Supplementary Appendix to
"Multi-dimensional latent Group Structure with Heterogeneous Distributions"

This technical appendix contains the proofs of the theorems and corollaries. Throughout this appendix, we suppress the dependence of the parameters and their estimates on τ_k wherever no confusion arises.

Proof of Theorem 1.

It follows from (2.2) that for each τ_k where $k = 1, \ldots, K$,

$$\big(\widehat{\gamma}_g, \widehat{\gamma}_h, \widehat{\boldsymbol{\theta}}(\tau_k)\big) = \underset{(\gamma_g,\gamma_h,\boldsymbol{\theta}(\tau_k))\in\Gamma_G\times\Gamma_H\times\Theta'}{\arg\min} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_{\tau_k}\big(y_{it} - \alpha_{h_i} - \lambda_t - x_{it}'\beta_{g_i}\big), \tag{A.1}$$

with $\Theta' := \mathbb{B}^G \times \mathbb{A}^H \times \mathbb{D}^T$. Let

$$\Delta_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) := \frac{1}{N}\sum_{i=1}^{N}\big(\rho_{\tau_k}(y_{it} - \alpha_{h_i} - \lambda - x_{it}'\beta_{g_i}) - \rho_{\tau_k}(y_{it} - \alpha^0_{h_i^0} - \lambda^0_t - x_{it}'\beta^0_{g_i^0})\big),$$

where with a little notation abuse, $\boldsymbol{\alpha} := (\alpha_{h_1}+\lambda, \ldots, \alpha_{h_N}+\lambda)'$ and $\boldsymbol{\beta} := (\beta'_{g_1}, \ldots, \beta'_{g_N})'$ denote the fixed effect and the coefficient vector, respectively. Put $l_t = \underset{i=1,\ldots,N}{\arg\max}\big\{(\widehat{\alpha}_{\widehat{h}_i}+\widehat{\lambda}_t-\alpha^0_{h_i^0}-\lambda^0_t)^2+\|\widehat{\beta}_{\widehat{g}_i}-\beta^0_{g_i^0}\|^2\big\}$(In the presence of non-unique solutions an arbitrary solution can be assigned to l_t.), let $(g_0, h_0) := (g^0_{l_t}, h^0_{l_t})$, and for any $\delta > 0$,

$$B_t(\delta) := \big\{(\boldsymbol{\alpha}, \boldsymbol{\beta}) : \max_{i=1,\ldots,N}\sqrt{(\alpha_{h_i} + \lambda - \alpha^0_{h_0} - \lambda^0_t)^2 + \|\beta_{g_i} - \beta^0_{g_0}\|^2} \le \delta\big\}$$

where $h_i, h_0 \in \{1, \ldots, H\}$, and $g_i, g_0 \in \{1, \ldots, G\}$. Note that $B_t(\delta)$ is independent of i. Following from (A.1) we have

$$\Big\{\max_{i=1,\ldots,N}\sqrt{(\widehat{\alpha}_{\widehat{h}_i} + \widehat{\lambda}_t - \alpha^0_{h_i^0} - \lambda^0_t)^2 + \|\widehat{\beta}_{\widehat{g}_i} - \beta^0_{g_i^0}\|^2} > \delta\Big\} \subset \Big\{\Delta_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) \le 0, (\boldsymbol{\alpha}, \boldsymbol{\beta}) \notin B_t(\delta), 1 \le \exists\, t \le T\Big\}. \tag{A.2}$$

For $(\boldsymbol{\alpha}, \boldsymbol{\beta}) \notin B_t(\delta)$, define $r = \delta/\big(\max_{i=1,\ldots,N}\sqrt{(\alpha_{h_i} + \lambda - \alpha^0_{h_0} - \lambda^0_t)^2 + \|\beta_{g_i} - \beta^0_{g_0}\|^2}\big)$, $\widetilde{\alpha}_{h_i} + \widetilde{\lambda} = r(\alpha_{h_i} + \lambda) + (1-r)(\alpha^0_{h_0} + \lambda^0_t)$ and $\widetilde{\beta}_{g_i} = r\beta_{g_i} + (1-r)\beta^0_{g_0}$ for $i = 1, \ldots, N$. Then $0 < r < 1$, $(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}) \in B_t(\delta)$ and $\Delta_t(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}) \le r\Delta_t(\boldsymbol{\alpha}, \boldsymbol{\beta})$ where $\widetilde{\boldsymbol{\alpha}} := (\widetilde{\alpha}_{h_1} + \widetilde{\lambda}, \ldots, \widetilde{\alpha}_{h_N} + \widetilde{\lambda})'$ and $\widetilde{\boldsymbol{\beta}} := (\widetilde{\beta}'_{g_1}, \ldots, \widetilde{\beta}'_{g_N})'$. It then follows from (A.2) that for some $\epsilon_\delta > 0$,

$$\begin{aligned}
&\Big\{\max_{i=1,\ldots,N}\sqrt{(\widehat{\alpha}_{\widehat{h}_i} + \widehat{\lambda}_t - \alpha^0_{h_i^0} - \lambda^0_t)^2 + \|\widehat{\beta}_{\widehat{g}_i} - \beta^0_{g_i^0}\|^2} > \delta\Big\} \subset \Big\{\Delta_t(\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}) \le 0, (\widetilde{\boldsymbol{\alpha}}, \widetilde{\boldsymbol{\beta}}) \in B_t(\delta), 1 \le \exists\, t \le T\Big\} \\
&\subset \Big\{\max_{1\le t\le T}\sup_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in B_t(\delta)}|\Delta_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) - \mathsf{E}[\Delta_t(\boldsymbol{\alpha}, \boldsymbol{\beta})]| > \epsilon_\delta\Big\},
\end{aligned}$$

provided that $\min_{1\le t\le T}\min_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in B_t(\delta)}\mathsf{E}[\Delta_t(\boldsymbol{\alpha}, \boldsymbol{\beta})] > 0$ by the identification condition Assumption

1(*iii*). Hence, it suffices to show that for any $\epsilon > 0$,

$$\max_{1\le t\le T} \mathsf{P}\Big(\sup_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in B(\delta)} |\Delta_t(\boldsymbol{\alpha},\boldsymbol{\beta}) - \mathsf{E}[\Delta_t(\boldsymbol{\alpha},\boldsymbol{\beta})]| > \epsilon \Big) = o(T^{-1}) \tag{A.3}$$

as N and $T \to \infty$. Without loss of generality, assume the true values of the regression parameters equal 0 and rewrite $B_t(\delta) = B(\delta) = \big\{(\boldsymbol{\alpha},\boldsymbol{\beta}) : \max_{i=1,\dots,N} \sqrt{(\alpha_{h_i}+\lambda)^2 + \|\beta_{g_i}\|^2} \le \delta\big\}$ and $\Delta_t(\boldsymbol{\alpha},\boldsymbol{\beta}) = \Delta(\boldsymbol{\alpha},\boldsymbol{\beta})$ for simplicity. We prove (A.3) using the similar arguments for (A.3) of Kato et al. (2012). Observing that $\{(\alpha,\beta) : \sqrt{(\alpha+\lambda)^2 + \|\beta\|^2} \le \delta\}$ is compact, there exists a finite open cover with centers $\{(\alpha+\lambda)^{(j)}, \beta^{(j)}\}_{j=1}^J$ and radius r and $J < \infty$. Then for every $(\boldsymbol{\alpha},\boldsymbol{\beta}) \in B(\delta)$, there exists some $j_{h_i,g_i} \in \{1,\dots,J\}$ such that $\sqrt{(\alpha_{h_i}+\lambda-(\alpha+\lambda)^{(j_{h_i,g_i})})^2 + \|\beta_{g_i} - \beta^{(j_{h_i,g_i})}\|^2} \le r$ for $i = 1,\dots,N$. Denote $\boldsymbol{\alpha}^{(J)} := \big((\alpha+\lambda)^{(j_{h_1,g_1})},\dots,(\alpha+\lambda)^{(j_{h_N,g_N})}\big)'$ and $\boldsymbol{\beta}^{(J)} := \big(\beta^{(j_{h_1,g_1})'},\dots,\beta^{(j_{h_N,g_N})'}\big)'$. We have the following decomposition.

$$\begin{aligned}
&\big|\Delta(\boldsymbol{\alpha},\boldsymbol{\beta}) - \mathsf{E}[\Delta(\boldsymbol{\alpha},\boldsymbol{\beta})]\big| \\
\le &\big|\Delta(\boldsymbol{\alpha},\boldsymbol{\beta}) - \Delta(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)})\big| + \big|\Delta(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)}) - \mathsf{E}[\Delta(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)})]\big| \\
&+ \mathsf{E}\big|\Delta(\boldsymbol{\alpha},\boldsymbol{\beta}) - \Delta(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)})\big|.
\end{aligned} \tag{A.4}$$

We bound the first term on the right side of (A.4) as follows.

$$\begin{aligned}
&\big|\Delta(\boldsymbol{\alpha},\boldsymbol{\beta}) - \Delta(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)})\big| \\
=&\Big|\frac{1}{N}\sum_{i=1}^N \big(\rho_{\tau_k}(y_{it} - (\alpha_{h_i}+\lambda) - x_{it}'\beta_{g_i}) - \rho_{\tau_k}(y_{it} - (\alpha+\lambda)^{(j_{h_i,g_i})} - x_{it}'\beta^{(j_{h_i,g_i})})\big)\Big| \\
\le&\frac{1}{N}\sum_{i=1}^N C_0(1+\sup_{i\ge1}\|x_{i1}\|)(|\alpha_{h_i}+\lambda-(\alpha+\lambda)^{(j_{h_i,g_i})}| + \|\beta_{g_i} - \beta^{(j_{h_i,g_i})}\|) \\
\le&C_0(1+\sup_{i\ge1}\|x_{i1}\|)\sqrt{2}r,
\end{aligned}$$

where C_0 is a universal constant. By setting $r = \epsilon/(3\sqrt{2}C_0(1+M))$ with $\sup_{i\ge1}\|x_{i1}\| \le M$, following from (A.4) we have

$$\big|\Delta(\boldsymbol{\alpha},\boldsymbol{\beta}) - \mathsf{E}[\Delta(\boldsymbol{\alpha},\boldsymbol{\beta})]\big| \le \epsilon/3 + \big|\Delta(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)}) - \mathsf{E}[\Delta(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)})]\big| + \epsilon/3.$$

Let $\mathcal{B}$ denote all the possible outcomes to put $\{((\alpha+\lambda)^{(j_{h_i,g_i})}, \beta^{(j_{h_i,g_i})})\}_{i=1}^N$ into the J balls, $\{(\alpha,\beta) : \sqrt{(\alpha+\lambda-(\alpha+\lambda)^{(j)})^2 + \|\beta-\beta^{(j)}\|^2} \le r\}$ for $j = 1,\dots,J$. Hence,

$$\begin{aligned}
\mathsf{P}\Big(\sup_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in B(\delta)} \big|\Delta(\boldsymbol{\alpha},\boldsymbol{\beta}) - \mathsf{E}[\Delta(\boldsymbol{\alpha},\boldsymbol{\beta})]\big| > \epsilon\Big) &\le \mathsf{P}\Big(\max_{(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)})\in\mathcal{B}} \big|\Delta(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)}) - \mathsf{E}[\Delta(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)})]\big| > \epsilon/3\Big) \\
&\le \sum_{(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)})\in\mathcal{B}} \mathsf{P}\Big(\big|\Delta(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)}) - \mathsf{E}[\Delta(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)})]\big| > \epsilon/3\Big).
\end{aligned} \tag{A.5}$$

Denote $\Delta_i(\alpha,\beta) := \rho_{\tau_k}(y_{it}-\alpha-\lambda-x_{it}'\beta)-\rho_{\tau_k}(y_{it})$. We apply Hoeffding's inequality to the right side of (A.5) to obtain that

$$\begin{aligned}
&\max_{1\le t\le T}\mathsf{P}\Big(\sup_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in B(\delta)}|\Delta_t(\boldsymbol{\alpha},\boldsymbol{\beta})-\mathsf{E}[\Delta_t(\boldsymbol{\alpha},\boldsymbol{\beta})]|>\epsilon\Big)\\
\le&\sum_{(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)})\in\mathcal{B}}\mathsf{P}\Big(\big|N^{-1}\sum_{i=1}^{N}\big(\Delta_i((\alpha+\lambda)^{(j_{h_i,g_i})},\beta^{(j_{h_i,g_i})})-\mathsf{E}[\Delta_i((\alpha+\lambda)^{(j_{h_i,g_i})},\beta^{(j_{h_i,g_i})})]\big)\big|>\epsilon/3\Big)\\
\le&\sum_{(\boldsymbol{\alpha}^{(J)},\boldsymbol{\beta}^{(J)})\in\mathcal{B}}2\exp(-cN)
\end{aligned}\tag{A.6}$$

where $c>0$ is a universal constant. Note that $|\mathcal{B}|=\binom{GH+N-1}{N}\binom{J+GH-1}{GH}$ determined by the double-groups assignment for units and the selection of the centers, $\{(\alpha+\lambda)^{(j)},\beta^{(j)}\}_{j=1}^{J}$, for each group pair (h,g). Combining with (A.6) yields

$$\begin{aligned}
&\max_{1\le t\le T}\mathsf{P}\Big(\sup_{(\boldsymbol{\alpha},\boldsymbol{\beta})\in B(\delta)}|\Delta_t(\boldsymbol{\alpha},\boldsymbol{\beta})-\mathsf{E}[\Delta_t(\boldsymbol{\alpha},\boldsymbol{\beta})]|>\epsilon\Big)\\
\le&2\tbinom{GH+N-1}{N}\tbinom{J+GH-1}{GH}\exp(-cN)=O\big(\exp\{(GH-1)\log N-cN\}\big),\quad\text{as } N\to\infty,
\end{aligned}$$

provided that G and H are finite. This shows (A.3) if $\log T/N\to 0$. Thus, for all $\delta>0$,

$$P\left(\max_{i=1,\dots,N}\sqrt{(\widehat{\alpha}_{\widehat{h}_i}+\widehat{\lambda}_t-\alpha^0_{h^0_i}-\lambda^0_t)^2+\|\widehat{\beta}_{\widehat{g}_i}-\beta^0_{g^0_i}\|^2}>\delta\right)\to 0$$

as N and $T\to\infty$, which implies

$$\max_{i=1,\dots,N}\|\widehat{\beta}_{\widehat{g}_i}-\beta^0_{g^0_i}\|\xrightarrow{\mathsf{P}}0\quad\text{and}\quad\max_{i=1,\dots,N}|\widehat{\alpha}_{\widehat{h}_i}+\widehat{\lambda}_t-(\alpha^0_{h^0_i}+\lambda^0_t)|\xrightarrow{\mathsf{P}}0.$$

This completes the proof. ∎

Define $\boldsymbol{\beta}^0:=(\beta^0_1,\dots,\beta^0_G)'$, $\widehat{\boldsymbol{\beta}}:=(\widehat{\beta}_1,\dots,\widehat{\beta}_G)'$, $\Phi^0_t=(\alpha^0_1+\lambda^0_t,\dots,\alpha^0_1+\lambda^0_t)'$ and $\widehat{\Phi}_t=(\widehat{\alpha}_1+\widehat{\lambda}_t,\dots,\widehat{\alpha}_H+\widehat{\lambda}_t)'$. The following lemma shows the convergence of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Phi}_t$ with respect to the Hausdorff distance, from which we obtain Corollary 1.

Lemma 2. *Let Assumption 1 hold. If* $\log T/N\to 0$ *we have*

$$d(\widehat{\boldsymbol{\beta}},\boldsymbol{\beta}^0)\xrightarrow{\mathsf{P}}0\quad\textit{and}\quad d(\widehat{\Phi}_t,\Phi^0_t)\xrightarrow{\mathsf{P}}0$$

as N *and* $T\to\infty$, *where*

$$d(\widehat{\boldsymbol{\beta}},\boldsymbol{\beta}^0)=\max\left\{\max_{g\in\{1,\dots,G\}}\min_{\widetilde{g}\in\{1,\dots,G\}}\|\widehat{\beta}_{\widetilde{g}}-\beta^0_g\|,\ \max_{\widetilde{g}\in\{1,\dots,G\}}\min_{g\in\{1,\dots,G\}}\|\widehat{\beta}_{\widetilde{g}}-\beta^0_g\|\right\},$$

and

$$d(\widehat{\Phi}_t, \Phi_t^0) = \max\left\{ \max_{h\in\{1,\dots,H\}} \min_{\widetilde{h}\in\{1,\dots,H\}} |\widehat{\alpha}_{\widetilde{h}} + \widehat{\lambda}_t - (\alpha_h^0 + \lambda_t^0)|, \max_{\widetilde{h}\in\{1,\dots,H\}} \min_{h\in\{1,\dots,H\}} |\widehat{\alpha}_{\widetilde{h}} + \widehat{\lambda}_t - (\alpha_h^0 + \lambda_t^0)| \right\}.$$

Proof. We prove the first result of Lemma 2 and the second result can be shown similarly. It follows from (3.1) of Theorem 1 that for all $g \in \{1, \dots, G\}$,

$$\begin{aligned}
&\left(\frac{1}{N}\sum_{i=1}^N I\{g_i^0 = g\}\right)\left(\min_{\widetilde{g}\in\{1,\dots,G\}}\left(\sum_{k=1}^K \|\widehat{\beta}_{\widetilde{g}}(\tau_k) - \beta_g^0(\tau_k)\|\right)\right) = \frac{1}{N}\sum_{i=1}^N \min_{\widetilde{g}\in\{1,\dots,G\}} I\{g_i^0 = g\}\left(\sum_{k=1}^K \|\widehat{\beta}_{\widetilde{g}}(\tau_k) - \beta_g^0(\tau_k)\|\right) \\
&\le \min_{\widetilde{g}\in\{1,\dots,G\}} \max_{i=1,\dots,N}\left(\sum_{k=1}^K \|\widehat{\beta}_{\widetilde{g}}(\tau_k) - \beta_{g_i^0}^0(\tau_k)\|\right) \\
&\le \max_{i=1,\dots,N}\left(\sum_{k=1}^K \|\widehat{\beta}_{\widehat{g}_i}(\tau_k) - \beta_{g_i^0}^0(\tau_k)\|\right) \\
&\le \left(\sum_{k=1}^K \max_{i=1,\dots,N} \|\widehat{\beta}_{\widehat{g}_i}(\tau_k) - \beta_{g_i^0}^0(\tau_k)\|\right) \xrightarrow{\mathsf{P}} 0,
\end{aligned}$$

as N and $T \to \infty$. By Assumption 1(*iv*), we hence have

$$\min_{\widetilde{g}\in\{1,\dots,G\}}\left(\sum_{k=1}^K \|\widehat{\beta}_{\widetilde{g}}(\tau_k) - \beta_g^0(\tau_k)\|\right) \xrightarrow{\mathsf{P}} 0. \tag{A.7}$$

Define $\sigma(g) = \arg\min_{\widetilde{g}\in\{1,\dots,G\}} \sum_{k=1}^K \|\widehat{\beta}_{\widetilde{g}}(\tau_k) - \beta_g^0(\tau_k)\|$ for all $g \in \{1, \dots, G\}$. We show that $\sigma : \{1, \dots, G\} \to \{1, \dots, G\}$ is a one-to-one mapping with probability approaching 1. Let $g \neq \widetilde{g}$ and by the triangle inequality, we obtain that

$$\begin{aligned}
\sum_{k=1}^K \|\widehat{\beta}_{\sigma(g)}(\tau_k) - \widehat{\beta}_{\sigma(\widetilde{g})}(\tau_k)\| &\ge \sum_{k=1}^K \|\beta_g^0(\tau_k) - \beta_{\widetilde{g}}^0(\tau_k)\| \\
&\quad - \sum_{k=1}^K \|\widehat{\beta}_{\sigma(g)}(\tau_k) - \beta_g^0(\tau_k)\| - \sum_{k=1}^K \|\widehat{\beta}_{\sigma(\widetilde{g})}(\tau_k) - \beta_{\widetilde{g}}^0(\tau_k)\|.
\end{aligned} \tag{A.8}$$

Because of Assumption 1(*v*) and (A.7) the right side of (A.8) converges in probability to a strictly positive constant. Hence σ is invertible and define its inverse as σ^{-1}.

For all $\widetilde{g} \in \{1, \dots, G\}$ note that $\widetilde{g} = \sigma[\sigma^{-1}(\widetilde{g})]$. We then use (A.7) to obtain that

$$\begin{aligned}
\min_{g\in\{1,\dots,G\}}\left(\sum_{k=1}^K \|\widehat{\beta}_{\widetilde{g}}(\tau_k) - \beta_g^0(\tau_k)\|\right) &\le \sum_{k=1}^K \|\widehat{\beta}_{\widetilde{g}}(\tau_k) - \beta_{\sigma^{-1}(\widetilde{g})}^0(\tau_k)\| \\
&= \min_{g\in\{1,\dots,G\}}\left(\sum_{k=1}^K \|\widehat{\beta}_g(\tau_k) - \beta_{\sigma^{-1}(\widetilde{g})}^0(\tau_k)\|\right) \xrightarrow{\mathsf{P}} 0.
\end{aligned} \tag{A.9}$$

It follows from (A.7) and (A.9) that for each $\tau_k, k = 1, \dots, K$, $\min_{\widetilde{g}\in\{1,\dots,G\}} \|\widehat{\beta}_{\widetilde{g}}(\tau_k) - \beta_g^0(\tau_k)\| \xrightarrow{\mathsf{P}} 0$ for all $g \in \{1, \dots, G\}$ and $\min_{g\in\{1,\dots,G\}} \|\widehat{\beta}_{\widetilde{g}}(\tau_k) - \beta_g^0(\tau_k)\| \xrightarrow{\mathsf{P}} 0$ for all $\widetilde{g} \in \{1, \dots, G\}$, respectively. This

completes the proof. ∎

Proof of Corollary 1. We first prove (3.3). The proof of Lemma 2 shows that there exists a one-to-one mapping $\sigma : \{1, \ldots, G\} \to \{1, \ldots, G\}$ such that

$$\|\widehat{\beta}_{\sigma(g)} - \beta_g^0\| \xrightarrow{\mathsf{P}} 0.$$

Because the objective function in (2.2) is invariant to relabeling the g-group labels, one may take $\sigma(g) = g$ for $g \in \{1, \ldots, G\}$. This shows (3.3). Similarly, we can show that

$$\max_{h \in \{1,\ldots,H\}} |\widehat{\alpha}_h + \widehat{\lambda}_t - (\alpha_h^0 + \lambda_t^0)| \xrightarrow{\mathsf{P}} 0. \tag{A.10}$$

Next we show (3.4). Put $\phi_{h,t} := \alpha_h + \lambda_t$, $\widehat{\phi}_{h,t} := \widehat{\alpha}_h + \widehat{\lambda}_t$ and $\phi_{h,t}^0 := \alpha_h^0 + \lambda_t^0$ for $h \in \{1, \ldots, H\}$ and $t = 1, \ldots, T$. Let $\Phi_t := (\phi_{1,t}, \ldots, \phi_{H,t})'$. Recall that $\widehat{\Phi}_t = (\widehat{\phi}_{1,t}, \ldots, \widehat{\phi}_{H,t})'$ and $\Phi_t^0 = (\phi_{1,t}^0, \ldots, \phi_{H,t}^0)'$. We have $\widehat{\Phi}_t = \underset{\Phi_t \in \Re^H}{\arg\min} \sum_{h \in \{1,\ldots,H\}} \sum_{i:\widehat{h}_i = h} \rho_{\tau_k}(y_{it} - \phi_{h,t} - x_{it}'\widehat{\beta}_{\widehat{g}_i})$. Hence, for all $\delta > 0$,

$$\begin{aligned}
&\left\{\max_{1 \le t \le T} \|\widehat{\Phi}_t - \Phi_t^0\| > \delta\right\} \\
\subset &\left\{\sum_{h \in \{1,\ldots,H\}} \sum_{i:\widehat{h}_i = h} \rho_{\tau_k}(y_{it} - \phi_{h,t} - x_{it}'\widehat{\beta}_{\widehat{g}_i}) \le \sum_{h \in \{1,\ldots,H\}} \sum_{i:\widehat{h}_i = h} \rho_{\tau_k}(y_{it} - \phi_{h,t}^0 - x_{it}'\widehat{\beta}_{\widehat{g}_i}), \|\Phi_t - \Phi_t^0\| > \delta, \right. \\
&\left. 1 \le \exists\, t \le T\right\}.
\end{aligned} \tag{A.11}$$

Define

$$\widehat{\Delta}(\Phi_t) = \frac{1}{N} \sum_{h \in \{1,\ldots,H\}} \sum_{i:\widehat{h}_i = h} \left(\rho_{\tau_k}(y_{it} - \phi_{h,t} - x_{it}'\widehat{\beta}_{\widehat{g}_i}) - \rho_{\tau_k}(y_{it} - \phi_{h,t}^0 - x_{it}'\widehat{\beta}_{\widehat{g}_i})\right).$$

Let $r_t := \delta / \|\Phi_t - \Phi_t^0\|$ and $\overline{\Phi}_t := r_t \Phi_t + (1 - r_t)\Phi_t^0$. From (A.11), we have $0 < r_t < 1$, $\|\overline{\Phi}_t - \Phi_t^0\| = \delta$ and $\widehat{\Delta}(\overline{\Phi}_t) \le r_t \widehat{\Delta}(\Phi_t)$ due to the convexity of function $\widehat{\Delta}$. Thus, combining with (A.11) yields that

$$\left\{\max_{1 \le t \le T} \|\widehat{\Phi}_t - \Phi_t^0\| > \delta\right\} \subset \left\{\widehat{\Delta}(\Phi_t) \le 0,\ \|\Phi_t - \Phi_t^0\| \le \delta,\ 1 \le \exists\, t \le T\right\}. \tag{A.12}$$

Denoting

$$\Delta^0(\Phi_t) := \frac{1}{N} \sum_{h \in \{1,\ldots,H\}} \sum_{i:h_i^0 = h} \left(\rho_{\tau_k}(y_{it} - \phi_{h,t} - x_{it}'\beta_{g_i^0}^0) - \rho_{\tau_k}(y_{it} - \phi_{h,t}^0 - x_{it}'\beta_{g_i^0}^0)\right),$$

we divide $\widehat{\Delta}(\Phi_t)$ into three terms.

$$\widehat{\Delta}(\Phi_t) = \left\{\widehat{\Delta}(\Phi_t) - \Delta^0(\Phi_t)\right\} + \left\{\Delta^0(\Phi_t) - \mathsf{E}[\Delta^0(\Phi_t)]\right\} + \mathsf{E}[\Delta^0(\Phi_t)].$$

Note that $\mathsf{E}[\Delta^0(\Phi_t)] > 0$ by Assumption 1(*iii*) provided $\|\Phi_t - \Phi_t^0\| \le \delta$. Thus, following from (A.12)

we have

$$\left\{\max_{1\le t\le T}\|\widehat{\Phi}_t-\Phi_t^0\|>\delta\right\}\subset\left\{\max_{1\le t\le T}\sup_{\|\Phi_t-\Phi_t^0\|\le\delta}\left|\left\{\widehat{\Delta}(\Phi_t)-\Delta^0(\Phi_t)\right\}+\left\{\Delta^0(\Phi_t)-\mathsf{E}[\Delta^0(\Phi_t)]\right\}\right|>\epsilon_\delta\right\}$$

$$\subset\left\{\max_{1\le t\le T}\sup_{\|\Phi_t-\Phi_t^0\|\le\delta}\left|\widehat{\Delta}(\Phi_t)-\Delta^0(\Phi_t)\right|>\epsilon_\delta/2\right\}\cup\left\{\max_{1\le t\le T}\sup_{\|\Phi_t-\Phi_t^0\|\le\delta}\left|\Delta^0(\Phi_t)-\mathsf{E}[\Delta^0(\Phi_t)]\right|>\epsilon_\delta/2\right\} \quad (A.13)$$

for some $\epsilon_\delta>0$. First,

$$\max_{1\le t\le T}\sup_{\|\Phi_t-\Phi_t^0\|\le\delta}\left|\widehat{\Delta}(\Phi_t)-\Delta^0(\Phi_t)\right|\le\frac{1}{N}\sum_{i=1}^N C_0(1+\|x_{it}\|)\big(|\alpha_{\widehat{h}_i}-\alpha_{h_i^0}|+|\alpha^0_{\widehat{h}_i}-\alpha^0_{h_i^0}|+2\|\widehat{\beta}_{\widehat{g}_i}-\beta^0_{g_i^0}\|\big)$$

$$\le C_0(1+\sup_{i\ge1}\|x_{i1}\|)\big(\max_{i=1,\dots,N}|\alpha_{\widehat{h}_i}-\alpha_{h_i^0}|+\max_{i=1,\dots,N}|\alpha^0_{\widehat{h}_i}-\alpha^0_{h_i^0}|+\max_{i=1,\dots,N}2\|\widehat{\beta}_{\widehat{g}_i}-\beta^0_{g_i^0}\|\big), \quad (A.14)$$

where C_0 is a universal constant. Moreover, by (A.10) and (3.2) of Theorem 1 we have

$$\max_{i=1,\dots,N}|\alpha^0_{\widehat{h}_i}-\alpha^0_{h_i^0}|\le\max_{h\in\{1,\dots,H\}}|\alpha^0_h+\lambda^0_t-(\widehat{\alpha}_h+\widehat{\lambda}_t)|+\max_{i=1,\dots,N}|\widehat{\alpha}_{\widehat{h}_i}+\widehat{\lambda}_t-(\alpha^0_{h_i^0}+\lambda^0_t)|=o_{\mathsf{P}}(1), \quad (A.15)$$

and hence

$$\begin{aligned}\max_{i=1,\dots,N}|\alpha_{\widehat{h}_i}-\alpha_{h_i^0}|&\le\max_{i=1,\dots,N}|\alpha_{\widehat{h}_i}-\alpha^0_{\widehat{h}_i}|+\max_{i=1,\dots,N}|\alpha^0_{\widehat{h}_i}-\alpha^0_{h_i^0}|+\max_{i=1,\dots,N}|\alpha^0_{h_i^0}-\alpha_{h_i^0}|\\&\le2\|\Phi_t-\Phi_t^0\|+\max_{i=1,\dots,N}|\alpha^0_{\widehat{h}_i}-\alpha^0_{h_i^0}|\le2\delta+o_{\mathsf{P}}(1)\end{aligned} \quad (A.16)$$

as N and $T\to\infty$. Combining (A.14), (A.15) and (A.16), for $\delta<\frac{\epsilon_\delta}{6C_0(1+M)}$ where $\sup_{i\ge1}\|x_{i1}\|\le M$ by Assumption 1(*ii*), we obtain that

$$\max_{1\le t\le T}\sup_{\|\Phi_t-\Phi_t^0\|\le\delta}\left|\widehat{\Delta}(\Phi_t)-\Delta^0(\Phi_t)\right|\xrightarrow{\mathsf{P}}0. \quad (A.17)$$

For the second event in right side of (A.13), using the similar arguments for proving (A.3) leads to

$$\max_{1\le t\le T}\mathsf{P}\left(\sup_{\|\Phi_t-\Phi_t^0\|\le\delta}\left|\Delta^0(\Phi_t)-\mathsf{E}[\Delta^0(\Phi_t)]\right|>\epsilon_\delta/2\right)=o(T^{-1})$$

from which we have

$$\max_{1\le t\le T}\sup_{\|\Phi_t-\Phi_t^0\|\le\delta}\left|\Delta^0(\Phi_t)-\mathsf{E}[\Delta^0(\Phi_t)]\right|\xrightarrow{\mathsf{P}}0. \quad (A.18)$$

Thus, combining (A.13), (A.17) and (A.18), we have $\max_{1\le t\le T}\|\widehat{\Phi}_t-\Phi_t^0\|\xrightarrow{\mathsf{P}}0$. This implies (3.4) provided that $H<\infty$. The proof is complete. ∎

Proof of Theorem 2. Define $W_{i,gh}(\boldsymbol{\theta}(\boldsymbol{\tau}))=T^{-1}\sum_{t=1}^T W_{it,gh}(\boldsymbol{\theta}(\boldsymbol{\tau}))$. For $(g,h)\neq(g_i^0,h_i^0)$, because

$I\{\widehat{g}_i = g, \widehat{h}_i = h\} \leq I\{W_{i,gh}(\boldsymbol{\theta}(\boldsymbol{\tau})) \leq 0\}$, we have

$$\frac{1}{N}\sum_{i=1}^{N}\big(1 - I\{\widehat{g}_i = g_i, \widehat{h}_i = h_i\}\big) = \frac{1}{N}\sum_{i=1}^{N}\sum_{g\in\{1,\ldots,G\}}\sum_{h\in\{1,\ldots,H\}} I\left\{(g,h) \neq (g_i^0, h_i^0)\right\} I\{\widehat{g}_i = g, \widehat{h}_i = h\}$$

$$\leq \frac{1}{N}\sum_{i=1}^{N}\sum_{g\in\{1,\ldots,G\}}\sum_{h\in\{1,\ldots,H\}} I\left\{(g,h) \neq (g_i^0, h_i^0)\right\} I\{W_{i,gh}(\boldsymbol{\theta}(\boldsymbol{\tau})) \leq 0\} \tag{A.19}$$

where $\boldsymbol{\theta}(\boldsymbol{\tau}) \in \mathbb{N}_\delta$. Let $\boldsymbol{\theta}^0(\boldsymbol{\tau})$ denote the true value of $\boldsymbol{\theta}(\boldsymbol{\tau})$. Note that

$$\begin{aligned} &I\{W_{i,gh}(\boldsymbol{\theta}(\boldsymbol{\tau})) \leq 0\} \leq I\{W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau})) \leq \left|W_{i,gh}(\boldsymbol{\theta}(\boldsymbol{\tau})) - W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right|\} \\ \leq &I\{W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau})) \leq 2KC_0\big(1 + \sup_{i\geq 1}\|x_{i1}\|\big)\cdot 2\delta\} \leq I\{W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau})) \leq 4KC_0\big(1+M\big)\cdot\delta\} \end{aligned} \tag{A.20}$$

where C_0 is a universal constant. Thus, combining (A.19) and (A.20) yields that

$$\begin{aligned} &\sup_{\boldsymbol{\theta}(\boldsymbol{\tau})\in\mathbb{N}_\delta}\frac{1}{N}\sum_{i=1}^{N}\big(1 - I\{\widehat{g}_i = g_i, \widehat{h}_i = h_i\}\big) \\ \leq &\frac{1}{N}\sum_{i=1}^{N}\sum_{g\in\{1,\ldots,G\}}\sum_{h\in\{1,\ldots,H\}} I\left\{(g,h) \neq (g_i^0, h_i^0)\right\} I\{W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau})) \leq 4KC_0\big(1+M\big)\delta\} \\ \leq &\sum_{g\in\{1,\ldots,G\}}\sum_{h\in\{1,\ldots,H\}}\frac{1}{N}\sum_{i=1}^{N} I\{W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau})) \leq 4KC_0\big(1+M\big)\delta\}. \end{aligned} \tag{A.21}$$

Denote $Z_i(g,h) := I\{W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau})) \leq 4KC_0\big(1+M\big)\delta\}$. Note that

$$\begin{aligned} \mathsf{E}[Z_i(g,h)] &= \mathsf{P}\big(W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau})) \leq 4KC_0\big(1+M\big)\delta\big) \\ &= \mathsf{P}\left(\frac{W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau})) - \mathsf{E}[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))]}{\sqrt{\mathsf{Var}\left[W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]}} \leq \frac{4KC_0\big(1+M\big)\delta - \mathsf{E}[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))]}{\sqrt{\mathsf{Var}\left[W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]}}\right), \end{aligned} \tag{A.22}$$

where

$$\mathsf{E}[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))] = \sum_{k=1}^{K}\mathsf{E}\left[\int_0^{(\alpha_h^0(\tau_k)-\alpha_{h_i^0}^0(\tau_k))+x_{i1}'(\beta_g^0(\tau_k)-\beta_{g_i^0}^0(\tau_k))}\big(F_{i,\tau_k}(u|x_{i1}) - \tau_k\big)\,du\right] > 0 \tag{A.23}$$

due to Assumption 1(v). By Assumption 2(i), applying Corollary 2.3 of Peligrad (1996) to the strongly mixing process, $\{W_{it,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau})), t \geq 1\}$, we obtain that for any $x \in \mathbb{R}$,

$$\lim_{T\to\infty}\mathsf{P}\left(\frac{W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau})) - \mathsf{E}[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))]}{\sqrt{\mathsf{Var}\left[W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]}} \leq x\right) = \Phi(x)$$

where Φ denotes the standard normal distribution. By Assumption 2(ii), the pointwise convergence

with respect to x above further implies

$$\lim_{T\to\infty}\sup_{x\in\mathbb{R}}\left|\mathsf{P}\left(\frac{W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))-\mathsf{E}[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))]}{\sqrt{\mathsf{Var}\left[W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]}}\leq x\right)-\Phi(x)\right|=0. \tag{A.24}$$

Moreover, from Lemma 1 of Bradley (1992) we have

$$\begin{aligned}T\,\mathsf{Var}\left[W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]&\leq\frac{1+\rho^*(1)}{1-\rho^*(1)}\mathsf{Var}\left[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]\\&\leq\frac{1+\rho^*(1)}{1-\rho^*(1)}\sup_{i\geq1}\max_{(g,h)}\mathsf{Var}\left[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]<\infty,\end{aligned} \tag{A.25}$$

by Assumption 1(*ii*) and Assumption 2(*i*). Hence, the convergence of (A.24) is uniform over i; i.e.,

$$\lim_{T\to\infty}\sup_{i\geq1}\sup_{x\in\mathbb{R}}\left|\mathsf{P}\left(\frac{W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))-\mathsf{E}[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))]}{\sqrt{\mathsf{Var}\left[W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]}}\leq x\right)-\Phi(x)\right|=0. \tag{A.26}$$

For any $0<\epsilon<1$, put $\delta=\epsilon\inf_{i\geq1}\min_{(g,h)\neq(g_i,h_i)}\frac{\mathsf{E}[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))]}{4KC_0(1+M)}$. It hence follows from (A.22) (A.25) and (A.26) that

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^N\mathsf{E}[Z_i(g,h)]\\=&\frac{1}{N}\sum_{i=1}^N\mathsf{P}\left(\frac{W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))-\mathsf{E}[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))]}{\sqrt{\mathsf{Var}\left[W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]}}\leq\frac{4KC_0\big(1+M\big)\delta-\mathsf{E}[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))]}{\sqrt{\mathsf{Var}\left[W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]}}\right)\\=&\frac{1}{N}\sum_{i=1}^N\Phi\left(\sqrt{T}\frac{4KC_0\big(1+M\big)\delta-\mathsf{E}[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))]}{\sqrt{T\mathsf{Var}\left[W_{i,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]}}\right)+o(1)\\\leq&\Phi\left(\sqrt{T}(\epsilon-1)\sqrt{\frac{1-\rho^*(1)}{1+\rho^*(1)}}\frac{\inf_{i\geq1}\min_{(g,h)\neq(g_i,h_i)}\mathsf{E}[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))]}{\sup_{i\geq1}\max_{(g,h)}\sqrt{\mathsf{Var}\left[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]}}\right)+o(1)\\=&O\left(T^{-1/2}\exp\left(-T\frac{(1-\epsilon)^2}{2}\frac{1-\rho^*(1)}{1+\rho^*(1)}\frac{\inf_{i\geq1}\min_{(g,h)\neq(g_i,h_i)}\mathsf{E}^2[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))]}{\sup_{i\geq1}\max_{(g,h)}\mathsf{Var}\left[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]}\right)\right)\end{aligned} \tag{A.27}$$

as $T\to\infty$, where the last equality is due to the Mills ratio and

$$\begin{aligned}&\mathsf{Var}\left[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))\right]\\=&\sum_{k=1}^K\sum_{l=1}^K\mathsf{E}\Bigg[\int_0^{(\alpha_h^0(\tau_k)-\alpha_{h_i^0}^0(\tau_k))+x_{i1}'(\beta_g^0(\tau_k)-\beta_{g_i^0}^0(\tau_k))}\int_0^{(\alpha_h^0(\tau_l)-\alpha_{h_i^0}^0(\tau_l))+x_{i1}'(\beta_g^0(\tau_l)-\beta_{g_i^0}^0(\tau_l))}\\&\qquad\mathsf{E}\left[\left(I(\epsilon_{i1}(\tau_k)\leq u)-\tau_k\right)\left(I(\epsilon_{i1}(\tau_l)\leq s)-\tau_l\right)|x_{i1}\right]\,duds\Bigg]-\mathsf{E}^2[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))].\end{aligned} \tag{A.28}$$

Since $\sup_{i\geq 1}|Z_i(g,h)|\leq 1$, by Lyapunov CLT we have

$$\sqrt{N}\,\frac{N^{-1}\sum_{i=1}^N Z_i(g,h)-N^{-1}\sum_{i=1}^N \mathsf{E}[Z_i(g,h)]}{\sqrt{N^{-1}\sum_{i=1}^N \mathsf{Var}\,[Z_i(g,h)]}}\xrightarrow{D}N(0,1),$$

as $N\to\infty$. This implies that

$$\frac{1}{N}\sum_{i=1}^N Z_i(g,h)=\frac{1}{N}\sum_{i=1}^N \mathsf{E}[Z_i(g,h)]+O_{\mathsf{P}}(1)\cdot\frac{1}{N}\sqrt{\sum_{i=1}^N \mathsf{Var}\,[Z_i(g,h)]}. \tag{A.29}$$

Noting that $\mathsf{Var}\,[Z_i(g,h)]=\mathsf{E}[Z_i(g,h)](1-\mathsf{E}[Z_i(g,h)])$, it follows from (A.27) that

$$\begin{aligned}\frac{1}{N}\sqrt{\sum_{i=1}^N \mathsf{Var}\,[Z_i(g,h)]}&=\frac{1}{N}\sqrt{\sum_{i=1}^N \mathsf{E}[Z_i(g,h)](1-\mathsf{E}[Z_i(g,h)])}\\ =&\frac{1}{N}\sqrt{\sum_{i=1}^N \mathsf{E}[Z_i(g,h)]}\cdot(1+o(1))=\frac{1}{\sqrt{N}}O(T^{-1/4}\exp(-T\zeta/2))\end{aligned} \tag{A.30}$$

as $T\to\infty$, where $\zeta=\frac{(1-\epsilon)^2}{2}\cdot\frac{1-\rho^*(1)}{1+\rho^*(1)}\cdot\frac{\inf_{i\geq 1}\min_{(g,h)\neq(g_i,h_i)}\mathsf{E}^2[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))]}{\sup_{i\geq 1}\max_{(g,h)}\mathsf{Var}\,[W_{i1,gh}(\boldsymbol{\theta}^0(\boldsymbol{\tau}))]}$. Hence, combining (A.21), (A.27), (A.29) and (A.30) obtains that

$$\begin{aligned}&\sup_{\boldsymbol{\theta}(\boldsymbol{\tau})\in\mathbb{N}_\delta}\frac{1}{N}\sum_{i=1}^N\left(1-I\{\widehat{g}_i=g_i,\widehat{h}_i=h_i\}\right)\leq\sum_{g\in\{1,\ldots,G\}}\sum_{h\in\{1,\ldots,H\}}\frac{1}{N}\sum_{i=1}^N Z_i(g,h)\\ =&O(T^{-1/2}\exp(-T\zeta))+O_{\mathsf{P}}(N^{-1/2}T^{-1/4}\exp(-T\zeta/2))\end{aligned}$$

as N and $T\to\infty$, provided that $G<\infty$ and $H<\infty$. This completes the proof of Theorem 2. ∎

Proof of Corollary 2. Following from the proof of Theorem 2 we have

$$\begin{aligned}&\mathsf{P}\left(\bigcup_{i=1}^N\{(\widehat{g}_i,\widehat{h}_i)\neq(g_i^0,h_i^0)\}\right)\leq\mathsf{P}\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{\tau})\notin\mathbb{N}_\delta\right)+\sum_{i=1}^N\mathsf{P}\left((\widehat{g}_i,\widehat{h}_i)\neq(g_i^0,h_i^0),\widehat{\boldsymbol{\theta}}(\boldsymbol{\tau})\in\mathbb{N}_\delta\right)\\ \leq&\mathsf{P}\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{\tau})\notin\mathbb{N}_\delta\right)+\sum_{g\in\{1,\ldots,G\}}\sum_{h\in\{1,\ldots,H\}}\sum_{i=1}^N\mathsf{E}[Z_i(g,h)]\leq\mathsf{P}\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{\tau})\notin\mathbb{N}_\delta\right)+O\left(NT^{-1/2}\exp(-\zeta T)\right).\end{aligned}$$

Moreover, Corollary 1 implies that $\mathsf{P}\big(\widehat{\boldsymbol{\theta}}(\boldsymbol{\tau})\notin\mathbb{N}_\delta\big)\to 0$. Thus, under the condition $N/(T^{1/2}\exp(T\epsilon))\to 0$ for all $\epsilon>0$, we obtain that

$$\mathsf{P}\left(\bigcup_{i=1}^N\{(\widehat{g}_i,\widehat{h}_i)\neq(g_i^0,h_i^0)\}\right)\to 0$$

as N and $T\to\infty$. This completes the proof. ∎

Proof of Corollary 3. Conditional on $\boldsymbol{\theta}(\boldsymbol{\tau}) \in \mathbb{N}_\delta$ for some $\delta > 0$ small enough, define

$$\widehat{Q}(\boldsymbol{\theta}(\boldsymbol{\tau})) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{K}\rho_{\tau_k}\big(y_{it} - \alpha_{\widehat{h}_i}(\tau_k) - \lambda_t(\tau_k) - x_{it}'\beta_{\widehat{g}_i}(\tau_k)\big),$$

and

$$\widetilde{Q}(\boldsymbol{\theta}(\boldsymbol{\tau})) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{K}\rho_{\tau_k}\big(y_{it} - \alpha_{h_i^0}(\tau_k) - \lambda_t(\tau_k) - x_{it}'\beta_{g_i^0}(\tau_k)\big).$$

Theorem 2 implies that

$$\sup_{\boldsymbol{\theta}(\boldsymbol{\tau})\in\mathbb{N}_\delta}\left|\widehat{Q}(\boldsymbol{\theta}(\boldsymbol{\tau})) - \widetilde{Q}(\boldsymbol{\theta}(\boldsymbol{\tau}))\right| = O_{\mathsf{P}}\left(\frac{\exp(-T\zeta/2)}{N^{1/2}T^{1/4}} + \frac{\exp(-T\zeta)}{T^{1/2}}\right).$$

From (2.2) and (3.8), we have that $\widehat{\boldsymbol{\theta}}(\boldsymbol{\tau}) = \arg\min_{\boldsymbol{\theta}(\boldsymbol{\tau})\in\mathbb{N}_\delta}\widehat{Q}(\boldsymbol{\theta}(\boldsymbol{\tau}))$ and $\widetilde{\boldsymbol{\theta}}(\boldsymbol{\tau}) = \arg\min_{\boldsymbol{\theta}(\boldsymbol{\tau})\in\mathbb{N}_\delta}\widetilde{Q}(\boldsymbol{\theta}(\boldsymbol{\tau}))$. We can use the similar arguments as in the proof of (S1.13) in Zhang et al. (2019b) or the proof of (B.17) in Bonhomme and Manresa (2015a) to show that

$$\widetilde{Q}(\widehat{\boldsymbol{\theta}}(\boldsymbol{\tau})) - \widetilde{Q}(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\tau})) = O_{\mathsf{P}}\left(\frac{\exp(-T\zeta/2)}{N^{1/2}T^{1/4}} + \frac{\exp(-T\zeta)}{T^{1/2}}\right), \tag{A.31}$$

as N and $T \to \infty$. For notational simplicity, we suppress the dependence of parameters on τ_k wherever no confusion arises. Then using the Knight (1998)'s identity, we depict the left side of (A.31) as

$$\begin{aligned}
\widetilde{Q}(\widehat{\boldsymbol{\theta}}(\boldsymbol{\tau})) - \widetilde{Q}(\widetilde{\boldsymbol{\theta}}(\boldsymbol{\tau})) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{K}\left[\rho_{\tau_k}\big(y_{it} - \widehat{\alpha}_{h_i^0} - \widehat{\lambda}_t - x_{it}'\widehat{\beta}_{g_i^0}\big) - \rho_{\tau_k}\big(y_{it} - \widetilde{\alpha}_{h_i^0} - \widetilde{\lambda}_t - x_{it}'\widetilde{\beta}_{g_i^0}\big)\right] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=1}^{K}\int_{\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t-(\alpha^0_{h_i^0}+\lambda^0_t)+x_{it}'(\widetilde{\beta}_{g_i^0}-\beta^0_{g_i^0})}^{\widehat{\alpha}_{h_i^0}+\widehat{\lambda}_t-(\alpha^0_{h_i^0}+\lambda^0_t)+x_{it}'(\widehat{\beta}_{g_i^0}-\beta^0_{g_i^0})}\{I(\varepsilon_{it}\le u) - \tau_k\}\,du \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\int_0^{\widehat{\alpha}_{h_i^0}+\widehat{\lambda}_t-(\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t)+x_{it}'(\widehat{\beta}_{g_i^0}-\widetilde{\beta}_{g_i^0})}\left\{I\big(\varepsilon_{it}\le u + \widetilde{\alpha}_{h_i^0} + \widetilde{\lambda}_t - (\alpha^0_{h_i^0}+\lambda^0_t) + x_{it}'(\widetilde{\beta}_{g_i^0} - \beta^0_{g_i^0})\big) - \tau_k\right\}du \\
&=: \frac{1}{N}\sum_{i=1}^{N}A_i,
\end{aligned} \tag{A.32}$$

Since A_i in (A.32) is bounded a.s. and independent across i, from Lyapunov CLT, we obtain that, as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N}A_i = \frac{1}{N}\sum_{i=1}^{N}\mathsf{E}(A_i|\widehat{\boldsymbol{\theta}}(\boldsymbol{\tau}),\widetilde{\boldsymbol{\theta}}(\boldsymbol{\tau})) + O_{\mathsf{P}}\left(\frac{1}{N}\sqrt{\sum_{i=1}^{N}\mathsf{Var}\left(A_i|\widehat{\boldsymbol{\theta}}(\boldsymbol{\tau}),\widetilde{\boldsymbol{\theta}}(\boldsymbol{\tau})\right)}\right). \tag{A.33}$$

We examine the two terms on the right side of (A.33) in turn. Without loss of generality, we assume $K = 1$. Again, the dependence on $\boldsymbol{\tau}$ is suppressed when no confusion arises. For the first term, using

the fact that $F_{i,\tau_1}(0|x_{it}) = \tau_1$, we note that

$$
\begin{aligned}
&\mathsf{E}(A_i|\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}) \\
=&\frac{1}{T}\sum_{t=1}^{T}\mathsf{E}\Big[\int_0^{\widehat{\alpha}_{h_i^0}+\widehat{\lambda}_t-(\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t)+x_{i1}'(\widehat{\beta}_{g_i^0}-\widetilde{\beta}_{g_i^0})} F_{i,\tau_1}(u+\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t-(\alpha_{h_i^0}^0+\lambda_t^0)+x_{it}'(\widetilde{\beta}_{g_i^0}-\beta_{g_i^0}^0)|x_{it})-\tau_1\,du\Big|\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}\Big] > 0.
\end{aligned}
$$

Using Taylor's theorem and Assumption 3, we can further show that

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\mathsf{E}(A_i|\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}) \\
&\geq \frac{1}{T}\sum_{t=1}^{T}\sum_{h,g}\frac{1}{N}\sum_{i:h_i^0=h,g_i^0=g} 2^{-1}\big(\widehat{\alpha}_h+\widehat{\lambda}_t-(\widetilde{\alpha}_h+\widetilde{\lambda}_t),(\widehat{\beta}_g-\widetilde{\beta}_g)'\big) \\
&\qquad \mathsf{E}\big[f_{i,\tau_1}\big(\widetilde{\alpha}_h+\widetilde{\lambda}_t-(\alpha_h^0+\lambda_t^0)+x_{it}'(\widetilde{\beta}_g-\beta_g^0)|x_{it}\big)(1,x_{it}')'(1,x_{it}')\big]\big(\widehat{\alpha}_h+\widehat{\lambda}_t-(\widetilde{\alpha}_h+\widetilde{\lambda}_t),(\widehat{\beta}_g-\widetilde{\beta}_g)'\big)' \\
&\qquad + o_{\mathsf{P}}\Big(\sum_{g\in\{1,\ldots,G\}}\|\widehat{\beta}_g-\widetilde{\beta}_g\|^2+T^{-1}\sum_{t=1}^{T}\sum_{h\in\{1,\ldots,H\}}|\widehat{\alpha}_h+\widehat{\lambda}_t-(\widetilde{\alpha}_h+\widetilde{\lambda}_t)|^2\Big) \\
&\geq \frac{1}{T}\sum_{t=1}^{T}\sum_{h,g}\underline{\omega}_{t,\tau_1}(h,g)\cdot 2^{-1}\left(\|\widehat{\beta}_g-\widetilde{\beta}_g\|^2+|\widehat{\alpha}_h+\widehat{\lambda}_t-(\widetilde{\alpha}_h+\widetilde{\lambda}_t)|^2\right) \\
&\quad + o_{\mathsf{P}}\Big(\sum_{g\in\{1,\ldots,G\}}\|\widehat{\beta}_g-\widetilde{\beta}_g\|^2+T^{-1}\sum_{t=1}^{T}\sum_{h\in\{1,\ldots,H\}}|\widehat{\alpha}_h+\widehat{\lambda}_t-(\widetilde{\alpha}_h+\widetilde{\lambda}_t)|^2\Big) \\
&\geq 2^{-1}\min_{h,g}\omega_{\tau_1}(h,g)\cdot\Big(H\sum_{g\in\{1,\ldots,G\}}\|\widehat{\beta}_g-\widetilde{\beta}_g\|^2+G\frac{1}{T}\sum_{t=1}^{T}\sum_{h\in\{1,\ldots,H\}}|\widehat{\alpha}_h+\widehat{\lambda}_t-(\widetilde{\alpha}_h+\widetilde{\lambda}_t)|^2\Big) \\
&\qquad + o_{\mathsf{P}}\Big(\sum_{g\in\{1,\ldots,G\}}\|\widehat{\beta}_g-\widetilde{\beta}_g\|^2+T^{-1}\sum_{t=1}^{T}\sum_{h\in\{1,\ldots,H\}}|\widehat{\alpha}_h+\widehat{\lambda}_t-(\widetilde{\alpha}_h+\widetilde{\lambda}_t)|^2\Big) \\
&\geq C\Big(\sum_{g\in\{1,\ldots,G\}}\|\widehat{\beta}_g-\widetilde{\beta}_g\|^2+\frac{1}{T}\sum_{t=1}^{T}\sum_{h\in\{1,\ldots,H\}}|\widehat{\alpha}_h+\widehat{\lambda}_t-(\widetilde{\alpha}_h+\widetilde{\lambda}_t)|^2\Big) \qquad \text{(A.34)}
\end{aligned}
$$

for some fixed constant $C > 0$, with probability approaching 1 as N and $T \to \infty$. Assumption 3 also implies that

$$
\frac{1}{N}\sum_{i=1}^{N}\mathsf{E}(A_i|\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}) \leq C'\Big(\sum_{g\in\{1,\ldots,G\}}\|\widehat{\beta}_g-\widetilde{\beta}_g\|^2+\frac{1}{T}\sum_{t=1}^{T}\sum_{h\in\{1,\ldots,H\}}|\widehat{\alpha}_h+\widehat{\lambda}_t-(\widetilde{\alpha}_h+\widetilde{\lambda}_t)|^2\Big) \qquad \text{(A.35)}
$$

for some constant $C' > 0$. Combining (A.34) and (A.35), we have

$$
\frac{1}{N}\sum_{i=1}^{N}\mathsf{E}(A_i|\widehat{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}}) = O_{\mathsf{P}}\Big(\sum_{g\in\{1,\ldots,G\}}\|\widehat{\beta}_g-\widetilde{\beta}_g\|^2+\frac{1}{T}\sum_{t=1}^{T}\sum_{h\in\{1,\ldots,H\}}|\widehat{\alpha}_h+\widehat{\lambda}_t-(\widetilde{\alpha}_h+\widetilde{\lambda}_t)|^2\Big). \qquad \text{(A.36)}
$$

Next, we study the second term on the right side of (A.33). Note that

$$\mathsf{Var}\left(A_i|\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\right) = \mathsf{Var}\left[\mathsf{E}(A_i|\{x_{it}\},\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})\big|\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\right] + \mathsf{E}\big[\mathsf{Var}\left(A_i|\{x_{it}\},\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\right)\big|\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\big], \tag{A.37}$$

where

$$\begin{aligned}
&\mathsf{Var}\left[\mathsf{E}(A_i|\{x_{it}\},\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})\big|\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\right]\\
=&\mathsf{Var}\left[\frac{1}{T}\sum_{t=1}^{T}\int_0^{\widehat{\alpha}_{h_i^0}+\widehat{\lambda}_t-(\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t)+x_{it}'(\widehat{\beta}_{g_i^0}-\widetilde{\beta}_{g_i^0})} F_{i,\tau_1}\left(u+\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t-(\alpha_{h_i^0}^0+\lambda_t^0)+x_{it}'(\widetilde{\beta}_{g_i^0}-\beta_{g_i^0}^0)|x_{it}\right)-\tau_1\,du\big|\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\right]\\
=&O_{\mathsf{P}}\left(\frac{1}{T}\sum_{t=1}^{T}\mathsf{E}\left[\left(\int_0^{(\widehat{\alpha}_{h_i^0}+\widehat{\lambda}_t)-(\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t)+x_{it}'(\widehat{\beta}_{g_i^0}-\widetilde{\beta}_{g_i^0})} F_{i,\tau_1}(u+\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t-(\alpha_{h_i^0}^0+\lambda_t^0)+x_{it}'(\widetilde{\beta}_{g_i^0}-\beta_{g_i^0}^0)|x_{it})-\tau_1\,du\right)^2\big|\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\right]\right)\\
=&O_{\mathsf{P}}\left(\frac{1}{T}\sum_{t=1}^{T}\mathsf{E}\left[\left(\widehat{\alpha}_{h_i^0}+\widehat{\lambda}_t-(\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t)+x_{it}'(\widehat{\beta}_{g_i^0}-\widetilde{\beta}_{g_i^0})\right)^4\big|\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\right]\right)\\
=&O_{\mathsf{P}}\left(\frac{1}{T}\sum_{t=1}^{T}\left|\widehat{\alpha}_{h_i^0}+\widehat{\lambda}_t-(\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t)\right|^4+\left\|\widehat{\beta}_{g_i^0}-\widetilde{\beta}_{g_i^0}\right\|^4\right),
\end{aligned} \tag{A.38}$$

and

$$\begin{aligned}
&\mathsf{E}\big[\mathsf{Var}\left(A_i|\{x_{it}\},\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\big|\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\right] = O_{\mathsf{P}}\big(\mathsf{E}\big[\mathsf{E}(A_i^2|\{x_{it}\},\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}})\big|\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\big]\big)\\
=&O_{\mathsf{P}}\left(\frac{1}{T}\sum_{t=1}^{T}\mathsf{E}\left\{\mathsf{E}\left[\left(\int_0^{\widehat{\alpha}_{h_i^0}+\widehat{\lambda}_t-(\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t)+x_{it}'(\widehat{\beta}_{g_i^0}-\widetilde{\beta}_{g_i^0})} I\big(\varepsilon_{it}\le u+\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t-(\alpha_{h_i^0}^0+\lambda_t^0)+x_{it}'(\widetilde{\beta}_{g_i^0}-\beta_{g_i^0}^0)\big)-\tau_1\,du\right)^2\right.\right.\\
&\qquad\left.\left.|\{x_{it}\},\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\right]\big|\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\right\}\right)\\
=&O_{\mathsf{P}}\left(\frac{1}{T}\sum_{t=1}^{T}\mathsf{E}\left[\int_0^{\widehat{\alpha}_{h_i^0}+\widehat{\lambda}_t-(\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t)+x_{it}'(\widehat{\beta}_{g_i^0}-\widetilde{\beta}_{g_i^0})}\int_0^{\widehat{\alpha}_{h_i^0}+\widehat{\lambda}_t-(\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t)+x_{it}'(\widehat{\beta}_{g_i^0}-\widetilde{\beta}_{g_i^0})}\left\{F_{i,\tau_1}(u+\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t-(\alpha_{h_i^0}^0+\lambda_t^0)\right.\right.\\
&\left.\left.+x_{it}'(\widetilde{\beta}_{g_i^0}-\beta_{g_i^0}^0)|x_{it})-\tau_1\right\}\left\{F_{i,\tau_1}(s+\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t-(\alpha_{h_i^0}^0+\lambda_t^0)+x_{it}'(\widetilde{\beta}_{g_i^0}-\beta_{g_i^0}^0)|x_{it})-\tau_1\right\}duds\big|\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\right]\right)\\
=&O_{\mathsf{P}}\left(\frac{1}{T}\sum_{t=1}^{T}\mathsf{E}\left[\left(\widehat{\alpha}_{h_i^0}+\widehat{\lambda}_t-(\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t)+x_{it}'(\widehat{\beta}_{g_i^0}-\widetilde{\beta}_{g_i^0})\right)^4\big|\widehat{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}}\right]\right)\\
=&O_{\mathsf{P}}\left(\frac{1}{T}\sum_{t=1}^{T}\left|\widehat{\alpha}_{h_i^0}+\widehat{\lambda}_t-(\widetilde{\alpha}_{h_i^0}+\widetilde{\lambda}_t)\right|^4+\left\|\widehat{\beta}_{g_i^0}-\widetilde{\beta}_{g_i^0}\right\|^4\right),
\end{aligned} \tag{A.39}$$

as N and $T\to\infty$. Thus, combining (A.37), (A.38) and (A.39) yields

$$\frac{1}{N}\sum_{i=1}^{N}\mathsf{Var}\left(A_i|\widehat{\boldsymbol{\theta}}(\boldsymbol{\tau}),\widetilde{\boldsymbol{\theta}}(\boldsymbol{\tau})\right) = O_{\mathsf{P}}\left(\sum_{g\in\{1,\ldots,G\}}\left\|\widehat{\beta}_g-\widetilde{\beta}_g\right\|^4+\frac{1}{T}\sum_{t=1}^{T}\sum_{h\in\{1,\ldots,H\}}\left|\widehat{\alpha}_h+\widehat{\lambda}_t-(\widetilde{\alpha}_h+\widetilde{\lambda}_t)\right|^4\right), \tag{A.40}$$

provided that $G < \infty$ and $H < \infty$. Finally, it follows from (A.31), (A.33), (A.36) and (A.40) that

$$
\begin{aligned}
&O_{\mathsf{P}}\left(\frac{\exp(-T\zeta/2)}{N^{1/2}T^{1/4}} + \frac{\exp(-T\zeta)}{T^{1/2}}\right) = \frac{1}{N}\sum_{i=1}^{N} A_i \\
&= O_{\mathsf{P}}\Bigg(\sum_{g\in\{1,\ldots,G\}} \left\|\widehat{\beta}_g - \widetilde{\beta}_g\right\|^2\Bigg) + O_{\mathsf{P}}\Bigg(\frac{1}{T}\sum_{t=1}^{T}\sum_{h\in\{1,\ldots,H\}} \left|\widehat{\alpha}_h + \widehat{\lambda}_t - (\widetilde{\alpha}_h + \widetilde{\lambda}_t)\right|^2\Bigg) \\
&\quad + O_{\mathsf{P}}\left(\frac{1}{N}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\sum_{h\in\{1,\ldots,H\}} \left|\widehat{\alpha}_{h_i^0} + \widehat{\lambda}_t - (\widetilde{\alpha}_{h_i^0} + \widetilde{\lambda}_t)\right|^4}\right),
\end{aligned}
$$

from which we have

$$
\sum_{g\in\{1,\ldots,G\}} \left\|\widehat{\beta}_g - \widetilde{\beta}_g\right\|^2 = O_{\mathsf{P}}\left(\frac{\exp(-T\zeta/2)}{N^{1/2}T^{1/4}} + \frac{\exp(-T\zeta)}{T^{1/2}}\right).
$$

This completes the proof. ∎

Proof of Lemma 1. We first show the asymptotic representation of $\widetilde{\beta}_g$ for $g = 1, \ldots, G$. Let $\phi_{h,t} := \alpha_h + \lambda_t$ and $\phi^0_{h,t} := \alpha^0_h + \lambda^0_t$ for $h \in \{1, \ldots, H\}$. Define

$$
\begin{aligned}
\mathbb{H}^{\alpha,\lambda}_{h,t}(\phi_{h,t}, \beta_g) &:= \frac{1}{N}\sum_{i:h_i^0=h, g_i^0=g} \{\tau_k - I(y_{it} \le \phi_{h,t} + x'_{it}\beta_g)\}, \\
\mathsf{H}^{\alpha,\lambda}_{h,t}(\phi_{h,t}, \beta_g) &:= \mathsf{E}\big[\mathbb{H}^{\alpha,\lambda}_{h,t}(\phi_{h,t}, \beta_g)\big] \\
&= \frac{1}{N}\sum_{i:h_i^0=h, g_i^0=g} \mathsf{E}\big[\tau_k - F_i(\phi_{h,t} - \phi^0_{h,t} + x'_{i1}(\beta_g - \beta^0_g)|x_{i1})\big], \qquad \text{(A.41)} \\
\mathbb{H}^{\beta}_{g}(\Phi_t, \beta_g) &:= \frac{1}{N}\sum_{h\in\{1,\ldots,H\}}\sum_{i:h_i^0=h, g_i^0=g} \{\tau_k - I(y_{it} \le \phi_{h,t} + x'_{it}\beta_g)\}x_{it}, \\
\mathsf{H}^{\beta}_{g}(\Phi_t, \beta_g) &:= \mathsf{E}\big[\mathbb{H}^{\beta}_{g}(\Phi_t, \beta_g)\big] \\
&= \frac{1}{N}\sum_{h\in\{1,\ldots,H\}}\sum_{i:h_i^0=h, g_i^0=g} \mathsf{E}\big[\{\tau_k - F_i(\phi_{h,t} - \phi^0_{h,t} + x'_{i1}(\beta_g - \beta^0_g)|x_{i1})\}x_{i1}\big], \qquad \text{(A.42)}
\end{aligned}
$$

with $\Phi_t := (\phi_{1,t}, \ldots, \phi_{H,t})'$ for $t = 1, \ldots, T$. Denote by $\widetilde{\phi}_{h,t} := \widetilde{\alpha}_h + \widetilde{\lambda}_t$ the infeasible estimator of $\phi^0_{h,t}$. By the subgradient condition in Koenker (2005), $\sup_{h,t} |\mathbb{H}^{\alpha,\lambda}_{h,t}(\widetilde{\phi}_{h,t}, \widetilde{\beta}_g)| = O_{\mathsf{P}}(N^{-1})$. Hence, expanding

$\mathsf{H}_{h,t}^{\alpha,\lambda}(\widetilde{\phi}_{h,t},\widetilde{\beta}_g)$ around $(\phi_{h,t}^0,\beta_g^0)$ yields that

$$\begin{aligned}\widetilde{\phi}_{h,t}-\phi_{h,t}^0=&\left(\frac{1}{N}\sum_{i:h_i^0=h,g_i^0=g}f_i(0)\right)^{-1}\mathbb{H}_{h,t}^{\alpha,\lambda}(\phi_{h,t}^0,\beta_g^0)-\iota_{h,g}'(\widetilde{\beta}_g-\beta_g^0)\\&+\left(\frac{1}{N}\sum_{i:h_i^0=h,g_i^0=g}f_i(0)\right)^{-1}\left\{\mathbb{H}_{h,t}^{\alpha,\lambda}(\widetilde{\phi}_{h,t},\widetilde{\beta}_g)-\mathsf{H}_{h,t}^{\alpha,\lambda}(\widetilde{\phi}_{h,t},\widetilde{\beta}_g)-\mathbb{H}_{h,t}^{\alpha,\lambda}(\phi_{h,t}^0,\beta_g^0)\right\}\\&+O_{\mathsf{P}}\left(\frac{1}{N}\vee\sup_{h,t}|\widetilde{\phi}_{h,t}-\phi_{h,t}^0|^2\vee\|\widetilde{\beta}_g-\beta_g^0\|^2\right).\end{aligned}\tag{A.43}$$

Moreover, let Φ_t^0 denote the true value of Φ_t, and expanding $\mathsf{H}_g^{\beta}(\widetilde{\Phi}_t,\widetilde{\beta}_g)$ around (Φ_t^0,β_g^0) yields that

$$\begin{aligned}\mathsf{H}_g^{\beta}(\widetilde{\Phi}_t,\widetilde{\beta}_g)=&-\sum_{h\in\{1,\ldots,H\}}\left(\frac{1}{N}\sum_{i:h_i^0=h,g_i^0=g}\mathsf{E}[f_i(0|x_{i1})x_{i1}]\right)\left(\widetilde{\phi}_{h,t}-\phi_{h,t}^0\right)\\&-\left(\frac{1}{N}\sum_{i:g_i^0=g}\mathsf{E}[f_i(0|x_{i1})x_{i1}x_{i1}']\right)\left(\widetilde{\beta}_g-\beta_g^0\right)\\&+o_{\mathsf{P}}\left(\|\widetilde{\beta}_g-\beta_g^0\|\right)+O_{\mathsf{P}}\left(\sup_{h,t}|\widetilde{\phi}_{h,t}-\phi_{h,t}^0|^2\right).\end{aligned}\tag{A.44}$$

Plugging (A.43) into (A.44) leads to

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^T\mathsf{H}_g^{\beta}(\widetilde{\Phi}_t,\widetilde{\beta}_g)=&-\frac{1}{T}\sum_{t=1}^T\sum_{h\in\{1,\ldots,H\}}\mathbb{H}_{h,t}^{\alpha,\lambda}(\phi_{h,t}^0,\beta_g^0)\iota_{h,g}-\Gamma_{Ng}(\widetilde{\beta}_g-\beta_g^0)\\&-\frac{1}{T}\sum_{t=1}^T\sum_{h\in\{1,\ldots,H\}}\left\{\mathbb{H}_{h,t}^{\alpha,\lambda}(\widetilde{\phi}_{h,t},\widetilde{\beta}_g)-\mathsf{H}_{h,t}^{\alpha,\lambda}(\widetilde{\phi}_{h,t},\widetilde{\beta}_g)-\mathbb{H}_{h,t}^{\alpha,\lambda}(\phi_{h,t}^0,\beta_g^0)\right\}\iota_{h,g}\\&+o_{\mathsf{P}}\left(\|\widetilde{\beta}_g-\beta_g^0\|\right)+O_{\mathsf{P}}\left(\frac{1}{N}\vee\sup_{h,t}|\widetilde{\phi}_{h,t}-\phi_{h,t}^0|^2\right).\end{aligned}\tag{A.45}$$

The subgradient condition in Koenker (2005) also implies that $\sup_{t\geq1}\mathbb{H}_g^{\beta}(\widetilde{\Phi}_t,\widetilde{\beta}_g)=O_{\mathsf{P}}(N^{-1}\sup_{i,t}\|\boldsymbol{x}_{it}\|)=O_{\mathsf{P}}(N^{-1})$, from which we have

$$\begin{aligned}O_{\mathsf{P}}(N^{-1})=&\frac{1}{T}\sum_{t=1}^T\left\{\mathbb{H}_g^{\beta}(\Phi_t^0,\beta_g^0)+\mathsf{H}_g^{\beta}(\widetilde{\Phi}_t,\widetilde{\beta}_g)\right\}\\&+\frac{1}{T}\sum_{t=1}^T\left\{\mathbb{H}_g^{\beta}(\widetilde{\Phi}_t,\widetilde{\beta}_g)-\mathsf{H}_g^{\beta}(\widetilde{\Phi}_t,\widetilde{\beta}_g)-\mathbb{H}_g^{\beta}((\Phi_t^0,\beta_g^0)\right\}.\end{aligned}\tag{A.46}$$

Denote $\mathbb{H}_t(\Phi_t,\beta_g):=-\sum_{h\in\{1,\ldots,H\}}\iota_{h,g}\mathbb{H}_{h,t}^{\alpha,\lambda}(\phi_{h,t},\beta_g)+\mathbb{H}_g^{\beta}(\Phi_t,\beta_g)$. It then follows from (A.45) and

(A.46) that

$$
\begin{aligned}
&\widetilde{\beta}_g - \beta_g^0 + o_{\mathsf{P}}(\|\widetilde{\beta}_g - \beta_g^0\|) \\
=&\Gamma_{Ng}^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbb{H}_t(\Phi_t^0,\beta_g^0) + \Gamma_{Ng}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left\{\mathbb{H}_t(\widetilde{\Phi}_t,\widetilde{\beta}_g) - \mathbb{H}_t(\Phi_t^0,\beta_g^0) - \mathsf{H}_t(\widetilde{\Phi}_t,\widetilde{\beta}_g)\right\} \\
&+ O_{\mathsf{P}}\left(\frac{1}{N} \vee \sup_{h,t}|\widetilde{\phi}_{h,t} - \phi_{h,t}^0|^2\right). \qquad \text{(A.47)}
\end{aligned}
$$

Using the similar arguments as in the proof of Theorem 3.2 of Kato et al. (2012) (see Step 3 therein), we can show that if N grows at most polynomially in T, then

$$
\sup_{h,t}|\widetilde{\phi}_{h,t} - \phi_{h,t}^0| = O_{\mathsf{P}}(N^{-1/2}(\log N)^{1/2}).
$$

Combining with (A.47), we can obtain the following representation:

$$
\begin{aligned}
&\widetilde{\beta}_g - \beta_g^0 + o_{\mathsf{P}}(\|\widetilde{\beta}_g - \beta_g^0\|) \\
=&\Gamma_{Ng}^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbb{H}_t(\Phi_t^0,\beta_g^0) + \Gamma_{Ng}^{-1}\frac{1}{T}\sum_{t=1}^{T}\left\{\mathbb{H}_t(\widetilde{\Phi}_t,\widetilde{\beta}_g) - \mathbb{H}_t(\Phi_t^0,\beta_g^0) - \mathsf{H}_t(\widetilde{\Phi}_t,\widetilde{\beta}_g)\right\} + O_{\mathsf{P}}\left(N^{-1}\log N\right). \qquad \text{(A.48)}
\end{aligned}
$$

Next, we show the rate of convergence for the second term on the right side of (A.48). Define $\phi_{h,t}^* := \arg\min_{\phi\in\mathbb{R}} \sum_{i:h_i^0=h,g_i^0=g}\rho_{\tau_k}(y_{it} - \phi - x_{it}'\beta_g^0)$ and $\Phi_t^* := (\phi_{1,t}^*,\ldots,\phi_{H,t}^*)'$, and then we can rewrite the second term as

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left\{\mathbb{H}_t(\widetilde{\Phi}_t,\widetilde{\beta}_g) - \mathbb{H}_t(\Phi_t^0,\beta_g^0) - \mathsf{H}_t(\widetilde{\Phi}_t,\widetilde{\beta}_g)\right\} \\
=&\frac{1}{T}\sum_{t=1}^{T}\mathbb{H}_t(\widetilde{\Phi}_t,\widetilde{\beta}_g) - \mathbb{H}_t(\Phi_t^*,\beta_g^0) - \left\{\mathsf{H}_t(\widetilde{\Phi}_t,\widetilde{\beta}_g) - \mathsf{H}_t(\Phi_t^*,\beta_g^0)\right\} \\
&+\frac{1}{T}\sum_{t=1}^{T}\left\{\mathbb{H}_t(\Phi_t^*,\beta_g^0) - \mathbb{H}_t(\Phi_t^0,\beta_g^0) - \mathsf{H}_t(\Phi_t^*,\beta_g^0)\right\}. \qquad \text{(A.49)}
\end{aligned}
$$

Provided that $\{\mathbb{H}_t(\Phi_t^*,\beta_g^0) - \mathbb{H}_t(\Phi_t^0,\beta_g^0), t = 1,\ldots,T\}$ are independent across t, the second average on the right side of (A.49) possesses the similar structure as $\frac{1}{n}\sum_{i=1}^{n}\mathbb{H}_{ni}^{(3)}(\widetilde{\alpha}_i,\boldsymbol{\beta}_0) - \mathbb{H}_{ni}^{(3)}(\alpha_{i0},\boldsymbol{\beta}_0) - \{\mathsf{H}_{ni}^{(3)}(\widetilde{\alpha}_i,\boldsymbol{\beta}_0) - \mathsf{H}_{ni}^{(3)}(\alpha_{i0},\boldsymbol{\beta}_0)\}$ in Galvao et al. (2020). Thus, following the similar arguments as in the proof of Theorem 1 of Galvao et al. (2020), we have

$$
\frac{1}{T}\sum_{t=1}^{T}\left\{\mathbb{H}_t(\Phi_t^*,\beta_g^0) - \mathbb{H}_t(\Phi_t^0,\beta_g^0) - \mathsf{H}_t(\Phi_t^*,\beta_g^0)\right\} = O_{\mathsf{P}}\left(N^{-2/3}T^{-1/2} + N^{-1}\log N\right).
$$

Combining with (A.48) and (A.49) yields that

$$\begin{aligned}\widetilde{\beta}_g - \beta_g^0 = &\Gamma_{Ng}^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbb{H}_t(\Phi_t^0, \beta_g^0)\\ &+ \Gamma_{Ng}^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbb{H}_t(\widetilde{\Phi}_t, \widetilde{\beta}_g) - \mathbb{H}_t(\Phi_t^*, \beta_g^0) - \left\{\mathsf{H}_t(\widetilde{\Phi}_t, \widetilde{\beta}_g) - \mathsf{H}_t(\Phi_t^*, \beta_g^0)\right\}\\ &+ O_{\mathsf{P}}\left(N^{-1}\log N\right) + o_{\mathsf{P}}\left(\|\widetilde{\beta}_g - \beta_g^0\|\right) + o_{\mathsf{P}}\left((NT)^{-1/2}\right). \end{aligned} \tag{A.50}$$

Moreover, using the similar arguments as in Lemma 7 and Lemma 5 of Galvao et al. (2020), we can show that

$$\sup_{h,t}|\phi_{h,t}^* - \widetilde{\phi}_{h,t}| = O_{\mathsf{P}}\left(\|\widetilde{\beta}_g - \beta_g^0\| + N^{-1}\log N\right), \tag{A.51}$$

and

$$\begin{aligned}&\sup_t\left|\mathbb{H}_t(\widetilde{\Phi}_t, \widetilde{\beta}_g) - \mathbb{H}_t(\Phi_t^*, \beta_g^0) - \left\{\mathsf{H}_t(\widetilde{\Phi}_t, \widetilde{\beta}_g) - \mathsf{H}_t(\Phi_t^*, \beta_g^0)\right\}\right|\\ &= O_{\mathsf{P}}\left(\left\{\|\widetilde{\beta}_g - \beta_g^0\| + \sup_{h,t}|\phi_{h,t}^* - \widetilde{\phi}_{h,t}|\right\}^{1/2} N^{-1/2}(\log N)^{1/2} + N^{-1}\log N\right), \end{aligned} \tag{A.52}$$

respectively. Plugging (A.51) into (A.52) further gives

$$\begin{aligned}&\sup_t\left|\mathbb{H}_t(\widetilde{\Phi}_t, \widetilde{\beta}_g) - \mathbb{H}_t(\Phi_t^*, \beta_g^0) - \left\{\mathsf{H}_t(\widetilde{\Phi}_t, \widetilde{\beta}_g) - \mathsf{H}_t(\Phi_t^*, \beta_g^0)\right\}\right|\\ &= O_{\mathsf{P}}\left(\|\widetilde{\beta}_g - \beta_g^0\|^{1/2}N^{-1/2}(\log N)^{1/2} + N^{-1}\log N\right). \end{aligned} \tag{A.53}$$

It hence follows from (A.50) and (A.53) that

$$\begin{aligned}\widetilde{\beta}_g - \beta_g^0 =&\Gamma_{Ng}^{-1}\frac{1}{T}\sum_{t=1}^{T}\mathbb{H}_t(\Phi_t^0, \beta_g^0) + O_{\mathsf{P}}\left(N^{-1}\log N\right) + \|\widetilde{\beta}_g - \beta_g^0\|^{1/2}O_{\mathsf{P}}\left(N^{-1/2}(\log N)^{1/2}\right)\\ &+ o_{\mathsf{P}}\left(\|\widetilde{\beta}_g - \beta_g\| + (NT)^{-1/2}\right)\\ =&O_{\mathsf{P}}\big((NT)^{-1/2} + N^{-1}\log N\big) + \|\widetilde{\beta}_g - \beta_g^0\|^{1/2}O_{\mathsf{P}}\big(N^{-1/2}(\log N)^{1/2}\big) + o_{\mathsf{P}}\big(\|\widetilde{\beta}_g - \beta_g\|\big). \end{aligned} \tag{A.54}$$

By **fact 1** in Galvao et al. (2020), (A.54) implies

$$\|\widetilde{\beta}_g - \beta_g^0\| = O_{\mathsf{P}}\left((NT)^{-1/2} + N^{-1}\log N\right).$$

Plugging this into (A.54) yields

$$\begin{aligned}\widetilde{\beta}_g - \beta_g^0 =& \Gamma_{Ng}^{-1} \frac{1}{T} \sum_{t=1}^{T} \mathbb{H}_t(\Phi_t^0, \beta_g^0) + O_{\mathsf{P}}\left(\left\{(NT)^{-1/4} + N^{-1/2}(\log N)^{1/2}\right\} N^{-1/2}(\log N)^{1/2}\right) \\ =& \Gamma_{Ng}^{-1}\left[\frac{1}{NT}\sum_{t=1}^{T}\sum_{h\in\{1,\ldots,H\}}\sum_{i:h_i^0=h,g_i^0=g}\{\tau_k - I(\varepsilon_{it}\le 0)\}\left(x_{it}-\iota_{h,g}\right)\right] + o_{\mathsf{P}}\left((NT)^{-1/2}\right) \quad \text{(A.55)}\end{aligned}$$

provided that $T(\log N)^2/N \to 0$ as N and $T \to \infty$, where $\varepsilon_{it} = y_{it} - \phi_{h_i^0,t}^0 - x_{it}'\beta_{g_i^0}^0$.

Finally, we derive the asymptotic distribution based on (A.55). Put $z_{it} := \{\tau_k - I(\varepsilon_{it} \le 0)\}(x_{it} - \iota_{h,g})$. We have $\mathsf{E}(z_{it}) = \mathsf{E}(z_{i1}) = 0$ and

$$\begin{aligned}\mathsf{Var}\left(z_{it}\right) = \mathsf{Var}\left(z_{i1}\right) &= \mathsf{E}\big[\mathsf{Var}\left(z_{i1}|x_{i1}\right)\big] + \mathsf{Var}\left[\mathsf{E}(z_{i1}|x_{i1})\right] \\ &= \mathsf{E}\big[(x_{i1}-\iota_{h,g})\mathsf{Var}\left(\tau_k - I(\varepsilon_{it}\le 0)|x_{i1}\right)(x_{i1}-\iota_{h,g})'\big] \\ &= \tau_k(1-\tau_k)\mathsf{E}\big[(x_{i1}-\iota_{h,g})(x_{i1}-\iota_{h,g})'\big].\end{aligned}$$

Denote $S_{NT,g} := \sum_{t=1}^{T}\sum_{h\in\{1,\ldots,H\}}\sum_{i:h_i^0=h,g_i^0=g} z_{it}$ and $B_{NT,g}^2 := \mathsf{Var}\left(S_{NT,g}\right)$. Provided that z_{it} is bounded and independent over i and t, by Lyapunov CLT we obtain that

$$\frac{S_{NT,g} - \mathsf{E}(S_{NT,g})}{B_{NT,g}} = \sqrt{NT}\frac{(NT)^{-1}S_{NT,g}}{\sqrt{\tau_k(1-\tau_k)V_{Ng}}} \xrightarrow{D} N(0,1),$$

where $V_{Ng} = N^{-1}\sum_{h\in\{1,\ldots,H\}}\sum_{i:h_i^0=h,g_i^0=g}\mathsf{E}[(x_{i1}-\iota_{h,g})(x_{i1}-\iota_{h,g})']$. This implies

$$\sqrt{NT}(\widetilde{\beta}_g(\tau_k) - \beta_g^0(\tau_k)) \xrightarrow{D} \mathbb{N}(\mathbf{0}, \tau_k(1-\tau_k)\Gamma_g^{-1}V_g\Gamma_g^{-1}),$$

as N and $T \to \infty$. The proof is complete. ∎

Proof of Theorem 3. When N grows at most polynomially in T, it follows from Corollary 3 that

$$\widehat{\beta}_g(\tau_k) - \widetilde{\beta}_g(\tau_k) = O_{\mathsf{P}}\left(\frac{\exp(-T\zeta/4)}{N^{1/4}T^{1/8}}\right),$$

for $g \in \{1,\ldots,G\}$, where $\zeta > 0$ defined in Theorem 2. Thus, from the asymptotic distribution of $\widetilde{\beta}_g(\tau_k)$ in Lemma 1, we have

$$\begin{aligned}\sqrt{NT}\big(\widehat{\beta}_g(\tau_k) - \beta_g^0(\tau_k)\big) &= \sqrt{NT}\big(\widehat{\beta}_g(\tau_k) - \widetilde{\beta}_g(\tau_k)\big) + \sqrt{NT}\big(\widetilde{\beta}_g(\tau_k) - \beta_g^0(\tau_k)\big) \\ &= o_{\mathsf{P}}(1) + \sqrt{NT}\big(\widetilde{\beta}_g(\tau_k) - \beta_g^0(\tau_k)\big) \\ &\xrightarrow{D} \mathbb{N}(\mathbf{0}, \tau_k(1-\tau_k)\Gamma_g^{-1}V_g\Gamma_g^{-1}).\end{aligned}$$

The proof is complete. ∎

Proof of Theorem 4. Let $\mathcal{S} = \{(G,H) : G = 1,\ldots,G_{\max}, H = 1,\ldots,H_{\max}\}$. We divide $\mathcal{S}$ into three subsets: $\mathcal{S}_1 = \{(G,H) : G < G_0 \text{ or } H < H_0\}$, $\mathcal{S}_2 = \{(G,H) : \{G \geq G_0, H > H_0\} \text{ or } \{H \geq H_0, G > G_0\}\}$ and $\mathcal{S}_3 = \{(G,H) : G = G_0, H = H_0\}$, which correspond to the cases of underspecified, overspecified and true number of groups, respectively. In the following, we show

$$\lim_{N,T\to\infty} \mathsf{P}\{\mathrm{IC}(G,H) > \mathrm{IC}(G_0,H_0)\} = 1, \tag{A.56}$$

for $(G,H) \in \mathcal{S}_1$ and $\mathcal{S}_2$, respectively, and hence $\mathsf{P}\big\{(\widehat{G},\widehat{H}) = (G_0,H_0)\big\} \to 1$.

For $G < G_0$, it follows from the uniform consistency result in Theorem 1 that

$$\begin{aligned}\mathrm{IC}(G_0,H_0) &= \frac{1}{NT}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{t=1}^{T}\rho_{\tau_k}\left(y_{it} - \widehat{\alpha}_{\widehat{h}_i}^{(G_0,H_0)}(\tau_k) - \widehat{\lambda}_t^{(G_0,H_0)}(\tau_k) - x_{it}'\widehat{\beta}_{\widehat{g}_i}^{(G_0,H_0)}(\tau_k)\right) + \kappa K np\,(G_0,H_0)\\ &= \sigma_0 + o_{\mathsf{P}}(1),\end{aligned}$$

provided that $\kappa \to 0$ as N and $T \to \infty$ and $K < \infty$. Therefore, by the first condition of Assumption 5(i), we have

$$\begin{aligned}\mathrm{IC}(G,H) &\geq \min_{1\leq G<G_0}\ \inf_{\gamma_g\in\Gamma_G,\gamma_h\in\Gamma_H} \widehat{\sigma}(\gamma_g,\gamma_h) + \kappa K np\,(G,H)\\ &> \sigma_0 + o_{\mathsf{P}}(1)\\ &> \mathrm{IC}(G_0,H_0) + o_{\mathsf{P}}(1),\end{aligned}$$

that is (A.56) for any $G < G_0$ as N and $T \to \infty$. Using the similar argument as above and the second condition of Assumption 5(i), we can have (A.56) for any $H < H_0$.

Next, we consider the case when $(G,H) \in \mathcal{S}_2$, where some groups are segmented into subgroups due to overspecifying the number of groups. Let (γ_g^0,γ_h^0) and $(\widehat{\gamma}_g,\widehat{\gamma}_h)$ denote the true and estimated group partitions, respectively. We have that

$$\begin{aligned}&\mathrm{IC}(G,H) - \mathrm{IC}(G_0,H_0)\\ =&\{\widehat{\sigma}(\widehat{\gamma}_g,\widehat{\gamma}_h)|_{(G,H)} - \widehat{\sigma}(\widehat{\gamma}_g,\widehat{\gamma}_h)|_{(G_0,H_0)}\} + \{np\,(G,H) - np\,(G_0,H_0)\}\kappa K\\ =&\{\widehat{\sigma}(\widehat{\gamma}_g,\widehat{\gamma}_h)|_{(G,H)} - \widehat{\sigma}(\gamma_g^0,\gamma_h^0)|_{(G_0,H_0)}\} + \{\widehat{\sigma}(\gamma_g^0,\gamma_h^0)|_{(G_0,H_0)} - \widehat{\sigma}(\widehat{\gamma}_g,\widehat{\gamma}_h)|_{(G_0,H_0)}\}\\ &+ \{np\,(G,H) - np\,(G_0,H_0)\}\kappa K,\end{aligned} \tag{A.57}$$

where $\widehat{\sigma}(\cdot,\cdot)|_{(G,H)}$ denotes the corresponding objective function obtained under certain specified number of groups (G,H). Following the proof of Theorem 2 we can show that for all $\epsilon > 0$,

$$\left|\widehat{\sigma}(\widehat{\gamma}_g,\widehat{\gamma}_h)|_{(G,H)} - \widehat{\sigma}(\gamma_g^0,\gamma_h^0)|_{(G_0,H_0)}\right| = o_{\mathsf{P}}(N^{-1/2}\exp(-T\epsilon)),$$

and

$$\widehat{\sigma}(\gamma_g^0, \gamma_h^0)|_{(G_0,H_0)} - \widehat{\sigma}(\widehat{\gamma}_g, \widehat{\gamma}_h)|_{(G_0,H_0)} = o_{\mathsf{P}}(N^{-1/2}\exp(-T\epsilon)),$$

provided that $N/(T^{1/2}\exp(T\epsilon)) \to 0$ as N and $T \to \infty$. Hence, from (A.57) and Assumption 5(*ii*) we obtain that

$$\begin{aligned}
&\mathsf{P}\big\{\mathrm{IC}(G,H) - \mathrm{IC}(G_0,H_0) > 0\big\} = \mathsf{P}\big\{\sqrt{N}\exp(T\epsilon)\big(\mathrm{IC}(G,H) - \mathrm{IC}(G_0,H_0)\big) > 0\big\} \\
=&\mathsf{P}\big\{\big(\sqrt{N}\exp(T\epsilon)\{\widehat{\sigma}(\widehat{\gamma}_g,\widehat{\gamma}_h)|_{(G,H)} - \widehat{\sigma}(\gamma_g^0,\gamma_h^0)|_{(G_0,H_0)}\} + \sqrt{N}\exp(T\epsilon)\{\widehat{\sigma}(\gamma_g^0,\gamma_h^0)|_{(G_0,H_0)} - \widehat{\sigma}(\widehat{\gamma}_g,\widehat{\gamma}_h)|_{(G_0,H_0)}\} \\
&\quad + \{np\,(G,H) - np\,(G_0,H_0)\}\sqrt{N}\exp(T\epsilon)\kappa K\big) > 0\big\} \to 1,
\end{aligned}$$

noting that $np\,(G,H) > np\,(G_0,H_0)$ when $(G,H) \in \mathcal{S}_2$. This completes the proof since $(\widehat{G},\widehat{H}) = \arg\min_{(G,H)\in\mathcal{S}} \mathrm{IC}(G,H)$. ∎

Derivations in Illustrative Example for Misclustering Probability.

Conditional on $\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau})$, we have

$$\sqrt{T}\,\frac{T^{-1}\sum_{t=1}^{T} W_{it,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau})) - \mathsf{E}[W_{i1,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}))]}{\sqrt{\mathsf{Var}\,[W_{i1,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}))]}} \xrightarrow{D} N(0,1),$$

as $T \to \infty$, for each fixed i. Let Φ denote the standard normal distribution function. The weak convergence above implies that

$$\mathsf{P}\left(T^{-1}\sum_{t=1}^{T} W_{it,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau})) \le 0\right) \sim \Phi\left(-\sqrt{T}\frac{\mathsf{E}[W_{i1,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}))]}{\sqrt{\mathsf{Var}\,[W_{i1,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}))]}}\right)$$

where for any two positive sequences a_T and b_T, $a_T \sim b_T$ means that $\lim_{T\to\infty} a_T/b_T = 1$. Moreover, by the Mills ratio,

$$\Phi\left(-\sqrt{T}\frac{\mathsf{E}[W_{i1,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}))]}{\sqrt{\mathsf{Var}\,[W_{i1,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}))]}}\right) = O\left(T^{-1/2}\exp\left(-\frac{\mathsf{E}^2[W_{i1,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}))]}{2\mathsf{Var}\,[W_{i1,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}))]}\right)\right),$$

as $T \to \infty$. Thus we obtain

$$\begin{aligned}
\mathsf{P}\big(\widehat{g}_i \ne g_i^0\big) &= \sum_{g\ne g_i^0} \mathsf{P}\big(\widehat{g}_i = g\big) \le \sum_{g\ne g_i^0} \mathsf{P}\left(T^{-1}\sum_{t=1}^{T} W_{it,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau})) \le 0\right) \\
&= O\left(T^{-1/2}\sum_{g\ne g_i^0}\exp\left(-T\frac{\mathsf{E}^2[W_{i1,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}))]}{2\mathsf{Var}\,[W_{i1,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}))]}\right)\right).
\end{aligned}$$

Let q_τ denote the $100\tau\%$ quantile of $N(0, \sigma_\varepsilon^2)$. Applying Knight (1998)'s identity and (3.3) of Corollary 1 yields that, as N and $T \to \infty$,

$$\begin{aligned}
\mathsf{E}[W_{i1,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}))] &= \sum_{k=1}^{K} \int_0^{\widehat{\beta}_g(\tau_k) - \widehat{\beta}_{g_i^0}(\tau_k)} \left\{ \Phi\left(\frac{u + q_{\tau_k}}{\sigma_\varepsilon} \right) - \tau_k \right\} du \\
&= \sum_{k=1}^{K} \int_0^{\beta_g^0 - \beta_{g_i^0}^0} \left\{ \Phi\left(\frac{u + q_{\tau_k}}{\sigma_\varepsilon} \right) - \tau_k \right\} du + o_{\mathsf{P}}(1),
\end{aligned}$$

and

$$\begin{aligned}
&\mathsf{Var}\left[W_{i1,g}(\widehat{\boldsymbol{\beta}}(\boldsymbol{\tau}))\right] \\
&= \sum_{k=1}^{K} \sum_{l=1}^{K} \int_0^{\widehat{\beta}_g(\tau_k) - \widehat{\beta}_{g_i^0}(\tau_k)} \int_0^{\widehat{\beta}_g(\tau_l) - \widehat{\beta}_{g_i^0}(\tau_l)} \left\{ \Phi\left(\frac{(u + q_{\tau_k}) \wedge (s + q_{\tau_l})}{\sigma_\varepsilon} \right) - \Phi\left(\frac{u + q_{\tau_k}}{\sigma_\varepsilon} \right) \Phi\left(\frac{s + q_{\tau_l}}{\sigma_\varepsilon} \right) \right\} du ds \\
&= \sum_{k=1}^{K} \sum_{l=1}^{K} \int_0^{\beta_g^0 - \beta_{g_i^0}^0} \int_0^{\beta_g^0 - \beta_{g_i^0}^0} \left\{ \Phi\left(\frac{(u + q_{\tau_k}) \wedge (s + q_{\tau_l})}{\sigma_\varepsilon} \right) - \Phi\left(\frac{u + q_{\tau_k}}{\sigma_\varepsilon} \right) \Phi\left(\frac{s + q_{\tau_l}}{\sigma_\varepsilon} \right) \right\} du ds + o_{\mathsf{P}}(1).
\end{aligned}$$

∎