{\mu}_i - \mu_{0i} = O(H^{-d} + \sqrt{\log(n)/T})$. Define

$$\begin{aligned}
\widetilde{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{T}\sum_{i=1}^{n}\Big\{ &- \boldsymbol{\Pi}(x_{it})^\top(\boldsymbol{\theta} - \boldsymbol{\theta}_0)(\tau - I\{e_{it} \le 0\})\\
&+\frac{1}{2}f_i(0|x_{it})[(\boldsymbol{\Pi}(x_{it})^\top(\boldsymbol{\theta} - \boldsymbol{\theta}))^2 + 2\boldsymbol{\Pi}(x_{it})^\top(\boldsymbol{\theta} - \boldsymbol{\theta}_0)\widetilde{R}_{it}]\Big\}.
\end{aligned}$$

We have obviously

$$\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + (\boldsymbol{Z}^\top\boldsymbol{f}\boldsymbol{Z})^{-1}(-\boldsymbol{Z}^\top\boldsymbol{f}\boldsymbol{R} + \boldsymbol{Z}^\top\boldsymbol{\epsilon}),$$

where $\boldsymbol{Z} = [\boldsymbol{\Pi}(x_{11}), \ldots, \boldsymbol{\Pi}(x_{1T}), \boldsymbol{\Pi}(x_{21}), \ldots, \boldsymbol{\Pi}(x_{nT})]^\top$, $\boldsymbol{f} = \mathrm{diag}(f_1(0|x_{11}), \ldots, f_n(0|x_{nT}))$, $\widetilde{\boldsymbol{R}} = (\widetilde{R}_{11}, \ldots, \widetilde{R}_{nT})^\top$, and $\boldsymbol{\epsilon} = ((\tau - I\{e_{11} \le 0\}), \ldots, (\tau - I\{e_{nT} \le 0\}))^\top$.

First consider $\boldsymbol{\Pi}(x)^\top(\boldsymbol{Z}^\top\boldsymbol{f}\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{\epsilon}$. Its conditional asymptotic variance is given by $\tau(1-\tau)\boldsymbol{\Pi}(x)^\top(\boldsymbol{Z}^\top\boldsymbol{f}\boldsymbol{Z})^{-1}(\boldsymbol{Z}^\top\boldsymbol{Z})(\boldsymbol{Z}^\top\boldsymbol{f}\boldsymbol{Z})^{-1}\boldsymbol{\Pi}(x) \asymp H/(nT)$. Using Lindeberg-Feller condition, similar to the proof of Theorem 3.1 of Zhou et al. (1998), and by a central limit theorem for α-mixing sequences, we have

$$\left[\tau(1-\tau)\boldsymbol{\Pi}(x)^\top(\boldsymbol{Z}^\top\boldsymbol{f}\boldsymbol{Z})^{-1}(\boldsymbol{Z}^\top\boldsymbol{Z})(\boldsymbol{Z}^\top\boldsymbol{f}\boldsymbol{Z})^{-1}\boldsymbol{\Pi}(x)\right]^{-\frac{1}{2}}\boldsymbol{\Pi}(x)^\top(\boldsymbol{Z}^\top\boldsymbol{f}\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{\epsilon} \xrightarrow{d} N(0,1).$$

By Lemma 3 and $|R_{it}| = O(H^{-d} + \sqrt{\log(n)/T})$,

$$\mathbf{\Pi}(x)^\top(\boldsymbol{Z}^\top\boldsymbol{f}\boldsymbol{Z})^{-1}\boldsymbol{Z}_k^\top\boldsymbol{R}_k = O_p(\sqrt{H/(nT)}(H^{-d} + \sqrt{\log(n)/T})) = o_p(\sqrt{H/(n_T)}).$$

Thus,

$$\frac{\mathbf{\Pi}(x)^\top(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)}{(\tau(1-\tau)\mathbf{\Pi}(x)^\top(\boldsymbol{Z}^\top\boldsymbol{f}\boldsymbol{Z})^{-1}(\boldsymbol{Z}^\top\boldsymbol{Z})(\boldsymbol{Z}^\top\boldsymbol{f}\boldsymbol{Z})^{-1}\mathbf{\Pi}(x))^{1/2}} \to N(0,1).$$

Denote

$$\begin{aligned} Q(\boldsymbol{\theta}) &= -\sum_{i=1}^n\sum_{t=1}^T[\mathbf{\Pi}(x_{it})^\top(\boldsymbol{\theta}-\boldsymbol{\theta}_0)](\tau - I(e_{it}\le 0)) \\ &+ \sum_{i=1}^n\sum_{t=1}^T\mathbb{E}[\rho_\tau(y_{it}-\widehat{\mu}_i-\mathbf{\Pi}(x_{it})^\top\boldsymbol{\theta})|x_{it}] - \sum_{i=1}^n\sum_{t=1}^T\mathbb{E}[\rho_\tau(y_{it}-\widehat{\mu}_i-\mathbf{\Pi}(x_{it})^\top\boldsymbol{\theta}_0)|x_{it}]. \end{aligned}$$

If $\|\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}\|_2 = \delta\xi(n,T)$ where δ is any positive constant, by a similar argument as Lemma 2 with all information of n individuals combined, we have

$$\begin{aligned} \sup_{\|\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}\|_2\le\delta\xi(n,T)}\Bigg|\sum_{i=1}^n\sum_{t=1}^T\rho_\tau(y_{it}-\widehat{\mu}_i-\mathbf{\Pi}(x_{it})^\top\boldsymbol{\theta}) - \sum_{i=1}^n\sum_{t=1}^T\rho_\tau(y_{it}-\widehat{\mu}_i-\mathbf{\Pi}(x_{it})^\top\widetilde{\boldsymbol{\theta}}) \\ - [Q(\boldsymbol{\theta}) - Q(\widetilde{\boldsymbol{\theta}})]\Bigg| = o_p(nT\xi^2(n,T)). \end{aligned}$$

By comparing $Q(\boldsymbol{\theta})$ with (A.1), $Q(\boldsymbol{\theta})$ is a quadratic function of $\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}$ after ignoring the small term $O_p(nT[\sqrt{H}(\sqrt{H/(nT)}+H^{-d})]^3)$. As $\widetilde{\boldsymbol{\theta}}$ is the minimizer of the quadratic function. When $\|\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}\|_2 = \delta\xi(n,T)$,

$$|Q(\boldsymbol{\theta}) - Q(\widetilde{\boldsymbol{\theta}})| \ge CnT\|\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}\|_2^2 - O_p(nT[\sqrt{H}(\sqrt{H/(nT)}+H^{-d})]^3) \ge CnT\|\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}\|_2^2.$$

Therefore, we have that with probablity approaching one

$$\inf_{\|\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}\|_2=\delta\xi(n,T)}\sum_{i=1}^n\sum_{t=1}^T\left[\rho_\tau(y_{it}-\widehat{\mu}_i-\mathbf{\Pi}(x_{it})^\top\boldsymbol{\theta}) - \rho_\tau(y_{it}-\widehat{\mu}_i-\mathbf{\Pi}(x_{it})^\top\widetilde{\boldsymbol{\theta}})\right] > 0.$$

By the convexity of $\rho_\tau(\cdot)$ function and the definition of $\widehat{\mu}_i$ and $\widehat{\boldsymbol{\theta}}$, this implies that $\|\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}\|_2 = o_p(\xi(n,T))$. Therefore, $\widehat{\boldsymbol{\theta}}$ has the same asymptotic properties as $\widetilde{\boldsymbol{\theta}}$.

Finally, by the B-spline approximation error, if $H(nT)^{-1/(2d+1)} \to \infty$, $|\mathbf{\Pi}(x)^\top\boldsymbol{\theta}_0 - m_i(x)| = o_p(\sqrt{H/nT})$, and the above results imply that

$$\frac{\mathbf{\Pi}(x)^\top\widehat{\boldsymbol{\theta}} - m_i(x)}{(\tau(1-\tau)\mathbf{\Pi}(x)^\top(\boldsymbol{Z}^\top\boldsymbol{f}\boldsymbol{Z})^{-1}(\boldsymbol{Z}^\top\boldsymbol{Z})(\boldsymbol{Z}^\top\boldsymbol{f}\boldsymbol{Z})^{-1}\mathbf{\Pi}(x))^{1/2}} \to N(0,1).$$

□

We state some auxiliary lemmas used for the proof of Theorem 1. The first lemma is the Bernstein-type inequality for the geometrically α-mixing sequence. It is a corollary of Theorem 2.19 in Fan and Yao (2008) by taking $q \asymp n/\log(n)$ in their theorem.

Lemma 1. *Let $\{x_{it}\}$ be a strictly stationary α-mixing process with mean zero and mixing coefficient $\alpha(l) \le r^l$ for some $r \in (0,1)$. Suppose that $\mathbb{E}|x_t|^k \le Ck!A^{k-2}D^2$, $k = 3, 4, \ldots$, then for any $\varepsilon > 0$,*

$$\mathbb{P}\left(\left|\sum_{t=1}^{T} x_t\right| > T\varepsilon\right) \le C\log(T)\exp\left[-C\frac{T}{\log(T)}\frac{\varepsilon^2}{\varepsilon A + D^2}\right].$$

Next, we state some intermediate results in the proof of Theorem 1 in the following lemmas and present their proofs.

Lemma 2. *Let $d(n,T)$ be a sequence depending on n and T such that $d(n,T) \to 0$ as $T \to \infty$ and $\sqrt{H\log(n)^2T^{-1}} + H^{-d} = O(\xi(n,T))$. Under the conditions in Theorem 1,*

$$\max_{1\le i\le n}\sup_{\|\boldsymbol{\vartheta}_i-\boldsymbol{\vartheta}_{0i}\|_2=d(n,T)}\left|\Delta_i^{(1)}(\boldsymbol{\vartheta}_i) + \Delta_i^{(2)}(\boldsymbol{\vartheta}_i) - \mathbb{E}[\Delta_i^{(1)}(\boldsymbol{\vartheta}_i)|\{x_{it}\}]\right| = o_p(d^2(n,T)).$$

Proof of Lemma 2. Let $\mathcal{N}_i = \{\boldsymbol{\vartheta}_i^{(1)}, \ldots, \boldsymbol{\vartheta}_i^{(N)}\}$ be a $\delta(n,T)$ covering of $\{\boldsymbol{\theta} : \|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}\|_2 \le d(n,T)\}$. The size of $\mathcal{N}_{(k)}$ is bounded by $N \le (Cd(n,T)/\delta(n,T))^H$ and thus $\log N \le CH\log(T)$ if we choose $\delta(n,T) \asymp T^{-a}d(n,T)$ for some $a > 0$.

Let $\Delta_{it}(\boldsymbol{\vartheta}_i) = \rho_\tau(y_{it} - \widetilde{\boldsymbol{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_i) - \rho_\tau(y_{it} - \widetilde{\boldsymbol{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_{0i}) + \widetilde{\boldsymbol{\Pi}}(x_{it})^\top(\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i})(\tau - I\{e_{it} \le 0\})$. Using the Lipschitz property of $\rho_\tau(\cdot)$, and that for any $\boldsymbol{\vartheta}_i$, there exists some $\boldsymbol{\vartheta}_i^{(l)}$ such that $\|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_i^{(l)}\|_2 \le \delta(n,T)$, we have

$$\begin{aligned}&\left|\Delta_i^{(1)}(\boldsymbol{\vartheta}_i) + \Delta_i^{(2)}(\boldsymbol{\vartheta}_i) - \mathbb{E}[\Delta_i^{(1)}(\boldsymbol{\vartheta}_i)|\{x_{it}\}] - \Delta_i^{(1)}(\boldsymbol{\vartheta}_i^{(l)}) - \Delta_i^{(2)}(\boldsymbol{\vartheta}_i^{(l)}) + \mathbb{E}[\Delta_i^{(1)}(\boldsymbol{\vartheta}_i^{(l)})|\{x_{it}\}]\right| \\ &= \left|\frac{1}{T}\sum_{t=1}^{T}\Delta_{it}(\boldsymbol{\vartheta}_i) - \mathbb{E}[\Delta_{it}(\boldsymbol{\vartheta}_i)|\{x_{it}\}]\right| \le \frac{C}{T}\sum_{t=1}^{T}|\widetilde{\boldsymbol{\Pi}}(x_{it})^\top(\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_i^{(l)})| = O(\sqrt{H}\delta(n,T)),\end{aligned}$$

which can obviously be made to be $o_p(\xi^2(n,T))$ by setting $\delta(n,T) \asymp T^{-a}\xi(n,T)$ for some a large enough.

Denote $m_{it} = m(x_{it})$. Using that $\rho_\tau(x) = |x|/2 + (\tau - 1/2)x$, by simple algebra,

$$\begin{aligned}|\Delta_{it}(\boldsymbol{\vartheta}_i)| = &\Big|\frac{1}{2}|e_{it} + m_{it} + \mu_{0i} - \widetilde{\boldsymbol{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_i| - \frac{1}{2}|e_{it} + m_{it} + \mu_{0i} - \widetilde{\boldsymbol{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_{0i}| \\ &+ \widetilde{\boldsymbol{\Pi}}(x_{it})^\top(\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i})(1/2 - I\{e_{it} \le 0\})\Big| \\ \le &|\widetilde{\boldsymbol{\Pi}}(x_{it})^\top(\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i})| \cdot I\{|e_{it}| \le |\widetilde{\boldsymbol{\Pi}}(x_{it})^\top(\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i})| + |m_{it} + \mu_{0i} - \widetilde{\boldsymbol{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_{0i}|\}.\end{aligned}$$

Thus, $|\Delta_{it}(\boldsymbol{\vartheta}_i)| \leq C\sqrt{H}\xi(n,T) := A$.

Furthermore, we have

$$
\begin{aligned}
&\mathbb{E}[(\Delta_{it}(\boldsymbol{\vartheta}_i) - \mathbb{E}[\Delta_{it}(\boldsymbol{\vartheta}_i)|\{x_{it}\}])^2] \leq \mathbb{E}|\Delta_{it}(\boldsymbol{\vartheta}_i)|^2 \\
\leq& \mathbb{P}\left\{|e_{it}| \leq |\widetilde{\boldsymbol{\Pi}}(x_{it})^\top(\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i})| + |m_{it} + \mu_{0i} - \widetilde{\boldsymbol{\Pi}}(x_{it})^\top \boldsymbol{\vartheta}_{0i}|\right\} \cdot \mathbb{E}|\widetilde{\boldsymbol{\Pi}}(x_{it})^\top(\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i})|^2 \\
\leq& [C\sqrt{H}d(n,T)] \cdot \mathbb{E}|\widetilde{\boldsymbol{\Pi}}(x_{it})^\top(\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i})|^2 \\
\leq& C\sqrt{H}d^3(n,T) := D^2,
\end{aligned}
$$

where the first factor $C\sqrt{H}d(n,T)$ comes from $\mathbb{P}(|e_{it}| \leq |\widetilde{\boldsymbol{\Pi}}(x_{it})^\top(\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i})| + |m_{it} + \mu_{0i} - \widetilde{\boldsymbol{\Pi}}(x_{it})^\top \boldsymbol{\vartheta}_{0i}|)$ by Assumption (A3).

Using Bernstein's inequality in Lemma 1, together with the union bound, we have that for any $a > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\vartheta}_i \in \mathcal{N}_i} \left|\Delta_i^{(1)}(\boldsymbol{\vartheta}_i) + \Delta_i^{(2)}(\boldsymbol{\vartheta}_i) - \mathbb{E}[\Delta_i^{(1)}(\boldsymbol{\vartheta}_i)|\{x_{it}\}]\right| > a\right) \\
\leq& C(T)^{CH}\log(T)\exp\left[-C\frac{T}{\log(T)}\frac{a^2}{aA + D^2}\right].
\end{aligned}
$$

Letting $a = Cd^2(n,T)$, we have

$$
\begin{aligned}
&\mathbb{P}\left(\max_{1\leq i\leq n}\sup_{\boldsymbol{\vartheta}_i \in \mathcal{N}_i} \left|\Delta_i^{(1)}(\boldsymbol{\vartheta}_i) + \Delta_i^{(2)}(\boldsymbol{\vartheta}_i) - \mathbb{E}[\Delta_i^{(1)}(\boldsymbol{\vartheta}_i)|\{x_{it}\}]\right| > Cd^2(n,T)\right) \\
\leq& Cn(T)^{CH}\log(T)\exp\left[-C\frac{T}{\log(T)}\frac{a^2}{aA + D^2}\right] \\
\leq& Cn(T)^{CH}\log(T)\exp\left[-C\frac{T}{\log(T)}H^{-1/2}\xi(n,T)\right] \\
\leq& \exp\left[\log(n) + CH\log(T) + C\log\log(T) - C\frac{\sqrt{T}\log(n)}{\log(T)} - C\frac{TH^{-d-1/2}}{\log(T)}\right] \to 0.
\end{aligned}
$$

□

Lemma 3. *Under the conditions of Theorem 1, the eigenvalues of* $T^{-1}\sum_{t=1}^T \boldsymbol{\Pi}(x_{it})\boldsymbol{\Pi}(x_{it})^\top$ *and* $T^{-1}\sum_{t=1}^T \widetilde{\boldsymbol{\Pi}}(x_{it})\widetilde{\boldsymbol{\Pi}}(x_{it})^\top$ *are bounded and bounded away from zero uniformly over* $i = 1, \ldots, n$*, with probability approaching one.*

Proof of Lemma 3. We focus on the proof of $T^{-1}\sum_{t=1}^T \boldsymbol{\Pi}(x_{it})\boldsymbol{\Pi}(x_{it})^\top$, since the statement for $T^{-1}\sum_{t=1}^T \widetilde{\boldsymbol{\Pi}}(x_{it})\widetilde{\boldsymbol{\Pi}}(x_{it})^\top$ can be proved in an analogous fashion.

Since $\boldsymbol{\Pi}(x_{it}) = \sqrt{H}\widetilde{\boldsymbol{O}}\boldsymbol{B}(x_{it})$ and the eigenvalues of $\mathbb{E}[\boldsymbol{\Pi}(x_{it})\boldsymbol{\Pi}(x_{it})^\top]$ are bounded away from zero and infinity, the desired statement for $\boldsymbol{\Pi}(x_{it})$ is implied by

$$
\left|\frac{H}{T}\sum_{t=1}^T \boldsymbol{B}_h(x_{it})\boldsymbol{B}_{h'}(x_{it}) - H\mathbb{E}\boldsymbol{B}_h(x_{it})\boldsymbol{B}_{h'}(x_{it})\right| = o_p(1/H)
$$

for all $1 \le h, h' \le H$. Denote $V_{it}^{h,h'} = H\boldsymbol{B}_h(x_{it})\boldsymbol{B}_{h'}(x_{it}) = \boldsymbol{e}_h^\top \sqrt{H}\boldsymbol{B}(x_{it})\sqrt{H}\boldsymbol{B}(x_{it})^\top \boldsymbol{e}_{h'}$, where $\boldsymbol{e}_i$ is the vector whose i-th entry is one and other entries are all zero. We have

$$\mathbb{E}[(V_{it}^{h,h'})^2] \le |e_h^\top \sqrt{H}\boldsymbol{B}(x_{it})|^2 \mathbb{E}[e_{h'}^\top \sqrt{H}\boldsymbol{B}(x_{it})\sqrt{H}\boldsymbol{B}(x_{it})^\top e_{h'}] \le CH,$$

as all eigenvalues of $\mathbb{E}[\sqrt{H}\boldsymbol{B}(x_{it})\sqrt{H}\boldsymbol{B}(x_{it})^\top]$ are bounded. Note that $|V_{it}^{h,h'}| \le |e_h^\top \sqrt{H}\boldsymbol{B}(x_{it})| \cdot |e_{h'}^\top \sqrt{H}\boldsymbol{B}(x_{it})| \le H$. By the α-mixing property of x_{it}, we know that $V_{it}^{h,h'}$ is also α-mixing with mixing coefficients bounded by those of x_{it}. By Lemma 1, for any fixed $\epsilon > 0$,

$$\mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T} V_{it}^{h,h'} - \mathbb{E}V_{it}^{h,h'}\right| \ge \frac{\epsilon}{H}\right) \le C\log(T)\exp\left[-C\frac{T}{\log(T)}\frac{\epsilon^2/H^2}{\epsilon + CH}\right]$$

Taking a union bound for all $1 \le h, h' \le H$, as $H^3\log(T)/T \to 0$,

$$\begin{aligned}&\mathbb{P}\left(\max_{1\le h,h'\le H}\left|\frac{1}{T}\sum_{t=1}^{T} V_{it}^{h,h'} - \mathbb{E}V_{it}^{h,h'}\right| \ge \frac{\epsilon}{H}\right)\\ \le& C\exp\left[2\log(H) + \log\log(T) - \frac{\epsilon^2 T/\log(T)}{CH^3}\right] \to 0.\end{aligned}$$

$\square$

Lemma 4. *For any positive sequence $d(n,T)$ depending on n and T,*

$$\begin{aligned}\inf_{\|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}\|_2 = d(n,T)} &\sum_{t=1}^{T}\mathbb{E}\left[\rho_\tau(y_{it} - \widetilde{\boldsymbol{\Pi}}(x_{it})^\top \boldsymbol{\vartheta}_i)\middle| x_{it}\right] - \sum_{t=1}^{T}\mathbb{E}\left[\rho_\tau(y_{it} - \widetilde{\boldsymbol{\Pi}}(x_{it})^\top \boldsymbol{\vartheta}_{0i})\middle| x_{it}\right]\\ \ge& CTd^2(n,T)\end{aligned}$$

with probability approaching 1.

Proof of Lemma 4. For convenience of notation, denote $m_{it} = m(x_{it})$. Using the Knight's identity, namely $\rho_\tau(x-y) - \rho_\tau(x) = -y(\tau - I(x \le 0)) + \int_0^y (I(x \le t) - I(x \le 0))dt$, and mean value expansion, we have that, for each $1 \le i \le n$ and $\tilde{z} \in [\boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0(k)} -$

$m_{it}, \mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{(k)} + \mu_i - \mu_{0i} - m_{it}]$,

$$
\begin{aligned}
&\sum_{t=1}^{T} \mathbb{E}[\rho_\tau(e_{it} + m_{it} + \mu_{0i} - \widetilde{\mathbf{\Pi}}(x_{it})^\top \boldsymbol{\vartheta}_i)|x_{it}] \\
&- \sum_{t=1}^{T} \mathbb{E}[\rho_\tau(e_{it} + m_{it} + \mu_{0i} - \widetilde{\mathbf{\Pi}}(x_{it})^\top \boldsymbol{\vartheta}_{0i})|x_{it}] \\
=& \sum_{t=1}^{T} \int_{\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i} - m_{it}}^{\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_i + \mu_i - \mu_{0i} - m_{it}} [F_k(z|x_{it}) - F_k(0|x_{it})] dz \\
=& \sum_{t=1}^{T} \int_{\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i} - m_{it}}^{\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_i + \mu_i - \mu_{0i} - m_{it}} \left[z f_k(0|x_{it}) + \frac{z^2}{2} f_i'(\widetilde{z}|x_{it}) \right] dz \\
\geq& \frac{1}{2} \sum_{t=1}^{T} f_i(0|x_{it}) \left[(\widetilde{\mathbf{\Pi}}(x_{it})^\top (\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}))^2 + 2 \widetilde{\mathbf{\Pi}}(x_{it})^\top (\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}) R_{it} \right] \\
&- \frac{\overline{f'}}{6} \sum_{t=1}^{T} |(R_{it} + \widetilde{\mathbf{\Pi}}(x_{it})^\top (\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}))^3 - R_{it}^3|,
\end{aligned}
$$

where $R_{it} = m_{it} - \mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_0$.

By the property of B-splines, we have $|R_{it}| = O(H^{-d})$. By Cauchy's inequality and Lemma 3,

$$
\begin{aligned}
&\sum_{t=1}^{T} \widetilde{\mathbf{\Pi}}(x_{it})^\top (\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}) R_{it} \\
\leq& \left[\sum_{t=1}^{T} \left(\widetilde{\mathbf{\Pi}}(x_{it})^\top (\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}) \right)^2 \right]^{1/2} \left[\sum_{t=1}^{T} R_{it}^2 \right]^{1/2} \\
=& \left[(\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i})^\top \left(\sum_{t=1}^{T} \widetilde{\mathbf{\Pi}}(x_{it}) \widetilde{\mathbf{\Pi}}(x_{it})^\top \right) (\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}) \right]^{1/2} \left[\sum_{t=1}^{T} R_{it}^2 \right]^{1/2} \\
=& C d(n,T) T H^{-d}.
\end{aligned}
$$

By Lemma 3, we have

$$
\sum_{t=1}^{T} \left(\widetilde{\mathbf{\Pi}}(x_{it})^\top (\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}) \right)^2 \asymp T \|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}\|_2^2 = T d^2(n,T).
$$

Since $f_k(0|x_{it}) \geq \underline{f}$, we have that

$$
\sum_{t=1}^{T} |(R_{it} + \widetilde{\mathbf{\Pi}}(x_{it})^\top (\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}))^3 - R_{it}^3| = O_p\left(T[\sqrt{H} d(n,T)]^3 \right) = o_p(d^2(n,T)),
$$

and with probability approaching one,

$$
\sum_{t=1}^{T} \mathbb{E}\left[\rho_\tau(y_{it} - \widetilde{\mathbf{\Pi}}(x_{it})^\top \boldsymbol{\vartheta}_i) \Big| x_{it} \right] - \mathbb{E}\left[\rho_\tau(y_{it} - \widetilde{\mathbf{\Pi}}(x_{it})^\top \boldsymbol{\vartheta}_{0i}) \Big| x_{it} \right] \geq C T d^2(n,T).
$$

□

Lemma 5. *Under the conditions of Theorem 1, for any constant $L > 0$ and any sequence $d(n,T)$ such that $d(n,T) \geq C\sqrt{H/T}$,*

$$\sup_{\|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}\|_2 = Ld(n,T)} \sum_{t=1}^{T} \widetilde{\boldsymbol{\Pi}}(x_{it})^\top (\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i})(\tau - I\{e_{it} \leq 0\}) = L \cdot O_p(Td^2(n,T)).$$

Proof of Lemma 5. The proof is straightforward using that

$$\mathbb{E}\left[\left\|\sum_{t=1}^{T} \widetilde{\Pi}(x_{it})(\tau - I\{e_{it} \leq 0\})\right\|_2^2\right] = O_p(TH).$$

By Markov's inequality, it is easy to check that

$$\begin{aligned}
&\sup_{\|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}\|_2 = Ld(n,T)} \sum_{t=1}^{T} \widetilde{\boldsymbol{\Pi}}(x_{it})^\top (\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i})(\tau - I\{e_{it} \leq 0\}) \\
=& L \cdot O_p(\sqrt{TH} d(n,T)) = L \cdot O_p(Td^2(n,T)).
\end{aligned}$$

□

Lemma 6. *Under the conditions of Theorem 1, we have the following asymptotic representations of the oracle estimator*

$$\begin{aligned}
&\widehat{\mu}_i - \mu_{0i} + o_p(|\widehat{\mu}_i - \mu_{0i}|) \\
=& -\boldsymbol{\gamma}_i^\top (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + f_i(0)^{-1} \left\{\mathbb{H}_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}}) - \mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0) - H_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}})\right\} \\
&+ f_i(0)^{-1} \mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0) + O_p(T^{-1} \vee H^{-d} \vee \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2^2),
\end{aligned}$$

for all $i = 1, 2, \ldots, n$, and

$$\begin{aligned}
&\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 + o_p(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2) \\
=& \boldsymbol{\Gamma}^{-1}\left[-\frac{1}{n}\sum_{i=1}^{n} \mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0)\boldsymbol{\gamma}_i + \mathbb{H}^{(2)}(\boldsymbol{\mu}_0, \boldsymbol{\theta}_0)\right] \\
&- \boldsymbol{\Gamma}^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\left\{\mathbb{H}_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}}) - \mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0) - H_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}})\right\}\boldsymbol{\gamma}_i\right] \\
&+ \boldsymbol{\Gamma}^{-1}\left[\mathbb{H}^{(2)}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}}) - \mathbb{H}^{(2)}(\boldsymbol{\mu}_0, \boldsymbol{\theta}_0) - H^{(2)}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}})\right] \\
&+ O_p\left(T^{-1}H^{1/2} \vee H^{-d} \vee \max_{1 \leq i \leq n} |\widehat{\mu}_i - \mu_{0i}|^2\right).
\end{aligned}$$

Proof of Lemma 6. By the computational property of the QR estimator (Kato et al., 2012), it is shown that $\max_{1 \leq i \leq n} |\mathbb{H}_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}})| = O_p(T^{-1})$. Thus, uniformly over $1 \leq i \leq n$, we have

$$O_p(T^{-1}) = \mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0) + H_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}}) + \left\{\mathbb{H}_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}}) - \mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0) - H_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}})\right\}.$$

Expanding $H_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}})$ around $(\mu_{0i}, \boldsymbol{\theta}_0)$, we have

$$H_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}}) = -f_i(0)(\widehat{\mu}_i - \mu_{0i}) - f_i(0)\boldsymbol{\gamma}_i^\top(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + O_p(H^{-d} \vee \max_{1\le i\le n} |\widehat{\mu}_i - \mu_{0i}|^2 \vee \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2^2),$$

and hence, for all $1 \le i \le n$,

$$\begin{aligned}\widehat{\mu}_i - \mu_{0i} = &-\boldsymbol{\gamma}_i^\top(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + f_i(0)^{-1}\left\{\mathbb{H}_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}}) - \mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0) - H_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}})\right\}\\ &+ f_i(0)^{-1}\mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0) + O_p(T^{-1} \vee H^{-d} \vee \max_{1\le i\le n} |\widehat{\mu}_i - \mu_{0i}|^2 \vee \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2^2).\end{aligned} \tag{A.2}$$

Similarly, we have $\|\mathbb{H}^{(2)}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}})\|_2 = O_p(T^{-1}\max_{1\le i\le n, 1\le t\le T} \|\boldsymbol{\Pi}(x_{it})\|_2) = O_p(T^{-1}H^{1/2})$, and

$$O_p(T^{-1}H^{1/2}) = \mathbb{H}^{(2)}(\boldsymbol{\mu}_0, \boldsymbol{\theta}_0) + H^{(2)}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}}) + \left\{\mathbb{H}^{(2)}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}}) - \mathbb{H}^{(2)}(\boldsymbol{\mu}_0, \boldsymbol{\theta}_0) - H^{(2)}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}})\right\}. \tag{A.3}$$

Expanding $H^{(2)}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}})$ around $(\boldsymbol{\mu}_0, \boldsymbol{\theta}_0)$, we have

$$\begin{aligned}H^{(2)}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}}) = &-\frac{1}{n}\sum_{i=1}^n \mathbb{E}[f_i(0|x_{it})\boldsymbol{\Pi}(x_{it})\boldsymbol{\Pi}(x_{it})^\top](\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\\ &-\frac{1}{n}\sum_{i=1}^n \mathbb{E}[f_i(0|x_{it})\boldsymbol{\Pi}(x_{it})](\widehat{\mu}_i - \mu_{0i}) + o_p(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2) + O_p(\max_{1\le i\le n} |\widehat{\mu}_i - \mu_{0i}|^2).\end{aligned} \tag{A.4}$$

By plugging (A.2) into (A.4), we have

$$\begin{aligned}H^{(2)}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}}) = &-\boldsymbol{\Gamma}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \frac{1}{n}\sum_{i=1}^n \mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0)\boldsymbol{\gamma}_i\\ &-\frac{1}{n}\sum_{i=1}^n \left\{\mathbb{H}_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}}) - \mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0) - H_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}})\right\}\boldsymbol{\gamma}_i\\ &+ o_p(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2) + O_p(T^{-1} \vee H^{-d} \vee \max_{1\le i\le n} |\widehat{\mu}_i - \mu_{0i}|^2).\end{aligned} \tag{A.5}$$

Combining (A.3) and (A.5), we can obtain

$$\begin{aligned}\boldsymbol{\Gamma}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = &-\frac{1}{n}\sum_{i=1}^n \mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0)\boldsymbol{\gamma}_i + \mathbb{H}^{(2)}(\boldsymbol{\mu}_0, \boldsymbol{\theta}_0)\\ &-\frac{1}{n}\sum_{i=1}^n \left\{\mathbb{H}_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}}) - \mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0) - H_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}})\right\}\boldsymbol{\gamma}_i\\ &+ \left\{\mathbb{H}^{(2)}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}}) - \mathbb{H}^{(2)}(\boldsymbol{\mu}_0, \boldsymbol{\theta}_0) - H^{(2)}(\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\theta}})\right\}\\ &+ O_p(T^{-1}H^{1/2} \vee H^{-d} \vee \max_{1\le i\le n} |\widehat{\mu}_i - \mu_{0i}|^2) + o_p(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2),\end{aligned}$$

which completes the proof.

$\square$

Lemma 7. *Take δ such that $\delta\sqrt{H} \to 0$ and $\max_{1\le i\le n} |\widehat{\mu}_i - \mu_{0i}| \vee \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 = O_p(\delta)$. We have*

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\gamma}_i \left\{ \mathbb{H}_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}}) - H_i^{(1)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}}) - \mathbb{H}_i^{(1)}(\mu_{0i}, \boldsymbol{\theta}_0) \right\} \right\|_2 = O_p(\sqrt{H} d(T,\delta) \vee H^{-d})$$

and

$$\left\| \mathbb{H}^{(2)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}}) - H^{(2)}(\widehat{\mu}_i, \widehat{\boldsymbol{\theta}}) - \mathbb{H}^{(2)}(\mu_{0i}, \boldsymbol{\theta}_0) \right\|_2 = O_p(\sqrt{H} d(T,\delta) \vee H^{-d}),$$

where $d(T,\delta) := [T^{-1}|\log(\sqrt{H}\delta)|] \vee [T^{-1/2}H^{1/4}\delta^{1/2}|\log(\sqrt{H}\delta)|^{1/2}]$.

Proof of Lemma 7. We focus on the proof of the first statement since the proof of the second one is analogous. Without loss of generality, we assume that $\mu_{0i} = 0$ and $\boldsymbol{\theta}_0 = \mathbf{0}$. Let $g_{\mu,\boldsymbol{\theta}}(u, \boldsymbol{x}) := I(u \le \mu + \boldsymbol{x}^\top\boldsymbol{\theta}) - I(u \le 0)$ and $\mathcal{G}_\delta := \{g_{\mu,\boldsymbol{\theta}} : |\mu| \le \delta, \|\boldsymbol{\theta}\|_2 \le \delta\}$ and $\xi_{it} = (u_{it}, \boldsymbol{\Pi}(x_{it}))$.

As $|m(x_{it}) - \boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}_{0i}| \le H^{-d}$ and $\|\boldsymbol{\gamma}_i\|_2 \le \sqrt{H}$ over $i = 1, \ldots, n$, it suffices to show

$$\max_{1\le i\le n} \mathbb{E}\left[\frac{1}{T} \sup_{g\in\mathcal{G}_\delta} \left| \sum_{t=1}^{T} \{g(\xi_{it}) - \mathbb{E}g(\xi_{it})\} \right| \right] = O(d(T,\delta)).$$

Denote $\widetilde{\mathcal{G}}_{i,\delta} := \{g - \mathbb{E}[g(\xi_{it})] : g \in \mathcal{G}_\delta\}$. Note that $\widetilde{\mathcal{G}}_{i,\delta}$ is pointwise measurable and each element is bounded by 2. By Lemmas 2.6.15 and 2.6.18 of van der Vaart and Wellner (1996), the class $\mathcal{G}_\infty$ is a VC subgraph class. By Theorem 2.6.7 of van der Vaart and Wellner (1996), there exists a constant $v > 1$ such that the covering number satisfies $N(\widetilde{\mathcal{G}}_{i,\delta}, L_2(Q), 2\epsilon) \le C\epsilon^{-v}$ for any $0 < \epsilon < 1$ and any probability measure Q on $\mathbb{R}^H$. In addition, as $\mathbb{E}[g_{\mu,\boldsymbol{\theta}}(\xi_{it})^2] = \mathbb{E}[|F_i(\mu + \boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}|x_{it}) - F_i(0|x_{it})] \le C(|\mu| + \sqrt{H}\|\boldsymbol{\theta}\|_2) \le C\sqrt{H}\delta$. By the Bernstein-type inequality for bounded empirical process, e.g., Proposition B.1 in Kato et al. (2012), we obtain the desired result.

□

B Proofs of Theorems 2 and 3

In this appendix, we present the proofs of Theorems 2 and 3 and relegate some auxiliary lemmas to the end of this appendix. For the brevity of notation, we simplity $\sum_{i=1}^n \sum_{t=1}^T$ to $\sum_{i,t}$.

Proof of Theorem 2. We define the oracle estimator to be that obtained from (5) assuming the groups are known and thus $\widehat{\boldsymbol{\theta}}^{\mathrm{o}}_{(k)}$ is obtained from only observations in G_k, separately for different groups. Similarly to the proof of Theorem 2, we denote $\xi(n,T) = \sqrt{H/(nT)} + H^{-d}$. It suffices to show that with probability approaching one, the oracle estimator is a local minimizer of the SCAD-penalized quantile regression (6).

Considering any $\boldsymbol{\theta}_i$ with $\|\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}^{\mathrm{o}}_i\|_2 \le c$ for all $1 \le i \le n$, with c sufficiently small, specifically $c = o(\lambda)$, and $(\mu_1, \ldots, \mu_n)$ with $\max_{1\le i\le n} |\mu_i - \widehat{\mu}^{\mathrm{o}}_i| \le d$, with d sufficiently small. We only need to show that uniformly over $\boldsymbol{\theta}_c := \{\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_n^\top)^\top : \|\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}^{\mathrm{o}}_i\|_2 \le c,\ \forall i\}$ and $\boldsymbol{\mu}_d := \{(\mu_1, \ldots, \mu_n)^\top : \max_{1\le i\le n} |\mu_i - \widehat{\mu}_i| \le d\}$

$$\begin{aligned}
&\frac{1}{nT}\sum_{i,t}\rho_\tau(y_{it} - \mu_i - \boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}_i) + \binom{n}{2}^{-1}\sum_{i<j} p_\lambda(\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_2)\\
\ge &\frac{1}{nT}\sum_{i,t}\rho_\tau(y_{it} - \widehat{\mu}^{\mathrm{o}}_i - \boldsymbol{\Pi}(x_{it})^\top\widehat{\boldsymbol{\theta}}^{\mathrm{o}}_i) + \binom{n}{2}^{-1}\sum_{i<j} p_\lambda(\|\widehat{\boldsymbol{\theta}}^{\mathrm{o}}_i - \widehat{\boldsymbol{\theta}}^{\mathrm{o}}_j\|_2).
\end{aligned}$$

Let $g_i = k$ if $i \in G_k$. That is, g_i is an indicator on the individual i's group identity. Let $\boldsymbol{O} := \{\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \ldots, \boldsymbol{\theta}_n^\top)^\top : \boldsymbol{\theta}_i = \boldsymbol{\theta}_j \text{ if } g_i = g_j\}$. That is, $\boldsymbol{O}$ consists of all coefficients that satisfy the group partition structure. For ease of presentation, define the mapping $\boldsymbol{\Gamma} : \mathbb{R}^{Hn} \to \boldsymbol{O}$ with $\Gamma(\boldsymbol{\theta}) = (\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_n^*)$, where $\boldsymbol{\theta}_i^* = \sum_{j:g_j=g_i} \boldsymbol{\theta}_i/|G_{g_i}|$. In other words, $\boldsymbol{\Gamma}$ can be the projected value of $\boldsymbol{\theta}$ to the space $\boldsymbol{O}$.

The proof of the displayed equation above can be achieved by the following two steps.

(a)

$$\begin{aligned}
&\inf_{\boldsymbol{\theta}^*=\boldsymbol{\Gamma}(\boldsymbol{\theta}),\boldsymbol{\theta}\in\boldsymbol{\theta}_c,\boldsymbol{\mu}\in\boldsymbol{\mu}_d} \frac{1}{nT}\sum_{i,t}\rho_\tau(y_{it} - \mu_i - \boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}^*_i) + \binom{n}{2}^{-1}\sum_{i<j} p_\lambda(\|\boldsymbol{\theta}^*_i - \boldsymbol{\theta}^*_j\|_2)\\
\ge &\frac{1}{nT}\sum_{i,t}\rho_\tau(y_{it} - \widehat{\mu}^{\mathrm{o}}_i - \boldsymbol{\Pi}(x_{it})^\top\widehat{\boldsymbol{\theta}}^{\mathrm{o}}_i) + \binom{n}{2}^{-1}\sum_{i<j} p_\lambda(\|\widehat{\boldsymbol{\theta}}^{\mathrm{o}}_i - \widehat{\boldsymbol{\theta}}^{\mathrm{o}}_j\|_2).
\end{aligned}$$

(b)

$$\begin{aligned}
&\inf_{\boldsymbol{\theta}^*=\boldsymbol{\Gamma}(\boldsymbol{\theta}),\boldsymbol{\theta}\in\boldsymbol{\theta}_c,\boldsymbol{\mu}\in\boldsymbol{\mu}_d} \frac{1}{nT}\sum_{i,t}\rho_\tau(y_{it} - \mu_i - \boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}_i) + \binom{n}{2}^{-1}\sum_{i<j} p_\lambda(\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_2)\\
-&\frac{1}{nT}\sum_{i,t}\rho_\tau(y_{it} - \mu_i - \boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}^*_i) - \binom{n}{2}^{-1}\sum_{i<j} p_\lambda(\|\boldsymbol{\theta}^*_i - \boldsymbol{\theta}^*_j\|_2) \ge 0.
\end{aligned}$$

For (a), by the definition of the local minimizer which minimizes the check loss subject to the grouping constraint, we have

$$\inf_{\boldsymbol{\theta}^*=\boldsymbol{\Gamma}(\boldsymbol{\theta}),\boldsymbol{\theta}\in\boldsymbol{\theta}_c,\boldsymbol{\mu}\in\boldsymbol{\mu}_d} \frac{1}{nT}\sum_{i,t}\rho_\tau(y_{i,t} - \mu_i - \boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}^*_i) \ge \frac{1}{nT}\sum_{i,t}\rho_\tau(y_{it} - \widehat{\mu}^{\mathrm{o}}_i - \boldsymbol{\Pi}(x_{it})^\top\widehat{\boldsymbol{\theta}}^{\mathrm{o}}_i). \quad (B.1)$$

If $g_i \neq g_j$, by our assumptions, we have $\lambda = o(\|\boldsymbol{\theta}_{0i} - \boldsymbol{\theta}_{0j}\|_2)$ and $\xi(n, T) = o(\lambda)$. Thus,

$$\begin{aligned} &\|\widehat{\boldsymbol{\theta}}_i^{\mathrm{o}} - \widehat{\boldsymbol{\theta}}_j^{\mathrm{o}}\|_2 \geq \|\boldsymbol{\theta}_{0i} - \boldsymbol{\theta}_{0j}\|_2 - \|\widehat{\boldsymbol{\theta}}_i^{\mathrm{o}} - \boldsymbol{\theta}_{0i}\|_2 - \|\widehat{\boldsymbol{\theta}}_j^{\mathrm{o}} - \boldsymbol{\theta}_{0j}\|_2 \\ \geq &3a\lambda - o_p(\lambda) \geq 2a\lambda. \end{aligned} \tag{B.2}$$

In addition,

$$\|\boldsymbol{\theta}_i^* - \widehat{\boldsymbol{\theta}}_i^{\mathrm{o}}\|_2 = \left\| \sum_{k: g_k = g_i} \boldsymbol{\theta}_k / |G_k| - \widehat{\boldsymbol{\theta}}_i^{\mathrm{o}} \right\|_2 \leq \max_{k: g_k = g_i} \|\boldsymbol{\theta}_k - \widehat{\boldsymbol{\theta}}_i^{\mathrm{o}}\|_2 \leq c,$$

which implies that

$$\|\boldsymbol{\theta}_i^* - \boldsymbol{\theta}_j^*\|_2 \geq \|\widehat{\boldsymbol{\theta}}_i^{\mathrm{o}} - \widehat{\boldsymbol{\theta}}_j^{\mathrm{o}}\|_2 - \|\boldsymbol{\theta}_i^* - \widehat{\boldsymbol{\theta}}_i^{\mathrm{o}}\|_2 - \|\boldsymbol{\theta}_j^* - \widehat{\boldsymbol{\theta}}_j^{\mathrm{o}}\|_2 \geq 2a\lambda - 2c \geq a\lambda.$$

Thus, by the definition of SCAD penalty function,

$$p_\lambda(\|\boldsymbol{\theta}_i^* - \boldsymbol{\theta}_j^*\|_2) = p_\lambda(\|\widehat{\boldsymbol{\theta}}_i^{\mathrm{o}} - \widehat{\boldsymbol{\theta}}_j^{\mathrm{o}}\|_2) = \frac{(a+1)\lambda^2}{2}, \quad \text{if } g_i \neq g_j. \tag{B.3}$$

On the other hand, if $g_i = g_j$, then $\widehat{\boldsymbol{\theta}}_i^{\mathrm{o}} = \widehat{\boldsymbol{\theta}}_j^{\mathrm{o}}$ and $\boldsymbol{\theta}_i^* = \boldsymbol{\theta}_j^*$ and thus we have

$$p_\lambda(\|\boldsymbol{\theta}_i^* - \boldsymbol{\theta}_j^*\|_2) = p_\lambda(\|\widehat{\boldsymbol{\theta}}_i^{\mathrm{o}} - \widehat{\boldsymbol{\theta}}_j^{\mathrm{o}}\|_2) = 0, \quad \text{if } g_i = g_j. \tag{B.4}$$

Combining these two cases (B.3) and (B.4), as well as (B.1), we proved (a).

In the rest of the proof we will show (b). Using the convexity of the check loss function, we have $\rho_\tau(x) - \rho_\tau(y) \geq (\tau - I\{y \leq 0\})(x - y)$. Thus for the difference of the loss terms, we have

$$\begin{aligned} &\sum_{t=1}^T \rho_\tau(y_{it} - \mu_i - \boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i) - \sum_{t=1}^T \rho_\tau(y_{it} - \mu_i - \boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i^*) \\ \geq &- \sum_{t=1}^T (\tau - 1\{y_{it} \leq \boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i + \mu_i\}) \boldsymbol{\Pi}(x_{it})^\top (\boldsymbol{\theta}_i - \boldsymbol{\theta}_i^*) \\ = &- \sum_{t=1}^T (\tau - 1\{e_{it} \leq 0\}) \boldsymbol{\Pi}(x_{it})^\top (\boldsymbol{\theta}_i - \boldsymbol{\theta}_i^*) \\ &- \sum_{t=1}^T (1\{e_{it} \leq 0\} - 1\{e_{it} \leq \boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it} + \mu_i - \mu_{0i}\}) \boldsymbol{\Pi}(x_{it})^\top (\boldsymbol{\theta}_i - \boldsymbol{\theta}_i^*). \end{aligned} \tag{B.5}$$

For the first term, using Bernstein's inequality in Lemma 1 in Appendix A, we have

$$\max_{1 \leq h \leq H, 1 \leq i \leq n} \sum_{t=1}^T (\tau - 1\{e_{it} \leq 0\}) \boldsymbol{\Pi}_h(x_{it}) = O_p(\sqrt{T \log(T) \log(nH \log(T))})$$

and thus, $\max_{1\le i\le n}\left\|\sum_{t=1}^{T}(\tau - 1\{e_{it}\le 0\})\boldsymbol{\Pi}(x_{it})\right\|_2 = O_p(\sqrt{TH\log(T)\log(nH\log(T))})$.

By Lemma 8, for sufficiently small c and d, we have

$$
\begin{aligned}
&\sup_{\substack{1\le i\le n, \|\boldsymbol{\theta}_i-\widehat{\boldsymbol{\theta}}_i^{\mathrm{o}}\|_2\le c\\ |\mu_i-\widehat{\mu}_i^{\mathrm{o}}|\le d}}\left\|\sum_{t=1}^{T}(1\{e_{it}\le 0\} - 1\{e_{it}\le \boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}_i - m_{it} + \mu_i - \mu_{0i}\})\boldsymbol{\Pi}(x_{it})\right\|_2\\
\le&\sup_{\substack{1\le i\le n, \|\boldsymbol{\theta}_i-\widehat{\boldsymbol{\theta}}_i^{\mathrm{o}}\|_2\le c\\ |\mu_i-\widehat{\mu}_i^{\mathrm{o}}|\le d}}\Bigg\|\sum_{t=1}^{T}(1\{e_{it}\le 0\} - 1\{e_{it}\le \boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}_i - m_{it} + \mu_i - \mu_{0i}\}\\
&\qquad + F(\boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}_i - m_{it} + \mu_i - \mu_{0i}) - F(0))\boldsymbol{\Pi}(x_{it})\Bigg\|_2\\
+&\sup_{\substack{1\le i\le n, \|\boldsymbol{\theta}_i-\widehat{\boldsymbol{\theta}}_i^{\mathrm{o}}\|_2\le c\\ |\mu_i-\widehat{\mu}_i^{\mathrm{o}}|\le d}}\left\|\sum_{t=1}^{T}(F(\boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}_i - m_{it} + \mu_i - \mu_{0i}) - F(0))\boldsymbol{\Pi}(x_{it})\right\|_2\\
=&O_p(H^{3/2}T^{1/2}\log(T)\log(nT)) + O_p(T\sqrt{H}\xi(n,T)) = O_p(T\sqrt{H}\xi(n,T)).
\end{aligned}
$$

We denote

$$
\begin{aligned}
\boldsymbol{w}_i =& -\sum_{t=1}^{T}(\tau - 1\{e_{it}\le 0\})\boldsymbol{\Pi}(x_{it})\\
&-\sum_{t=1}^{T}(1\{e_{it}\le 0\} - 1\{e_{it}\le \boldsymbol{\Pi}(x_{it})^\top\boldsymbol{\theta}_i - m_{it} + \mu_i - \mu_{0i}\})\boldsymbol{\Pi}(x_{it}).
\end{aligned}
$$

Then, the last line in (B.5), after summing over i, can be written as

$$
\begin{aligned}
&\frac{1}{nT}\sum_{i=1}^{n}\boldsymbol{w}_i^\top(\boldsymbol{\theta}_i - \boldsymbol{\theta}_i^*) = \frac{1}{nT}\sum_{i=1}^{n}\boldsymbol{w}_i^\top(\boldsymbol{\theta}_i - \sum_{j:g_j=g_i}\boldsymbol{\theta}_j/|G_{g_i}|)\\
=&\frac{1}{nT}\sum_{i=1}^{n}\sum_{j:g_j=g_i}\frac{\boldsymbol{w}_i^\top(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j)}{|G_{g_i}|} = \frac{1}{nT}\sum_{(i,j):i<j \text{ and } g_i=g_j}\frac{(\boldsymbol{w}_i - \boldsymbol{w}_j)^\top(\boldsymbol{\theta}_i - \boldsymbol{\theta}_j)}{|G_{g_i}|}\\
=&O_p(n^{-2}\sqrt{H}\xi(n,T) + n^{-2}\sqrt{(H/T)\log(T)\log(nH\log(T))}) \times \left(\sum_{i<j \text{ and } g_i=g_j}\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|\right).
\end{aligned}
$$

When $g_i\ne g_j$, by (B.2),

$$
\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_2 \ge \|\widehat{\boldsymbol{\theta}}_i^{\mathrm{o}} - \widehat{\boldsymbol{\theta}}_j^{\mathrm{o}}\|_2 - \|\widehat{\boldsymbol{\theta}}_i - \widehat{\boldsymbol{\theta}}_i^{\mathrm{o}}\|_2 - \|\widehat{\boldsymbol{\theta}}_j - \widehat{\boldsymbol{\theta}}_j^{\mathrm{o}}\|_2 \ge \|\widehat{\boldsymbol{\theta}}_i^{\mathrm{o}} - \widehat{\boldsymbol{\theta}}_j^{\mathrm{o}}\|_2 - 2c \ge a\lambda.
$$

Also, $\|\boldsymbol{\theta}_i^* - \boldsymbol{\theta}_j^*\|_2 \geq a\lambda$, when $g_i \neq g_j$. So the difference of penalty terms of (b) is

$$
\begin{aligned}
&\binom{n}{2}^{-1}\sum_{i<j} p_\lambda(\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_2) - \binom{n}{2}^{-1}\sum_{i<j} p_\lambda(\|\boldsymbol{\theta}_i^* - \boldsymbol{\theta}_j^*\|_2)\\
=&\binom{n}{2}^{-1}\sum_{i<j, g_i\neq g_j} [p_\lambda(\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_2) - p_\lambda(\|\boldsymbol{\theta}_i^* - \boldsymbol{\theta}_j^*\|_2)]\\
+&\binom{n}{2}^{-1}\sum_{i<j, g_i= g_j} [p_\lambda(\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_2) - p_\lambda(\|\boldsymbol{\theta}_i^* - \boldsymbol{\theta}_j^*\|_2)]\\
=&\binom{n}{2}^{-1}\sum_{i<j, g_i= g_j} [p_\lambda(\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_2) - p_\lambda(\|\boldsymbol{\theta}_i^* - \boldsymbol{\theta}_j^*\|_2)].
\end{aligned}
$$

When $g_i = g_j$, we have $\boldsymbol{\theta}_i^* = \boldsymbol{\theta}_j^*$. Furthermore,

$$
\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_2 \leq \|\widehat{\boldsymbol{\theta}}_i^{\mathrm{o}} - \widehat{\boldsymbol{\theta}}_j^{\mathrm{o}}\|_2 + \|\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_i^{\mathrm{o}}\|_2 + \|\boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_j^{\mathrm{o}}\|_2 \leq 2c \leq \lambda,
$$

and since the SCAD penalty $p_\lambda(x) = \lambda x$ when $x \in [0, \lambda]$, we have

$$
\begin{aligned}
&\binom{n}{2}^{-1}\sum_{i<j} p_\lambda(\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_2) - \binom{n}{2}^{-1}\sum_{i<j} p_\lambda(\|\boldsymbol{\theta}_i^* - \boldsymbol{\theta}_j^*\|_2)\\
\geq&\binom{n}{2}^{-1}\sum_{i<j, g_i= g_j} p_\lambda(\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_2) = \binom{n}{2}^{-1}\lambda\sum_{i<j, g_i= g_j} \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_j\|_2.
\end{aligned}
$$

Thus, by our assumption, the difference of the penalties is positive and dominant in the left hand side of (b), which implies that (b) holds.

□

Proof of Theorem 3. In the proof, we denote the true number of groups by K_0 and the true partition is $\mathcal{G}_0 = (G_{01}, \ldots, G_{0n})$ with true group indicators $g_{0i} = k$ if $i \in G_k$. Let $\mathcal{G} = \{G_1, \ldots, G_K\}$ be any partition for $\{1, \ldots, n\}$ with K groups, with group indicators $g_i, i = 1, \ldots, n$. Define

$$
\boldsymbol{\theta}^{\mathcal{G}_0} = \{\boldsymbol{\theta}_{01}^{\mathcal{G}_0}, \ldots, \boldsymbol{\theta}_{0n}^{\mathcal{G}_0}\} = \min_{\boldsymbol{\theta}_i = \boldsymbol{\theta}_j \text{ if } g_i = g_j} \mathbb{E}\left[\sum_{i=1}^n \sum_{t=1}^T \rho_\tau(y_{it} - \mu_{0i} - \boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i)\right].
$$

We note that $\boldsymbol{\theta}_{0i}^{\mathcal{G}_0}$ under the true partition is different from the $\boldsymbol{\theta}_{0i}$ we defined previously, as the minimizer of $\mathbb{E}[f_{(k)}(0|x_{it})(\boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it})^2]$, where $f_{(k)}(\cdot|x_{it})$ is the average conditional density function for the group G_k. However, we first show that they are close enough. By Knight's identity, for any $i \in G_k$,

$$
\begin{aligned}
&\rho_\tau(y_{it} - \mu_{0i} - \boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i) - \rho_\tau(y_{it} - \mu_{0i} - m_{it})\\
=&(\boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it})[I\{e_{it} \leq 0\} - \tau] + \int_0^{\boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it}} [I\{e_{it} \leq u\} - I\{e_{it} \leq 0\}]du.
\end{aligned}
$$

Hence, as $f_i(0|x_{it}) = f_{(k)}(0|x_{it})$ for all $i \in G_k$,

$$
\begin{aligned}
&\mathbb{E}[\rho_\tau(y_{it} - \mu_{0i} - \mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_i)] - \mathbb{E}[\rho_\tau(y_{it} - \mu_{0i} - m_{it})] \\
=&\mathbb{E}\left\{\int_0^{\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta} - m_{it}} F_i[u|x_{it}] - F[0|x_{it}] du\right\} \\
=&\mathbb{E}\left[\frac{1}{2} f_{(k)}(0|x_{it})[\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it}]^2\right] + O(\mathbb{E}[|\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it}|^3]).
\end{aligned}
$$

So for $\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_{0i}\|_2 \le M_n H^{-(2/3)d}$ with $M_n \to \infty$ arbitrarily slowly, we have $\mathbb{E}[|\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it}|^3] \le \mathbb{E}[(|\mathbf{\Pi}(x_{it})^\top (\boldsymbol{\theta}_i - \boldsymbol{\theta}_{0i})| + |\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i} - m_{it}|)^3] = O(M_n^3 H^{-(9/2)d+3/2} + H^{-3d}) = O(H^{-3d})$ and thus

$$
\begin{aligned}
&\mathbb{E}[\rho_\tau(y_{it} - \mu_{0i} - \mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_i)] - \mathbb{E}[\rho_\tau(y_{it} - \mu_{0i} - m_{it})] \\
=&\mathbb{E}\left[\frac{1}{2} f_{(k)}(0|x_{it})[\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it}]^2\right] + O(H^{-3d}).
\end{aligned}
$$

For $\|\boldsymbol{\theta}_i - \boldsymbol{\theta}_{0i}\|_2 = M_n H^{-(2/3)d}$, we have

$$
\begin{aligned}
&\mathbb{E}[\rho_\tau(y_{it} - \mu_{0i} - \mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{(k)})] - \mathbb{E}[\rho_\tau(y_{it} - \mu_{0i} - \mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i})] \\
=&\mathbb{E}\left[\frac{1}{2} f_{(k)}(0|x_{it})[\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it}]^2\right] - \mathbb{E}\left[\frac{1}{2} f_{(k)}(0|x_{it})[\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i} - m_{it}^2]\right] + O(H^{-3d}) \\
=&\mathbb{E}\left[\frac{1}{2} f_{(k)}(0|x_{it})[\mathbf{\Pi}(x_{it})^\top (\boldsymbol{\theta}_i - \boldsymbol{\theta}_{0i})]^2\right] + \mathbb{E}\left[\frac{1}{2} f_{(k)}(0|x_{it})\mathbf{\Pi}(x_{it})^\top (\boldsymbol{\theta}_i - \boldsymbol{\theta}_{0i})(\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i} - m_{it})\right] \\
&+ O(H^{-3d}) \\
=&\mathbb{E}[\frac{1}{2} f_{(k)}(0|x_{it})[\mathbf{\Pi}(x_{it})^\top (\boldsymbol{\theta}_i - \boldsymbol{\theta}_{0i})]^2] + O(H^{-3d}) \\
\ge& C M_n H^{-3d} - O(H^{-3d}) > 0,
\end{aligned}
$$

where the third equality above results from $\boldsymbol{\theta}_{0i}$ minimizes $\mathbb{E}[\frac{1}{2} f_{(k)}(0|x_{it})(\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it})^2]$, which means $\mathbb{E}[f_{(k)}(0|x_{it})\mathbf{\Pi}(x_{it})^\top (\boldsymbol{\theta}_i - \boldsymbol{\theta}_{0i})(\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i} - m_{it})] = 0$. This means $\|\boldsymbol{\theta}_{0i}^{\mathcal{G}_0} - \boldsymbol{\theta}_{0i}\|_2 \le M_n H^{-3d} \le H^{-d-1/2}$ and thus $\boldsymbol{\theta}_{0i}^{\mathcal{G}_0}$ still satisfies the approximation property $\sup_x |\mathbf{\Pi}(x)^\top \boldsymbol{\theta}_{0i}^{\mathcal{G}_0} - m_i(x)| \le C H^{-d}$.

Case 1. ($K < K_0$, under-fitted model)

In this case, let $\mathcal{G}$ be the partition that minimizes

$$
\min_{\boldsymbol{\theta}_i = \boldsymbol{\theta}_j,\ \text{if}\ G_i = G_j, |\mathcal{G}| = K} \mathbb{E}\left[\sum_{i=1}^n \sum_{t=1}^T \rho_\tau(y_{it} - \mu_{0i} - \mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_i)\right].
$$

By definition, it is obvious that if i, j belongs to the same group in the true partition $\mathcal{G}_0$ so that the distribution of (y_{it}, x_{it}) and (y_{jt}, x_{jt}) are the same, they are still in the same group in the partition $\mathcal{G}$. In other words, $\mathcal{G}$ is formed by combining some groups in $\mathcal{G}_0$. In particular, given K_0 is fixed, there are only a fixed number of such possible partitions $\mathcal{G}$.

Suppose G_{0k}, $G_{0k'}$ are combined into $G_{k''}$, then

$$\sum_{i\in G_{k''}} \mathbb{E}[\rho_\tau(y_{it} - \mu_{0i} - \mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i}^{\mathcal{G}})] - \sum_{i\in G_{k''}} \mathbb{E}[\rho_\tau(y_{it} - \mu_{0i} - \mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i}^{\mathcal{G}_0})]$$
$$= \sum_{i\in G_{k''}} \mathbb{E}\int_{\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i}^{\mathcal{G}_0} - m_{it}}^{\mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i}^{\mathcal{G}} - m_{it}} [I\{e_{it} \le u\} - I\{e_{it} \le 0\}]du \ge \sum_{i\in G_{k''}} \|\boldsymbol{\theta}_{0i}^{\mathcal{G}} - \boldsymbol{\theta}_{0i}^{\mathcal{G}_0}\|_2^2 \ge Cn_{k''}\rho$$

where at least one of the distance $\|\boldsymbol{\theta}_{0i}^{\mathcal{G}} - \boldsymbol{\theta}_{0i}^{\mathcal{G}_0}\|_2$ for $i \in G_{0k}$ and $\|\boldsymbol{\theta}_{0i}^{\mathcal{G}} - \boldsymbol{\theta}_{0i}^{\mathcal{G}_0}\|_2$ for $i \in G_{0k'}$ is larger than, say $\rho/2$. By summing over different groups, we get

$$\sum_{i=1}^n \mathbb{E}[\rho_\tau(y_{it} - \mu_{0i} - \mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i}^{\mathcal{G}})] - \sum_{i=1}^n \mathbb{E}[\rho_\tau(y_{it} - \mu_{0i} - \mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i}^{\mathcal{G}_0})] \ge Cn\rho.$$

By following the proof of Theorem 1, in particular Lemma 2, we can show that $\|\widehat{\boldsymbol{\vartheta}}_i^{\mathcal{G}} - \boldsymbol{\vartheta}_{0i}^{\mathcal{G}_0}\|_2 = O_p(\sqrt{H/(nT)} + H^{-d})$. Similarly to Lemma 4, we have $\mathbb{E}[\sum_{t=1}^T \rho_\tau(y_{it} - \widetilde{\mathbf{\Pi}}(x_{it})^\top \widehat{\boldsymbol{\vartheta}}_i^{\mathcal{G}})] - \mathbb{E}[\sum_t \rho_\tau(y_{it} - \mathbf{\Pi}(x_{it})^\top \boldsymbol{\vartheta}_{0i}^{\mathcal{G}_0})] = O_p(T\|\widehat{\boldsymbol{\vartheta}}_i^{\mathcal{G}} - \boldsymbol{\vartheta}_{0i}^{\mathcal{G}_0}\|_2^2) = O_p(nT\xi^2(n,T))$. By the definition of $\boldsymbol{\vartheta}_{0i}^{\mathcal{G}}$, we have $\sum_{i,t} \widetilde{\mathbf{\Pi}}(x_{it})^\top (\widehat{\boldsymbol{\vartheta}}_i^{\mathcal{G}} - \boldsymbol{\vartheta}_{0i}^{\mathcal{G}_0})[I\{e_{it} \le \widetilde{\mathbf{\Pi}}(x_{it})^\top \boldsymbol{\theta}_{0i}^{\mathcal{G}} - \mu_{0i} - m_{it}\} - \tau]$ has mean zero and thus of order $O_p(nT\xi^2(n,T))$. Thus,

$$\Big|\sum_{i,t} \rho_\tau(y_{it} - \mathbf{\Pi}(x_{it})^\top \widehat{\boldsymbol{\theta}}_i^{\mathcal{G}}) - \sum_{i,t} \rho_\tau(y_{it} - \mathbf{\Pi}(x_{it})^\top \boldsymbol{\theta}_{0i}^{\mathcal{G}})\Big| = O_p(nT\xi^2(n,T)).$$

Note that

$$\left|\frac{1}{nT}\sum_{i,t} \rho_\tau(y_{it} - \widetilde{\mathbf{\Pi}}(x_{it})^\top \boldsymbol{\vartheta}_{0i}^{\mathcal{G}}) - \mathbb{E}\left[\frac{1}{nT}\sum_{i,t} \rho_\tau(y_{it} - \widetilde{\mathbf{\Pi}}(x_{it})^\top \boldsymbol{\vartheta}_{0i}^{\mathcal{G}})\right]\right| = O_p((nT)^{-1/2}),$$

and

$$\left|\frac{1}{nT}\sum_{i,t} \rho_\tau(y_{it} - \widetilde{\mathbf{\Pi}}(x_{it})^\top \boldsymbol{\vartheta}_{0i}^{\mathcal{G}_0}) - \mathbb{E}\left[\frac{1}{nT}\sum_{i,t} \rho_\tau(y_{it} - \widetilde{\mathbf{\Pi}}(x_{it})^\top \boldsymbol{\vartheta}_{0i}^{\mathcal{G}_0})\right]\right| = O_p((nT)^{-1/2}).$$

So for the SIC, denote $\eta = KH\log(nT)/(nT)$ and we can write

$$
\begin{aligned}
&\mathrm{SIC}(K) - \mathrm{SIC}(K_0)\\
=&\log\left(1+\frac{\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\widehat{\boldsymbol{\vartheta}}_i^{\mathcal{G}})/(nT)-\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\widehat{\boldsymbol{\vartheta}}_i^{\mathcal{G}_0})/(nT)}{\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\widehat{\boldsymbol{\vartheta}}_i^{\mathcal{G}_0})/(nT)}\right)+O(\eta)\\
=&\log\left(1+\frac{\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_{0i}^{\mathcal{G}})/(nT)-\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_{0i}^{\mathcal{G}_0})/(nT)+O_p(\xi^2)}{\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_{0i}^{\mathcal{G}_0})/(nT)+O_p(\xi^2)}\right)+O(\eta)\\
=&\log\left(1+\frac{\mathbb{E}[\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_{0i}^{\mathcal{G}})]-\mathbb{E}[\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_{0i}^{\mathcal{G}_0})]+O_p(nT\xi^2+\sqrt{nT})}{\mathbb{E}[\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_{0i}^{\mathcal{G}_0})]+O_p(nT\xi^2+\sqrt{nT})}\right)\\
&+O(\eta)\geq\log(1+C\rho)+O\left(\eta\right)>0.
\end{aligned}
$$

Case 2. ($K_0 < K \leq K_{max}$, over-fitted model)

Again, let $\mathcal{G}$ be the partition that minimizes

$$
\min_{\boldsymbol{\theta}_i=\boldsymbol{\theta}_j \text{ if } G_i=G_j, |\mathcal{G}|=K} \mathbb{E}\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\rho_\tau(y_{it}-\mu_{0i}-\mathbf{\Pi}(x_{it})^\top\boldsymbol{\theta}_i)\right].
$$

Obviously, we will have $\boldsymbol{\theta}_{0i}^{\mathcal{G}} = \boldsymbol{\theta}_{0i}^{\mathcal{G}_0}$. By the same argument in case 1, we have

$$
\begin{aligned}
&\Big|\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\widehat{\boldsymbol{\vartheta}}_i^{\mathcal{G}})-\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_{0i}^{\mathcal{G}})\Big| = O_p(H+nTH^{-2d}),\\
\text{and } &\Big|\frac{1}{nT}\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_{0i}^{\mathcal{G}_0})-\mathbb{E}[\frac{1}{nT}\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\boldsymbol{\vartheta}_{0i}^{\mathcal{G}_0})]\Big| = O_p((nT)^{-1/2}).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\mathrm{SIC}(K) - \mathrm{SIC}(K_0)\\
=&\log\left(1+\frac{\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\widehat{\boldsymbol{\vartheta}}_i^{\mathcal{G}})/(nT)-\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\widehat{\boldsymbol{\vartheta}}_i^{\mathcal{G}_0})/(nT)}{\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\widehat{\boldsymbol{\vartheta}}_i^{\mathcal{G}_0})/(nT)}\right)\\
&+\frac{(K-K_0)H\log(nT)}{nT}\\
=&\log\left(1+\frac{O_p(H+nTH^{-2d})}{\sum_{i,t}\rho_\tau(y_{it}-\widetilde{\mathbf{\Pi}}(x_{it})^\top\widehat{\boldsymbol{\vartheta}}_i^{\mathcal{G}_0})/(nT)}\right)+\frac{(K-K_0)H\log(nT)}{nT}\\
=&O_p(H+nTH^{-2d})+\frac{(K-K_0)H\log(nT)}{nT}>0.
\end{aligned}
$$

□

Finally, we present an auxiliary lemma used in the proof of Theorem 2.

Lemma 8. *For $c > 0$ and $d > 0$ sufficiently small,*

$$\sup_{\substack{1 \le i \le n, 1 \le h \le H \\ \|\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_i^{\text{o}}\| \le c, |\mu_i - \widehat{\mu}_i^o| \le d}} \left| \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\Pi}_h(x_{it})[I\{e_{it} \le 0\} - I\{e_{it} \le \boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it} + \mu_i - \mu_{0i}\} \right.$$

$$\left. - F(0|x_{it}) + F(\boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it} + \mu_i - \mu_{0i}|x_{it})] \right|$$

$$= O_p(H^{3/2} T^{-1/2} \log(T) \log(nT)).$$

Proof of Lemma 8. We consider the upper bound for $\sum_{t=1}^{T} \boldsymbol{\Pi}_h(x_{it})[I\{e_{it} \le 0\} - I\{e_{it} \le \boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it} + \mu_i - \mu_{0i}\} - F_i(0|x_{it}) + F_i(\boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i - m_{it} + \mu_i - \mu_{0i}|x_{it})$ only since the lower bound can be derived similarly. Letting $m_{it}(\boldsymbol{\theta}_i) = \boldsymbol{\Pi}(x_{it})^\top \boldsymbol{\theta}_i$ and t_n satisfy that $|m_{it}(\boldsymbol{\theta}_i) - m_{it}(\widehat{\boldsymbol{\theta}}_i^{\text{o}}) + \mu_i - \widehat{\mu}_i^{\text{o}}| \le t_n$, we have

$$\sup_{\substack{1 \le i \le n, 1 \le h \le H \\ \|\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_i^{\text{o}}\| \le c, |\mu_i - \widehat{\mu}_i^{\text{o}}| \le d}} \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\Pi}_h(x_{it}) \left[I\{e_{it} \le m_{it}(\boldsymbol{\theta}_i) - m_{it} + \mu_i - \mu_{0i}\} - I\{e_{it} \le 0\}\right.$$

$$\left. + F(0|x_{it}) - F(m_{it}(\boldsymbol{\theta}_i) - m_{it} + \mu_i - \mu_{0i}|x_{it})\right]$$

$$\le \sup_{\substack{1 \le i \le n, 1 \le h \le H \\ \|\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_i^{\text{o}}\| \le c, |\mu_i - \widehat{\mu}_i^{\text{o}}| \le d}} \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\Pi}_h(x_{it}) \Big[I\{e_{it} \le m_{it}(\widehat{\boldsymbol{\theta}}_i^{\text{o}}) - m_{it} + \widehat{\mu}_i^{\text{o}} - \mu_{0i} + t_n\} - I\{e_{it} \le 0\}$$

$$+ F(0|x_{it}) - F(m_{it}(\boldsymbol{\theta}_i) - m_{it} + \mu_i - \mu_{0i}|x_{it})]$$

$$\le \sup_{1 \le i \le n, 1 \le h \le H} \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\Pi}_h(x_{it}) \Big[I\{e_{it} \le m_{it}(\widehat{\boldsymbol{\theta}}_i^{\text{o}}) - m_{it} + \widehat{\mu}_i^{\text{o}} - \mu_{0i} + t_n\} - I\{e_{it} \le 0\}$$

$$+ F(0|x_{it}) - F(m_{it}(\widehat{\boldsymbol{\theta}}_i^{\text{o}}) - m_{it} + \widehat{\mu}_i^{\text{o}} - \mu_{0i} + t_n|x_{it})\Big]$$

$$+ \sup_{\substack{1 \le i \le n, 1 \le h \le H \\ \|\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_i^{\text{o}}\| \le c, |\mu_i - \widehat{\mu}_i^{\text{o}}| \le d}} \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{\Pi}(x_{it})[F(m_{it}(\widehat{\boldsymbol{\theta}}_i^{\text{o}}) - m_{it} + \widehat{\mu}_i^{\text{o}} - \mu_{0i} + t_n|x_{it})$$

$$- F(m_{it}(\boldsymbol{\theta}_i) - m_{it} + \mu_i - \mu_{0i})],$$

where the first inequality stems from the increasing monotonicity of the indicator function. The second term in the last line can be arbitrarily small since $|m_{it}(\boldsymbol{\theta}_i) - m_{it}(\widehat{\boldsymbol{\theta}}_i^{\text{o}}) + \mu_i - \widehat{\mu}_i^{\text{o}}| \le t_n$ while t_n is arbitrarily small when we choose c and d to be sufficiently small.

Note that $\mathbb{E}[|\boldsymbol{\Pi}_h(x_{it})(1\{e_{it} \le a_n + \delta_n\} - 1\{e_{it} \le a_n\})|^q] \le (C\sqrt{H})^{q-2}\delta_n$, for $q = 3, 4, \ldots$

By Lemma 1, for any non-negative sequences $a_n \to 0$, $\delta_n \to 0$, we have that for any $u > 0$,

$$\begin{aligned}
&\mathbb{P}\left[\left|\frac{1}{T}\sum_{t=1}^{T}\mathbf{\Pi}_h(x_{it})[1\{e_{it} \le a_n + \delta_n\} - 1\{e_{it} \le a_n\} - F(a_n + \delta_n|x_{it}) + F(a_n|x_{it})]\right| > u\right] \\
\le & C\log(T)\exp\left[-C\frac{Tu^2}{\log(T)(u\sqrt{H} + \delta_n)}\right].
\end{aligned} \tag{B.6}$$

Denote $\boldsymbol{\vartheta}_i = (\mu_i, \boldsymbol{\theta}_i^\top)^\top$, $\boldsymbol{\vartheta}_{0i} = (\mu_{0i}, \theta_{0i}^\top)^\top$ and $\widetilde{m}(\boldsymbol{\vartheta}_i) = m_{it}(\boldsymbol{\theta}_i) + \mu_i$. Let $A_i = \{\boldsymbol{\vartheta}_i : \|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_{0i}\|_2 \le C\xi(n,T)\}$. Similarly to Lemma 2, we construct an $(nT)^{-\delta}$ covering of A_i with size $R = O((nT)^{CH})$, with elements denoted by $\{\boldsymbol{\vartheta}_i^{(1)}, \ldots, \boldsymbol{\vartheta}_i^{(R)}\}$. Then, we have

$$\begin{aligned}
&\sup_{\substack{1\le i\le n, 1\le h\le H \\ \boldsymbol{\vartheta}_i \in A_i}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{\Pi}_h(x_{it})\Big[1\{e_{it} \le \widetilde{m}_{it}(\boldsymbol{\vartheta}_i) - m_{it} - \mu_{0i} + t_n\} - 1\{e_{it} \le 0\} \\
&\qquad - F(\widetilde{m}_{it}(\boldsymbol{\vartheta}_i) - m_{it} - \mu_{0i} + t_n|x_{it}) + F(0|x_{it})\Big] \\
\le & \max_{\substack{1\le i\le n, 1\le h\le H \\ 1\le r\le R}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{\Pi}_h(x_{it})\Big[1\{e_{it} \le \widetilde{m}_{it}(\boldsymbol{\vartheta}_i^{(r)}) - m_{it} - \mu_{0i} + t_n\} - 1\{e_{it} \le 0\} \\
&\qquad - F(\widetilde{m}_{it}(\boldsymbol{\vartheta}_i^{(r)}) - m_{it} - \mu_{0i} + t_n|x_{it}) + F(0|x_{it})\Big] \\
+ & \sup_{\substack{1\le i\le n, 1\le h\le H \\ 1\le r\le R, \|\boldsymbol{\theta}_i - \boldsymbol{\theta}_i^{(r)}\| \le C(nT)^{-\delta}}} \frac{1}{T}\sum_{t=1}^{T}\mathbf{\Pi}_h(x_{it})\Big[1\{e_{it} \le \widetilde{m}_{it}(\boldsymbol{\vartheta}_i) - m_{it} - \mu_{0i} + t_n\} \\
&\qquad - 1\{e_{it} \le \widetilde{m}_{it}(\boldsymbol{\vartheta}_i^{(r)}) - m_{it} - \mu_{0i} + t_n\} \\
&\qquad - F(\widetilde{m}_{it}(\boldsymbol{\vartheta}_i) - m_{it} - \mu_{0i} + t_n|x_{it}) + F(\widetilde{m}_{it}(\boldsymbol{\vartheta}_i^{(r)}) - m_{it} + t_n|x_{it})\Big] \\
&\qquad := I_1 + I_2.
\end{aligned}$$

By (B.6), using the union bound and that $|\widetilde{m}_{it}(\boldsymbol{\vartheta}_i^{(r)}) - m_{it} - \mu_{0i} + t_n| \le C\sqrt{H}\xi(n,T)$, we have

$$I_1 = O_p((\sqrt{H}/T)\log(T)(H\log(nT))).$$

For I_2, using the monotonicity of the indicator function and define t_n' such that

$|\widetilde{m}_{it}(\boldsymbol{\vartheta}_i) - \widetilde{m}_{it}(\boldsymbol{\vartheta}_i^{(r)})| \le t_n'$ for all $\|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_i^{(r)}\|_2 \le C(nT)^{-\delta}$, we have

$$
\begin{aligned}
I_2 \le & \sup_{\substack{1\le r\le R, 1\le i\le n, 1\le h\le H \\ \boldsymbol{\theta}_i \in A_i, \|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_i^{(r)}\|_2 \le C(nT)^{-\delta}}} \frac{1}{T}\sum_t \boldsymbol{\Pi}_h(x_{it})\Big[1\{e_{it} \le \widetilde{m}_{it}(\boldsymbol{\vartheta}_i^{(r)}) - m_{it} - \mu_{0i} + t_n' + t_n\} \\
& - 1\{e_{it} \le \widetilde{m}_{it}(\boldsymbol{\vartheta}_i^{(r)}) - m_{it} - \mu_{0i} + t_n\} - F(\widetilde{m}_{it}(\boldsymbol{\vartheta}_i) - m_{it} - \mu_{0i} + t_n | x_{it}) \\
& + F(\widetilde{m}_{it}(\boldsymbol{\vartheta}_i^{(r)}) - m_{it} - \mu_{0i} + t_n | x_{it})\Big] \\
\le & \sup_{1\le r\le R, 1\le i\le n, 1\le h\le H} \frac{1}{T}\sum_t \boldsymbol{\Pi}_h(x_{it})\Big[1\{e_{it} \le \widetilde{m}_{it}(\boldsymbol{\vartheta}_i^{(r)}) - m_{it} - \mu_{0i} + t_n' + t_n\} \\
& - 1\{e_{it} \le \widetilde{m}_{it}(\boldsymbol{\vartheta}_i^{(r)}) - m_{it} - \mu_{0i} + t_n\} - F(\widetilde{m}_{it}(\boldsymbol{\vartheta}_i) - m_{it} - \mu_{0i} + t_n' + t_n | x_{it}) \\
& + F(\widetilde{m}_{it}(\boldsymbol{\vartheta}_i^{(r)}) - m_{it} - \mu_{0i} + t_n | x_{it})\Big] \\
+ & \sup_{\substack{1\le r\le R, 1\le i\le n, 1\le h\le H \\ \boldsymbol{\vartheta}_i \in A_i, \|\boldsymbol{\vartheta}_i - \boldsymbol{\vartheta}_i^{(r)}\|_2 \le C(nT)^{-\delta}}} \frac{1}{T}\sum_t \boldsymbol{\Pi}_h(x_{it})\Big[F(\widetilde{m}_{it}(\boldsymbol{\vartheta}_i) - m_{it} - \mu_{0i} + t_n' + t_n | x_{it}) \\
& - F(\widetilde{m}_{it}(\boldsymbol{\vartheta}_i) - m_{it} - \mu_{0i} + t_n | x_{it})\Big] \\
=: & I_{21} + I_{22}.
\end{aligned}
$$

Again by (B.6) for I_{21} with union bound, and that I_{22} is arbitrarily small by the smoothness of $F(\cdot)$, we obtain

$$
I_2 = O_p((\sqrt{H}/T)\log(T)(H\log(nT))).
$$

□

C Additional simulation results and information

This appendix presents detailed information on the data generating process (DGP) and some additional simulation results.

C.1 Detailed information on DGP in Section 5.2

To better illustrate the varying subgroup structures of the DGP across τ, the conditional quantile functions for two lower quantile subgroups and three upper quantile subgroups are presented in Figures 1(a)–1(b) and Figures 2(a)–2(c), respectively. In addition, by the definition of $m_{i,\tau}(\cdot)$, we need to add a constant independent of x_{it} to ensure that

$\int_0^1 m_{i,\tau}(x)dx = 0$. Therefore, for $\tau = 0.1$ and 0.9, $m_{i,\tau}$ is defined as

$$m_{i,0.1}(x_{it}) = \begin{cases} \sin(2\pi x_{it}) - (0.4 + 0.8x_{it}) \times 1.281552 + 1.0252412, & i \in G_{1,\mathrm{L}}; \\ \sin(2\pi x_{it}) - (1.2 - 0.8x_{it}) \times 1.281552 + 1.0252412, & i \in G_{2,\mathrm{L}}, \end{cases}$$

and

$$m_{i,0.9}(x_{it}) = \begin{cases} \sin(2\pi x_{it}) + (0.4 + 0.8x_{it}) \times 1.281552 - 1.0252412, & i \in G_{1,\mathrm{U}}; \\ \sin(2\pi x_{it}) + (1.2 - 0.8x_{it}) \times 1.281552 - 1.0252412, & i \in G_{2,\mathrm{U}}; \\ \sin(2\pi x_{it}), & i \in G_{3,\mathrm{U}}. \end{cases}$$

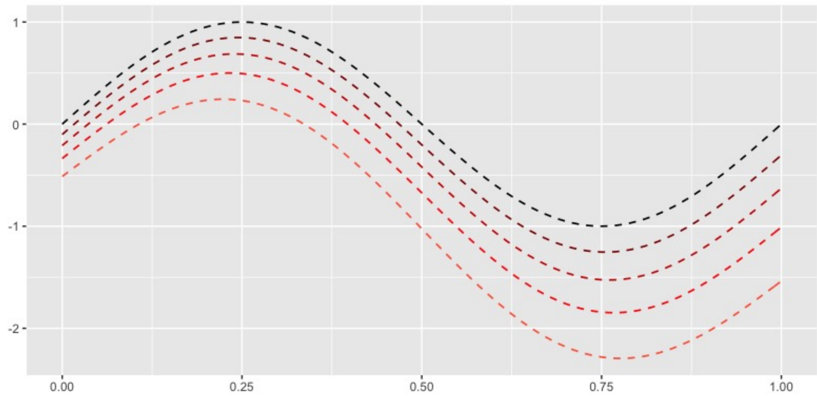
C.2 Additional simulation results

We conduct an additional simulation experiment with varying subgroup structure at different quantile levels. The data generating process is identical to Experiment 4, only with $(n, T) = (60, 1000)$.

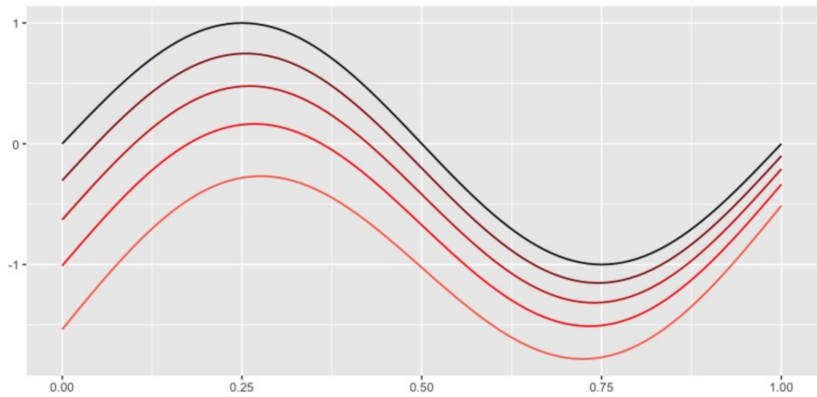
The MSEs and the percentages of correct subgroup recovery for this experiment are summarized in Table 1, and the empirical coverage probabilities for the pointwise confidence intervals are summarized in Figure 3. The findings are similar to Experiment 4 in the main paper. With T increased to 1000, we have the percentages of correct subgroup recovery for three quantile levels increase accordingly, and the MSEs decrease accordingly. The empirical coverage probabilities are closer to the target line for three quantile levels than the case with $(n, T) = (60, 100)$.

Table 1: Mean squared errors (MSEs) of the oracle and SCAD-penalized estimators, and percentages of correct subgroup recovery for the additional experiment. We consider three quantile levels of $\tau = 0.1$, 0.5 and 0.9.

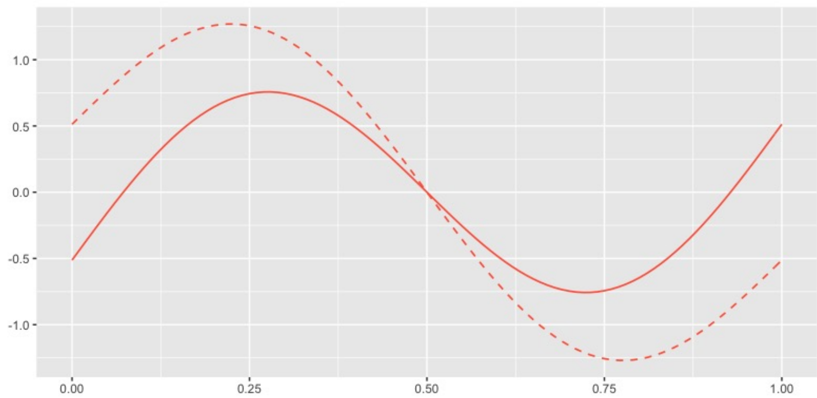
$(n, T) = (60, 1000)$		$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$
% of correct subgroup		99.2%	100%	94.4%
MSE ($\times 10^{-4}$)	Oracle	0.443	0.0913	0.682
	SCAD	0.460	0.0911	0.744



(a) $Q_\tau(y_{it}|x_{it})$ of the 1st lower subgroup

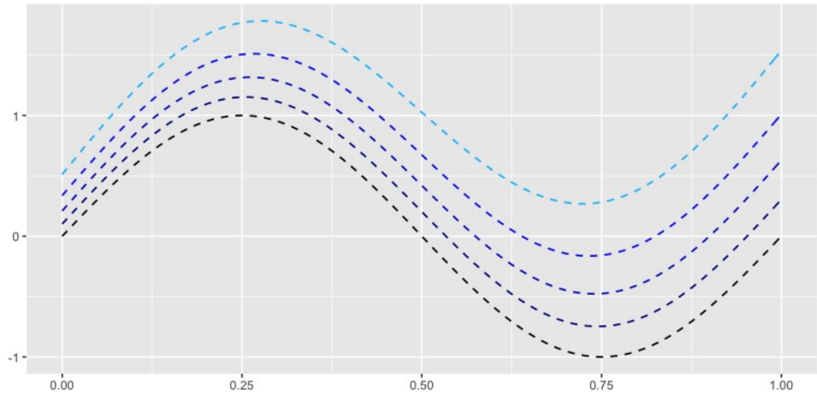


(b) $Q_\tau(y_{it}|x_{it})$ of the 2nd lower subgroup

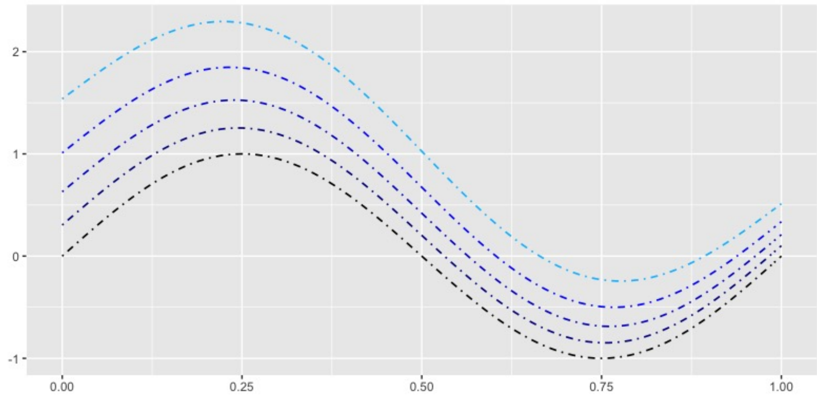


(c) $m_{(k),0.1}(\cdot)$ for $k = 1, 2$

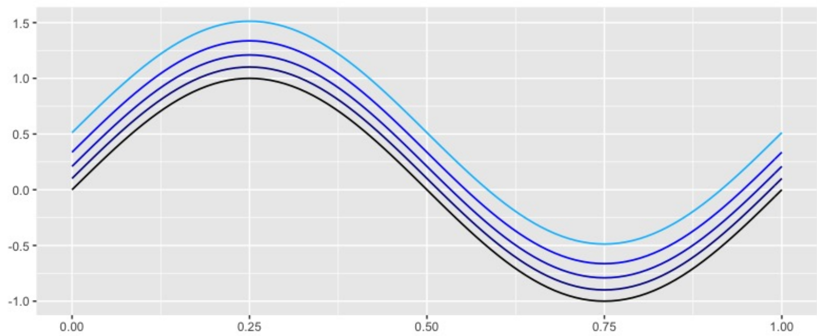
Figure 1: Conditional quantile functions at lower quantiles and $m_{(k),0.1}$ of two lower subgroups. - - - -: 1st subgroup; ——: 2nd subgroup. ■ $\tau = 0.5$; ■ $\tau = 0.4$; ■ $\tau = 0.3$; ■ $\tau = 0.2$; ■ $\tau = 0.1$.



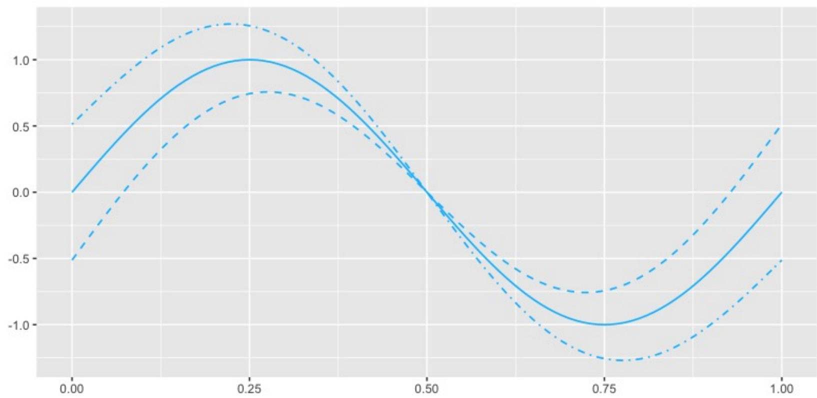
(a) $Q_\tau(y_{it}|x_{it})$ of the 1st upper subgroup



(b) $Q_\tau(y_{it}|x_{it})$ of the 2nd upper subgroup



(c) $Q_\tau(y_{it}|x_{it})$ of the 3rd upper subgroup



(d) $m_{(k),0.9}(\cdot)$ for $k = 1, 2, 3$

Figure 2: Conditional quantile functions at lower quantiles and $m_{(k),0.9}$ of three upper subgroups. - - - -: 1st subgroup; —·—: 2nd subgroup; ——: 3rd subgroup. ■ $\tau = 0.5$; ■ $\tau = 0.6$; ■ $\tau = 0.7$; ■ $\tau = 0.8$; ■ $\tau = 0.9$.

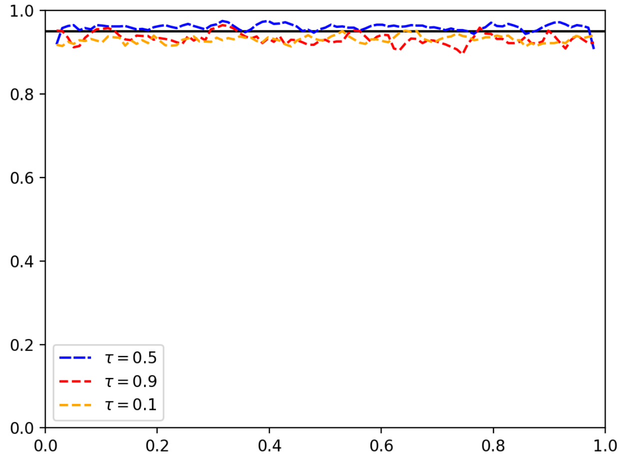


Figure 3: Empirical coverage probabilities for pointwise confidence intervals of $m_{i,\tau}(\cdot)$ for the case with varying subgroup structures at different quantile levels. We consider $(n, T) = (60, 1000)$ and three quantile levels of $\tau = 0.1$, 0.5 and 0.9.

References

Fan, J. and Yao, Q. (2008). *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media.

Kato, K., Galvao Jr, A. F., and Montes-Rojas, G. V. (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics*, 170:76–91.

van der Vaart, A. W. and Wellner, J. A. (1996). Weak convergence and empirical processes with applications to statistics.

Zhou, S., Shen, X., and Wolfe, D. (1998). Local asymptotics for regression splines and confidence regions. *Annals of Statistics*, 26:1760–1782.