Online Supplement to
"Panel Threshold Regressions With Latent Group Structures"

Ke Miao[a], Liangjun Su[a] and Wendun Wang[b]

[a] School of Economics, Singapore Management University, Singapore
[b] Econometric Institute, Erasmus University Rotterdam and Tinbergen Institute, The Netherlands

This supplement is composed of four parts. Section B contains the proofs of Lemmas A.-A.8 in the above paper. Section C contains the full analysis of the infeasible estimators. Section D provides some additional assumptions for the determination of the true number of groups and a new proposition. Section E studies the consistency of the panel threshold estimators in the framework of fixed threshold effects.

B Proof of Lemmas A.1-A.8 in Appendix A

Proof of Lemma A.1. Note that

$$\begin{aligned}
&\frac{1}{NT}\left[\mathcal{Q}(\Theta,\mathbf{D},\mathbf{G})-\tilde{\mathcal{Q}}(\Theta,\mathbf{D},\mathbf{G})\right]\\
=&\ \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(\beta_{g_i^0}^0-\beta_{g_i})'\tilde{x}_{it}+\delta_{g_i^0}^{0\prime}\tilde{x}_{it}(\gamma_{g_i^0}^0)-\delta_{g_i}'\tilde{x}_{it}(\gamma_{g_i})\right]\tilde{\varepsilon}_{it}\\
=&\ \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\theta_{g_i^0}^{0\prime}\tilde{z}_{it}(\gamma_{g_i^0}^0)\varepsilon_{it}-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\theta_{g_i}'\tilde{z}_{it}(\gamma_{g_i})\varepsilon_{it}\\
=&\ \sum_{g=1}^{G}\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(g_i^0=g)\theta_g^{0\prime}\tilde{z}_{it}(\gamma_g^0)\varepsilon_{it}-\sum_{g=1}^{G}\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(g_i=g)\theta_g'\tilde{z}_{it}(\gamma_g)\varepsilon_{it}.
\end{aligned}$$

It suffices to show that the second term in the last line is $o_p(1)$ uniformly in $(\Theta,\mathbf{D},\mathbf{G})\in\mathcal{B}^G\times\Gamma^G\times\mathcal{G}^N$. For each $g\in\mathcal{G}$, we have

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(g_i=g)\theta_g'\tilde{z}_{it}(\gamma_g)\varepsilon_{it}=&\ \frac{1}{NT}\sum_{i=1}^{N}\mathbf{1}(g_i=g)\sum_{t=1}^{T}\theta_g'z_{it}(\gamma_g)\varepsilon_{it}\\
&-\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{1}(g_i=g)\sum_{t,s=1}^{T}\theta_g'z_{is}(\gamma_g)\varepsilon_{it}\equiv A_1\left(\theta_g,\gamma_g\right)-A_2\left(\theta_g,\gamma_g\right),
\end{aligned}$$

where $\sum_{t,s=1}^{T}=\sum_{t=1}^{T}\sum_{s=1}^{T}$. Then by the compactness of $\mathcal{B}$ in Assumption A.1(iv) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\sup_{(\theta,\gamma)\in\mathcal{B}\times\Gamma}\left|\frac{1}{NT}\sum_{i=1}^{N}\mathbf{1}(g_i=g)\sum_{t=1}^{T}\theta'z_{it}(\gamma)\varepsilon_{it}\right|\leq&\ C\sup_{\gamma\in\Gamma}\left\|\frac{1}{NT}\sum_{i=1}^{N}\mathbf{1}(g_i=g)\sum_{t=1}^{T}z_{it}(\gamma)\varepsilon_{it}\right\|\\
\leq&\ C\left\{\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(g_i=g)\right\}^{1/2}\frac{1}{N}\left\{\sum_{i=1}^{N}\sup_{\gamma\in\Gamma}\left\|\frac{1}{T}\sum_{t=1}^{T}z_{it}(\gamma)\varepsilon_{it}\right\|^2\right\}^{1/2}.
\end{aligned}$$

Following similar arguments used in the proof of Lemma A.3 in Hansen (2000), we can show that

$\max_{1\leq i\leq N}\sup_{\gamma\in\Gamma}\left\|\frac{1}{T}\sum_{t=1}^{T}z_{it}(\gamma)\varepsilon_{it}\right\|^2=o_p(1)$. It follows that $\sup_{(\theta,\gamma)\in\mathcal{B}\times\Gamma}|A_1(\theta,\gamma)|=o_p(1)$. Similarly, by the repeated use of Cauchy-Schwarz inequality

$$\begin{aligned}\sup_{(\theta,\gamma)\in\mathcal{B}\times\Gamma}\left|\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{1}(g_i=g)\sum_{t,s=1}^{T}\theta'z_{is}(\gamma)\varepsilon_{it}\right| &\leq C\left\{\sup_{\gamma\in\Gamma}\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{s=1}^{T}z_{is}(\gamma)\right\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\right|^2\right\}^{1/2}\\ &\leq 2C\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{s=1}^{T}\|x_{is}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\right|^2\right\}^{1/2}\\ &= O_P(T^{-1/2}).\end{aligned}$$

where we use the fact that $E|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}|^2=O\left(T^{-1}\right)$ under Assumption A.1(i.1) or A.1(i.2) and Assumption A.1(iii). Then $\sup_{(\theta,\gamma)\in\mathcal{B}\times\Gamma}|A_2(\theta,\gamma)|=o_p(1)$. Consequently, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(g_i=g)\theta_g'\tilde{z}_{it}(\gamma_g)\varepsilon_{it}=o_p(1)$ uniformly in $(\Theta,\mathbf{D},\mathbf{G})\in\mathcal{B}^G\times\Gamma^G\times\mathcal{G}^N$. ■

Proof of Lemma A.2. It suffices to show (i) $\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|\right)=o_p(1)$ and (ii) $\max_{\tilde{g}\in\mathcal{G}}\left(\min_{g\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|\right)=o_p(1)$.

We first show (i). By Lemma A.1, we have

$$\begin{aligned}\frac{1}{NT}\tilde{\mathcal{Q}}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}}) &= \frac{1}{NT}\mathcal{Q}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}})+o_p(1)\leq\frac{1}{NT}\mathcal{Q}(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)+o_p(1)\\ &= \frac{1}{NT}\tilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)+o_p(1),\end{aligned}$$

where the inequality holds by the definition of least squares estimator. On the other hand, noting that $\tilde{\mathcal{Q}}(\Theta,\mathbf{D},\mathbf{G})$ is minimized at $(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)$, we have $\frac{1}{NT}[\tilde{\mathcal{Q}}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}})-\tilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)]\geq 0$. It follows that $\frac{1}{NT}[\tilde{\mathcal{Q}}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}})-\tilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)]=o_p(1)$. By direct calculation, we have uniformly in $(\Theta,\mathbf{D},\mathbf{G})$,

$$\begin{aligned}&\frac{1}{NT}\left[\tilde{\mathcal{Q}}(\Theta,\mathbf{D},\mathbf{G})-\tilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}_g^0)\right]\\ =\;&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\theta_{g_i^0}^{0\prime}\tilde{z}_{it}(\gamma_{g_i^0}^0)-\theta_{g_i}'\tilde{z}_{it}(\gamma_{g_i})\right\}^2\\ =\;&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{(\theta_{g_i^0}^0-\theta_{g_i})'\tilde{z}_{it}(\gamma_{g_i})+\delta_{g_i^0}^{0\prime}[\tilde{x}_{it}(\gamma_{g_i^0}^0)-\tilde{x}_{it}(\gamma_{g_i})]\right\}^2\\ =\;&\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(g_i^0=g)\mathbf{1}(g_i=\tilde{g})\left[(\theta_g^0-\theta_{\tilde{g}})'\tilde{z}_{it}(\gamma_{\tilde{g}})\right]^2+o_p(1)\end{aligned}$$

where the last equality follows from the fact that $\sup_i\left\|\delta_{g_i^0}^0\right\|=o(1)$ under Assumption A.1(vi),

$$\begin{aligned}\sup_{\gamma\in\Gamma}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\tilde{x}_{it}(\gamma)\|^2 &\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\tilde{x}_{it}\|^2=O_p(1),\text{ and}\\ \sup_{\gamma,\gamma^*\in\Gamma}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\tilde{z}_{it}(\gamma)\|\,\|\tilde{x}_{it}(\gamma^*)\| &\leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\|\tilde{x}_{it}\|^2=O_p(1).\end{aligned}$$

By the definition of $M_{NT}(g,\tilde{g},\mathbf{D},\mathbf{G})$ in section 3.1, we have

$$\begin{aligned}
o_p(1) &= \frac{1}{NT}\left[\widetilde{\mathcal{Q}}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}})-\widetilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}_g^0)\right] \\
&= \sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}(\theta_g^0-\hat{\theta}_{\tilde{g}})'M_{NT}(g,\tilde{g},\hat{\mathbf{D}},\hat{\mathbf{G}})(\theta_g^0-\hat{\theta}_{\tilde{g}})+o_p(1) \\
&\geq \max_{g\in\mathcal{G}}\sum_{\tilde{g}=1}^{G}(\theta_g^0-\hat{\theta}_{\tilde{g}})'M_{NT}(g,\tilde{g},\hat{\mathbf{D}},\hat{\mathbf{G}})(\theta_g^0-\hat{\theta}_{\tilde{g}})+o_p(1) \\
&\geq \max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|^2\right)\sum_{\tilde{g}=1}^{G}\lambda_{\min}[M_{NT}(g,\tilde{g},\hat{\mathbf{D}},\hat{\mathbf{G}})]+o_p(1) \\
&\geq \max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|^2\right)\underline{c}_\lambda+o_p(1).
\end{aligned}$$

where the last equality follows from Assumption A.2 which says that there exists a group $\tilde{g}^*\in\mathcal{G}$ such that $\lambda_{\min}[M_{NT}(g,\tilde{g}^*,\hat{\mathbf{D}},\hat{\mathbf{G}})]>\underline{c}_\lambda>0$ with probability approaching 1. Consequently, we have $\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|\right)$ $=o_p(1)$.

To show (ii), let $\sigma(g)\equiv\sigma_{\hat{\Theta}}(g)\equiv\text{argmin}_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|$. Then by the triangle inequality, we have for any $\tilde{g}\neq g$

$$\left\|\hat{\theta}_{\sigma(g)}-\hat{\theta}_{\sigma(\tilde{g})}\right\|\geq\|\theta_g^0-\theta_{\tilde{g}}^0\|-\left\|\hat{\theta}_{\sigma(\tilde{g})}-\theta_{\tilde{g}}^0\right\|-\left\|\hat{\theta}_{\sigma(g)}-\theta_g^0\right\|.$$

The first term on the right hand side (RHS) of the last inequality is larger than $c_{g,\tilde{g}}$ by Assumption A.3(a) and the second and third terms are $o_p(1)$ by the above arguments. Then we can conclude that $\sigma_{\hat{\Theta}}(g)\neq\sigma_{\hat{\Theta}}(\tilde{g})$ w.p.a.1, implying that $\sigma_{\hat{\Theta}}(\cdot)$ is bijective and has the inverse $\sigma_{\hat{\Theta}}^{-1}$. Thus, we have for all $\tilde{g}\in\mathcal{G}$,

$$\min_{g\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|\leq\left\|\theta^0_{\sigma^{-1}(\tilde{g})}-\hat{\theta}_{\tilde{g}}\right\|=\min_{h\in\mathcal{G}}\left\|\theta^0_{\sigma^{-1}(\tilde{g})}-\hat{\theta}_h\right\|=o_p(1).$$

Therefore we have $\max_{\tilde{g}\in\mathcal{G}}\left(\min_{g\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|\right)=o_p(1)$. This completes the proof of Lemma A.2. ■.

Proof of Lemma A.3. For all $g\in\mathcal{G}$, we have

$$\mathbf{1}(\hat{g}_i(\Theta,\mathbf{D})=g)\leq\mathbf{1}\left\{\sum_{t=1}^{T}[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_g)'\theta_g]^2\leq\sum_{t=1}^{T}[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_{g_i^0})'\theta_{g_i^0}]^2\right\}.$$

Thus we have

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(\hat{g}_i(\Theta,\mathbf{D})\neq g_i^0)=\sum_{g=1}^{G}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(\hat{g}_i(\Theta,\mathbf{D})=g)\mathbf{1}(g_i^0\neq g)\leq\sum_{g=1}^{G}\frac{1}{N}\sum_{i=1}^{N}\mathcal{Z}_{ig}(\Theta,\mathbf{D}),$$

where

$$\mathcal{Z}_{ig}(\Theta,\mathbf{D})\equiv\mathbf{1}(g_i^0\neq g)\mathbf{1}\left\{\sum_{t=1}^{T}[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_g)'\theta_g)]^2\leq\sum_{t=1}^{T}[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_{g_i^0})'\theta_{g_i^0}]^2\right\}.$$

For $\mathcal{Z}_{ig}(\Theta,\mathbf{D})$, we have

$$\begin{aligned}
\mathcal{Z}_{ig}(\Theta,\mathbf{D}) &= \mathbf{1}\left(g_i^0\neq g\right)\mathbf{1}\left\{\sum_{t=1}^{T}\left[\tilde{z}_{it}(\gamma_{g_i^0})'(\theta_{g_i^0}-\theta_g)+[\tilde{x}_{it}(\gamma_{g_i^0})-\tilde{x}_{it}(\gamma_{g_i})]'\delta_g\right]\times\right. \\
&\qquad\left.\left[\tilde{z}_{it}(\gamma^0_{g_i^0})'\theta^0_{g_i^0}+\tilde{\varepsilon}_{it}-\frac{1}{2}[\tilde{z}_{it}(\gamma_g)'\theta_g+\tilde{z}_{it}(\gamma_{g_i^0})'\theta_{g_i^0}]\right]\leq 0\right\} \\
&\leq \max_{\tilde{g}\in\mathcal{G}\setminus\{g\}}\mathbf{1}\left(L_i(g,\tilde{g})\leq 0\right),
\end{aligned}$$

where

$$L_i(g,\tilde{g}) = \sum_{t=1}^{T} \left\{\tilde{z}_{it}(\gamma_{\tilde{g}})'(\theta_{\tilde{g}} - \theta_g) + [\tilde{x}_{it}(\gamma_{\tilde{g}}) - \tilde{x}_{it}(\gamma_g)]'\delta_g\right\} \{\tilde{z}_{it}(\gamma^0_{\tilde{g}})'\theta^0_{\tilde{g}} + \tilde{\varepsilon}_{it} - \frac{1}{2}[\tilde{z}_{it}(\gamma_g)'\theta_g + \tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}})]\}.$$

By adding and subtracting some terms, we have

$$L_i(g,\tilde{g}) = (\beta^0_{\tilde{g}} - \beta^0_g)' \sum_{t=1}^{T} \tilde{x}_{it} \left[\frac{1}{2}\tilde{x}'_{it}(\beta^0_{\tilde{g}} - \beta^0_g) + \tilde{\varepsilon}_{it}\right] + A_{iT}(g,\tilde{g}) + B_{iT}(g,\tilde{g}) + C_{iT}(g,\tilde{g}),$$

where

$$\begin{aligned}
A_{iT}(g,\tilde{g}) &= [(\beta_{\tilde{g}} - \beta^0_{\tilde{g}}) - (\beta_g - \beta^0_g)]' \sum_{t=1}^{T} \tilde{x}_{it} \left\{\tilde{z}_{it}(\gamma^0_{\tilde{g}})'\theta^0_{\tilde{g}} + \tilde{\varepsilon}_{it} - \frac{1}{2}[\tilde{z}_{it}(\gamma_g)'\theta_g + \tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}})]\right\}, \\
B_{iT}(g,\tilde{g}) &= \sum_{t=1}^{T} \left(\delta'_{\tilde{g}}\tilde{x}_{it}(\gamma_{\tilde{g}}) - \delta'_g\tilde{x}_{it}(\gamma_g)\right)' \left\{\tilde{z}_{it}(\gamma^0_{\tilde{g}})'\theta^0_{\tilde{g}} + \tilde{\varepsilon}_{it} - \frac{1}{2}[\tilde{z}_{it}(\gamma_g)'\theta_g + \tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}})]\right\}, \text{ and} \\
C_{iT}(g,\tilde{g}) &= (\beta^0_{\tilde{g}} - \beta^0_g)' \sum_{t=1}^{T} \tilde{x}_{it} \left\{\tilde{z}_{it}(\gamma^0_{\tilde{g}})'\theta^0_{\tilde{g}} + \tilde{\varepsilon}_{it} - \frac{1}{2}[\tilde{z}_{it}(\gamma_g)'\theta^0_g + \tilde{z}_{it}(\gamma_{\tilde{g}})'\theta^0_{\tilde{g}})]\right\} \\
&\quad -(\beta^0_{\tilde{g}} - \beta^0_g)' \sum_{t=1}^{T} \tilde{x}_{it}(\beta^0_{\tilde{g}} - \beta^0_g) \left[\frac{1}{2}\tilde{x}_{it}'(\beta^0_{\tilde{g}} - \beta^0_g) + \tilde{\varepsilon}_{it}\right].
\end{aligned}$$

For A_{iT}, have

$$\begin{aligned}
|A_{iT}(g,\tilde{g})| &\leq \left|[(\beta_{\tilde{g}} - \beta^0_{\tilde{g}}) - (\beta_g - \beta^0_g)]' \sum_{t=1}^{T} \tilde{x}_{it}\tilde{\varepsilon}_{it}\right| + \left|[(\beta_{\tilde{g}} - \beta^0_{\tilde{g}}) - (\beta_g - \beta^0_g)]' \sum_{t=1}^{T} \tilde{x}_{it}\tilde{z}_{it}(\gamma^0_{\tilde{g}})'\theta^0_{\tilde{g}}\right| \\
&\quad +\frac{1}{2}\left|[(\beta_{\tilde{g}} - \beta^0_{\tilde{g}}) - (\beta_g - \beta^0_g)]' \sum_{t=1}^{T} \tilde{x}_{it}(\tilde{z}_{it}(\gamma_g)'\theta_g + \tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}))\right| \\
&\equiv A_{iT,1}(g,\tilde{g}) + A_{T,2}(g,\tilde{g}) + A_{T,3}(g,\tilde{g}).
\end{aligned}$$

For $A_{T,1}$, we have

$$\begin{aligned}
A_{iT,1}(g,\tilde{g}) &\leq (\|\theta_{\tilde{g}} - \theta^0_{\tilde{g}}\| + \|\theta_g - \theta^0_g\|) \left\|\sum_{t=1}^{T} \tilde{x}_{it}\tilde{\varepsilon}_{it}\right\| \\
&\leq 2\sqrt{\eta}\left(\left\|\sum_{t=1}^{T} x_{it}\varepsilon_{it}\right\| + \frac{1}{T}\left\|\sum_{t=1}^{T}\sum_{s=1}^{T} x_{it}\varepsilon_{is}\right\|\right) \\
&\leq 2\sqrt{\eta}T\left\{\left\|\frac{1}{T}\sum_{t=1}^{T} x_{it}\varepsilon_{it}\right\| + \left\|\frac{1}{T}\sum_{t=1}^{T} x_{it}\right\|\left\|\frac{1}{T}\sum_{s=1}^{T}\varepsilon_{is}\right\|\right\} \\
&\leq 4\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}\|x_{it}\|^2 + \frac{1}{T}\sum_{t=1}^{T}\varepsilon^2_{it}\right)
\end{aligned}$$

where we used the fact that $\left\|\theta_g - \theta_g^0\right\| \leq \sqrt{\eta}$ for all $g \in \mathcal{G}$. Similarly, we have

$$\begin{aligned}
A_{iT,2}(g,\tilde{g}) &\leq 2\sqrt{\eta}T\left\|\theta_{\tilde{g}}^0\right\|\left\|\frac{1}{T}\sum_{t=1}^T \tilde{x}_{it}\tilde{z}_{it}(\gamma_{\tilde{g}}^0)'\right\| \leq 4\sqrt{\eta}T\left\|\theta_{\tilde{g}}^0\right\|\left(\frac{1}{T}\sum_{t=1}^T \left\|\tilde{x}_{it}\right\|^2\right) \\
&\leq 4\sqrt{\eta}T\left\|\theta_{\tilde{g}}^0\right\|\left(\frac{1}{T}\sum_{t=1}^T \left\|x_{it}\right\|^2\right), \text{ and} \\
A_{iT,3}(g,\tilde{g}) &\leq 4T\sqrt{\eta}(\|\theta_{\tilde{g}}\| + \|\theta_g\|)\frac{1}{T}\sum_{t=1}^T \left\|\tilde{x}_{it}\right\|^2 \leq 4T\sqrt{\eta}(\|\theta_{\tilde{g}}\| + \|\theta_g\|)\frac{1}{T}\sum_{t=1}^T \left\|x_{it}\right\|^2.
\end{aligned}$$

Thus, for any $\theta \in \mathcal{N}_\eta$,

$$|A_{iT}(g,\tilde{g})| \leq C_1\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^T(\left\|x_{it}\right\|^2 + \varepsilon_{it}^2)\right) \equiv H_{1,iT},$$

where C_1 is a positive constant independent of η and T.

For $B_{iT}(g,\tilde{g})$, we have

$$|B_{iT}(g,\tilde{g})| \leq (\|\delta_{\tilde{g}}\| + \|\delta_g\|)\sup_{\gamma\in\Gamma}\left\|\sum_{t=1}^T \tilde{x}_{it}(\gamma)\tilde{\varepsilon}_{it}\right\| + 2(\|\delta_{\tilde{g}}\| + \|\delta_g\|)\sup_{\gamma\in\Gamma}\left\|\sum_{t=1}^T \tilde{x}_{it}(\gamma_{\tilde{g}}^0)\tilde{z}_{it}(\gamma)\right\|.$$

Due to the fact $\theta \in \mathcal{N}_\eta$, we have $\|\delta_g\| \leq \left\|\delta_g - \delta_g^0\right\| + \left\|\delta_g^0\right\| \leq 2\sqrt{\eta}$ for all $g \in \mathcal{G}$. Following the analysis of A_{iT}, we can show that

$$|B_{iT}(g,\tilde{g})| \leq C_2\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^T(\left\|x_{it}\right\|^2 + \varepsilon_{it}^2)\right) \equiv H_{2,iT},$$

where C_2 is a positive constant independent of η and T. Analogously, we can show that

$$\begin{aligned}
|C_{iT}(g,\tilde{g})| &\leq \left|(\beta_{\tilde{g}}^0 - \beta_g^0)'\sum_{t=1}^T \tilde{x}_{it}\tilde{x}_{it}(\gamma_{\tilde{g}}^0)'\delta_{\tilde{g}}^0\right| + \frac{1}{2}\left|(\beta_{\tilde{g}}^0 - \beta_g^0)'\sum_{t=1}^T \tilde{x}_{it}\tilde{x}_{it}(\gamma_g)'\theta_g^0\right| + \frac{1}{2}\left|(\beta_{\tilde{g}}^0 - \beta_g^0)'\sum_{t=1}^T \tilde{x}_{it}\tilde{x}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}^0\right| \\
&\leq C_3\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^T(\left\|x_{it}\right\|^2 + \varepsilon_{it}^2)\right) \equiv H_{3,iT},
\end{aligned}$$

where C_3 is a positive constant independent of η and T. It follows that

$$\mathcal{Z}_{ig}(\Theta,\mathbf{D}) \leq \max_{\tilde{g}\in\boldsymbol{G}\backslash\{g\}} \mathbf{1}\left((\beta_{\tilde{g}}^0 - \beta_g^0)'\sum_{t=1}^T \tilde{x}_{it}\left[\frac{1}{2}\tilde{x}_{it}'(\beta_{\tilde{g}}^0 - \beta_g^0) + \tilde{\varepsilon}_{it}\right] \leq H_{iT}\right) \equiv \tilde{\mathcal{Z}}_{ig},$$

where $H_{iT} = C\sqrt{\eta}T[\frac{1}{T}\sum_{t=1}^T(\left\|x_{it}\right\|^2 + \varepsilon_{it}^2)]$ with $C = C_1 + C_2 + C_3$. Hence, we can conclude that

$$\sup_{(\Theta,\mathbf{D})\in\mathcal{N}_\eta\times\Gamma^G} \frac{1}{N}\sum_{i=1}^N \mathbf{1}\left(\hat{g}_i(\Theta,\mathbf{D}) \neq g_i^0\right) \leq \frac{1}{N}\sum_{g=1}^G\sum_{i=1}^N \tilde{\mathcal{Z}}_{ig}.$$

Noting that $\tilde{\mathcal{Z}}_{ig}$ does not depend on $(\Theta,\mathbf{D})$ and G is fixed, we are left to bound $\Pr(\tilde{\mathcal{Z}}_{ig} = 1)$.

Observe that

$$\Pr(\tilde{\mathcal{Z}}_{ig} = 1) \leq \sum_{\tilde{g}\in\boldsymbol{G}\backslash\{g\}} \Pr\left\{\xi_{iT}(g,\tilde{g}) \leq H_{iT}\right\},$$

where $\xi_{iT}(g,\tilde{g})=(\beta^0_{\tilde{g}}-\beta^0_g)'\sum_{t=1}^T \tilde{x}_{it}\left[\frac{1}{2}\tilde{x}_{it}'(\beta^0_{\tilde{g}}-\beta^0_g)+\tilde{\varepsilon}_{it}\right]$. Letting $C_4=2\max_{i,t}E(\|x_{it}\|^2+\varepsilon_{it}^2)$, we have

$$\begin{aligned}
\Pr(\tilde{\mathcal{Z}}_{ig} &= 1)\leq \sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{\xi_{iT}(g,\tilde{g})\leq H_{iT}\right\}\\
&\leq \sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{\xi_{iT}(g,\tilde{g})\leq H_{iT}, H_{iT}\leq 2E\left(H_{iT}\right)\right\}+\sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{\xi_{iT}(g,\tilde{g})\leq H_{iT}, H_{iT}> 2E\left(H_{iT}\right)\right\}\\
&\leq \sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{\xi_{iT}(g,\tilde{g})\leq 2E\left(H_{iT}\right)\right\}+\sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{H_{iT}> 2E\left(H_{iT}\right)\right\}\\
&\leq \sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{\xi_{iT}(g,\tilde{g})\leq CC_4\sqrt{\eta}T\right\}+\sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left(\frac{1}{T}\sum_{t=1}^T(\|x_{it}\|^2+\varepsilon_{it}^2)\geq C_4\right).
\end{aligned}$$

Using the fact $(\beta^0_{\tilde{g}}-\beta^0_g)'\sum_{t=1}^T\tilde{x}_{it}\tilde{\varepsilon}_{it}=(\beta^0_{\tilde{g}}-\beta^0_g)'[\sum_{t=1}^T x_{it}\varepsilon_{it}-\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T x_{is}\varepsilon_{it}]$, we have

$$\begin{aligned}
\Pr(\tilde{\mathcal{Z}}_{ig} &= 1)\leq \sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\left\{\Pr\left(\frac{1}{T}\sum_{t=1}^T(\|x_{it}\|^2+\varepsilon_{it}^2)\geq C_4\right)\right.\\
&\quad+\Pr\left(\frac{1}{T}\sum_{t=1}^T[\tilde{x}_{it}'(\beta^0_{\tilde{g}}-\beta^0_g)]^2\leq\frac{c_{g,\tilde{g}}}{2}\right)\\
&\quad+\Pr\left((\beta^0_{\tilde{g}}-\beta^0_g)'\frac{1}{T}\sum_{t=1}^T x_{it}\varepsilon_{it}\leq-\frac{c_{g,\tilde{g}}}{4}+\frac{CC_4}{2}\sqrt{\eta}\right)\\
&\quad\left.+\Pr\left(-(\beta^0_{\tilde{g}}-\beta^0_g)'\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T x_{is}\varepsilon_{it}\leq-\frac{c_{g,\tilde{g}}}{4}+\frac{CC_4}{2}\sqrt{\eta}\right)\right\}.
\end{aligned}\tag{B.1}$$

By Assumptions A.1 and A.3, we can use Lemma B.1 in the next section to show the first two terms to be $o(T^{-4})$. To study the third term on the RHS of the last inequality, we take η such that $\eta\leq\left[\min_{g\in\mathcal{G}}\left(\frac{\min_{\tilde{g}\in\mathcal{G}\backslash\{g\}}c_{g,\tilde{g}}}{8CC_4}\right)\right]^2$. Then we have that for any $g\neq\tilde{g}\in\mathcal{G}$,

$$\begin{aligned}
\Pr\left((\beta^0_{\tilde{g}}-\beta^0_g)'\frac{1}{T}\sum_{t=1}^T x_{it}\varepsilon_{it}\leq-\frac{c_{g,\tilde{g}}}{4}+\frac{CC_4}{2}\sqrt{\eta}\right) &\leq \Pr\left((\beta^0_{\tilde{g}}-\beta^0_g)'\frac{1}{T}\sum_{t=1}^T x_{it}\varepsilon_{it}\leq-\frac{c_{g,\tilde{g}}}{8}\right)\\
&\leq \Pr\left(\left|(\beta^0_{\tilde{g}}-\beta^0_g)'\frac{1}{T}\sum_{t=1}^T x_{it}\varepsilon_{it}\right|\geq\frac{c_{g,\tilde{g}}}{8}\right)\\
&= o(T^{-4}),
\end{aligned}$$

where the last equality follows by another application of Lemma B.1 and the fact that $\left\|\beta^0_{\tilde{g}}-\beta^0_g\right\|\geq\underline{c}_\beta>0$ under Assumption A.3(i). Similarly, we can show that the last term on the RHS of (B.1) is $o(T^{-4})$. Then we have

$$E\left(\sup_{(\Theta,\mathbf{D})\in\mathcal{N}_\eta\times\Gamma^G}\frac{1}{N}\sum_{i=1}^N\mathbf{1}\{\hat{g}_i(\Theta,\mathbf{D})\neq g_i^0\}\right)\leq E\left(\frac{1}{N}\sum_{g=1}^G\sum_{i=1}^N\tilde{\mathcal{Z}}_{ig}\right)=\frac{1}{N}\sum_{g=1}^G\sum_{i=1}^N\Pr(\tilde{\mathcal{Z}}_{ig}=1)=o(T^{-4}).$$

Lastly, by Markov inequality,

$$\begin{aligned}
\Pr\left(\sup_{(\Theta,\mathbf{D})\in\mathcal{N}_\eta\times\Gamma^G}\frac{1}{N}\sum_{i=1}^N\mathbf{1}(\hat{g}_i(\Theta,\mathbf{D})\neq g_i^0)>\epsilon T^{-4}\right) &\leq \Pr\left(\frac{1}{N}\sum_{g=1}^G\sum_{i=1}^N\Pr(\tilde{\mathcal{Z}}_{ig}=1)>\epsilon T^{-4}\right)\\
&\leq \frac{E\left(\frac{1}{N}\sum_{g=1}^G\sum_{i=1}^N\tilde{\mathcal{Z}}_{ig}\right)}{\epsilon T^{-4}}=o(1),
\end{aligned}$$

for any constant $\epsilon > 0$. This completes our proof. ■.

Proof of Lemma A.4. By Markov inequality, we have

$$\Pr(\sup_{i,t} \|w_{it}\| > \eta\,(NT)^{1/3}) \leq \frac{1}{\eta^{3+\epsilon}\,(NT)^{(3+\epsilon)/3}} \sum_{i,t} E\,\|w_{it}\|^{3+\epsilon} = o\,(1)\,,$$

implying that $\sup_{i,t} \|w_{it}\| = o_p((NT)^{1/3})$. By Lemma A.3 and the order requirement on N and T, we have

$$\begin{aligned} \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(\hat{g}_i \neq g_i^0) w_{it} \right\| &\leq \left(\sup_{(\Theta,\mathbf{D}) \in \mathcal{N}_\eta \times \Gamma^G} \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\hat{g}_i(\Theta,\mathbf{D}) \neq g_i^0) \right) \sup_{i,t} \|w_{it}\| \\ &= o_p(T^{-4}(NT)^{1/3}) = o_p((NT)^{-1}).\ ■. \end{aligned}$$

Proof of Lemma A.5. By direct calculations, we have

$$\begin{aligned} \sup_{\gamma\in\Gamma} \left(\frac{1}{NT} \sum_{i,t} E\,\|\tilde{z}_{it}(\gamma)\tilde{z}_{it}(\gamma)'\|^4 \right) &= \sup_{\gamma\in\Gamma} \left(\frac{1}{NT} \sum_{i,t} E\,\|\tilde{z}_{it}(\gamma)\|^8 \right) \leq \sup_{\gamma\in\Gamma} \left(\frac{1}{NT} \sum_{i,t} E\,\|\tilde{z}_{it}(\gamma)\|^8 \right) \\ &\leq \frac{256}{NT} \sum_{i,t} E\,\|x_{it}\|^8 \leq C < \infty \text{ by Assumption A.1(v).} \end{aligned}$$

Similarly, $\sup_{\gamma\in\Gamma} \left(\frac{1}{NT} \sum_{i,t} E\,\|\tilde{z}_{it}(\gamma)\tilde{y}_{it}\|^4 \right) \leq C < \infty$. Then we apply Lemma A.4 with $\epsilon = 1$ to obtain

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(\hat{g}_i = g)\tilde{z}_{it}(\hat{\gamma}_g)\tilde{y}_{it} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(g_i^0 = g)\tilde{z}_{it}(\hat{\gamma}_g)\tilde{y}_{it} + o_p((NT)^{-1}).$$

Analogously, we have

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(\hat{g}_i = g)\tilde{z}_{it}(\hat{\gamma}_g)\tilde{z}_{it}(\hat{\gamma}_g)' = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(g_i^0 = g)\tilde{z}_{it}(\hat{\gamma}_g)\tilde{z}_{it}(\hat{\gamma}_g)' + o_p((NT)^{-1}).$$

To sum up, we have $\hat{\theta}_g = \check{\theta}_g(\hat{\gamma}_g) + o_p((NT)^{-1})$. ■.

Proof of Lemma A.6. Note that

$$\begin{aligned} \frac{1}{NT}\check{\mathcal{Q}}(\check{\Theta}(\hat{\mathbf{D}}),\hat{\mathbf{D}}) &= \frac{1}{NT}\check{\mathcal{Q}}(\hat{\Theta},\hat{\mathbf{D}}) + o_p((NT)^{-1}) = \frac{1}{NT}\mathcal{Q}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}}) + o_p((NT)^{-1}) \\ &\leq \frac{1}{NT}\check{\mathcal{Q}}(\check{\Theta},\check{\mathbf{D}}) + o_p((NT)^{-1}), \end{aligned}$$

where the first and second equalities hold by Lemmas A.5 and A.4, respectively, and the inequality holds by the definition of least squares estimator $(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}})$. On the other hand,

$$\frac{1}{NT}\check{\mathcal{Q}}_g(\check{\Theta},\check{\mathbf{D}}) = \frac{1}{NT}\check{\mathcal{Q}}_g(\check{\Theta}(\check{\mathbf{D}}),\check{\mathbf{D}}) \leq \frac{1}{NT}\check{\mathcal{Q}}_g(\check{\Theta}(\hat{\mathbf{D}}),\hat{\mathbf{D}})$$

by the fact that $\check{\mathcal{Q}}_g(\check{\theta}_g,\check{\gamma}_g) = \inf_{(\theta,\gamma)} \check{\mathcal{Q}}_g(\theta,\gamma))$. It follows that

$$\frac{1}{NT}\left[\check{\mathcal{Q}}_g(\check{\theta}_g(\check{\gamma}_g),\check{\gamma}_g) - \check{\mathcal{Q}}_g(\check{\theta}_g(\hat{\gamma}_g),\hat{\gamma}_g)\right] = o_p((NT)^{-1}), \text{ for all } g \in \mathcal{G}. \tag{B.2}$$

Following the analysis of the infeasible estimator $\check{\gamma}_g$ in Lemma C.10 in the online Supplementary Material, we can also show that $\hat{\gamma}_g - \gamma_g^0 = O_p(1/\alpha_{NT})$ based on (B.2). ■

Proof of Lemma A.7. For all $g \in \mathcal{G}$, we have

$$\check{\theta}_g(\gamma) - \theta_g^0 = [\Phi_{1g}(\gamma)]^{-1} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 \varepsilon_i - [\Phi_{1g}(\gamma)]^{-1} \Phi_{2g}(\gamma) \delta_g^0,$$

where $\Phi_{1g}(\gamma) \equiv \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 Z_i(\gamma)$, $\Phi_{2g}(\gamma) \equiv \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 X_i(\gamma, \gamma_g^0)$ and $X_i(\gamma, \gamma_g^0) = X_i(\gamma) - X_i(\gamma_g^0)$. It is easy to show that

$$\begin{aligned}
\Phi_{1g}(\gamma) &= \Phi_{1g}(\gamma_g^0) + O_p(\alpha_{N_g T}^{-1}) = O_p(1), \\
\frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma_g^0)' \mathbb{M}_0 \varepsilon_i &= O_p((NT)^{-1/2} + T^{-1}), \\
\frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \left[Z_i(\gamma) - Z_i(\gamma_g^0) \right]' \mathbb{M}_0 \varepsilon_i &= O_p(\alpha_{N_g T}^{-1} [(NT)^{-1/2} + T^{-1}]),
\end{aligned}$$

and $\Phi_{2g}(\gamma) \delta_g^0 = (NT)^{-\alpha} \Phi_{2g}(\gamma) C_g^0 = (NT)^{-\alpha} O_p(\alpha_{N_g T}^{-1}) = O_p((NT)^{-1+\alpha})$, where we use the fact that $\alpha_{N_g T} = (N_g T)^{1-2\alpha}$ and $N_g / N \to \pi_g > 0$. With these results, we can readily show that

$$\begin{aligned}
\check{\theta}_g(\gamma) - \check{\theta}_g(\gamma_g^0) &= \left\{ [\Phi_{1g}(\gamma)]^{-1} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 \varepsilon_i - \left[\Phi_{1g}(\gamma_g^0) \right]^{-1} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma_g^0)' \mathbb{M}_0 \varepsilon_i \right\} \\
&\quad - [\Phi_{1g}(\gamma)]^{-1} \Phi_{2g}(\gamma) \delta_g^0 \\
&= \left\{ [\Phi_{1g}(\gamma)]^{-1} - \left[\Phi_{1g}(\gamma_g^0) \right]^{-1} \right\} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma_g^0)' \mathbb{M}_0 \varepsilon_i \\
&\quad + [\Phi_{1g}(\gamma)]^{-1} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \left[Z_i(\gamma) - Z_i(\gamma_g^0) \right] \mathbb{M}_0 \varepsilon_i - [\Phi_{1g}(\gamma)]^{-1} \Phi_{2g}(\gamma) \delta_g^0 \\
&= O_p(\alpha_{N_g T}^{-1} [(NT)^{-1/2} + T^{-1}]) + O_p(\alpha_{N_g T}^{-1} [(NT)^{-1/2} + T^{-1}]) + O_p((NT)^{-1+\alpha}) \\
&= O_p(\alpha_{N_g T}^{-1} [(NT)^{-1/2} + T^{-1}]) + O_p((NT)^{-1+\alpha}) = o_p((NT)^{-1/2})
\end{aligned}$$

where the last equality follows from the fact that $\alpha \in (0, 1/3)$ and $N = O(T^2)$. The above analysis also shows that $\check{\theta}_g(\gamma_g^0) - \theta_g^0 = O_p\left((NT)^{-1/2} + T^{-1} \right)$.

Next, noting that

$$\check{Q}_g(\theta, \gamma) = \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} [\tilde{y}_{it} - \tilde{z}_{it}(\gamma)' \theta]^2 = \sum_{i \in \mathbf{G}_g^0} [Y_i - Z_i(\gamma) \theta]' \mathbb{M}_0 [Y_i - Z_i(\gamma) \theta],$$

$A_i' \mathbb{M}_0 A_i - B_i' \mathbb{M}_0 B_i = (A_i - B_i)' \mathbb{M}_0 (A_i - B_i) + 2(A_i - B_i)' \mathbb{M}_0 B_i$ for any two $T \times 1$ vectors A_i and B_i, and

$Y_i - Z_i(\gamma)\theta = [X_i(\gamma) - X_i(\gamma_g^0)]\delta_g^0 + \mu_i\iota_T + \varepsilon_i$ with ι_T being a $T \times 1$ vector of ones, we have

$$\begin{aligned}
\check{Q}_g(\check{\theta}_g(\gamma), \gamma) - \check{Q}_g(\check{\theta}_g, \gamma) &= \sqrt{NT}[\check{\theta}_g(\gamma) - \check{\theta}_g]' \frac{1}{NT} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 Z_i(\gamma) \sqrt{NT}[\check{\theta}_g(\gamma) - \check{\theta}_g] \\
&\quad + 2NT[\check{\theta}_g(\gamma) - \check{\theta}_g]' \frac{1}{NT} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 [X_i(\gamma) - X_i(\gamma_g^0)]\delta_g^0 \\
&\quad + 2NT[\check{\theta}_g(\gamma) - \check{\theta}_g]' \frac{1}{NT} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 \varepsilon_i \\
&= o_p(1) + NT o_p((NT)^{-1/2}) O_p((NT)^{-1+\alpha}) \\
&\quad + NT[O_p(\alpha_{N_gT}^{-1}[(NT)^{-1/2} + T^{-1}]) + O_p((NT)^{-1+\alpha})] O_p((NT)^{-1/2} + T^{-1}) \\
&= o_p(1),
\end{aligned}$$

where the last equality follows from the fact that $\alpha \in (0, 1/3)$ and $N = O(T^2)$. ■

Proof of Lemma A.8. (i) Let $\mathbb{P}_0 = \frac{1}{T}\iota_T\iota_T'$. Note that

$$\begin{aligned}
\frac{1}{\sqrt{N_gT}} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma_g^0)' \mathbb{M}_0 \varepsilon_i &= \frac{1}{\sqrt{N_gT}} \sum_{i \in \mathbf{G}_g^0} [Z_i(\gamma_g^0) - \mathbb{P}_0 E(Z_i(\gamma_g^0))]' \varepsilon_i \qquad \text{(B.3)} \\
&\quad - \frac{1}{T\sqrt{N_gT}} \sum_{i \in \mathbf{G}_g^0} \sum_{s,t=1}^{T} \{z_{it}(\gamma_g^0) - E\left[z_{it}(\gamma_g^0)\right]\} \varepsilon_{is} \equiv A_1 - A_2.
\end{aligned}$$

It suffices to show that (i1) $A_1 \xrightarrow{d} N(0, \Omega_{g,1}^0)$ and (i2) $A_2 = \sqrt{\frac{N_g}{T}}\mathbb{B}_{g,NT} + o_p(1)$. To prove (i1), we relabel the index $\mathbf{G}_g^0 = \{i_1, ..., i_{N_g}\}$ to $\{1, ..., N_g\}$. Let c denote a $2K \times 1$ nonrandom vector with $\|c\| = 1$. For $m = (i-1)T + t$ for $t = 1, ..., T$ and $i = 1, ..., N_g$, let $\zeta_m = \left[z_{it}(\gamma_g^0) - \frac{1}{T}\sum_{t=1}^T E(z_{it}(\gamma_g^0))\right]\varepsilon_{it}$. Let $M = N_gT$. Then we have

$$c'A_1 = \frac{1}{\sqrt{M}} \sum_{m=1}^{M} c'\zeta_m.$$

Immediately, $\{\zeta_m\}_{m=1}^M$ is a martingale difference sequence (m.d.s.) under the filtration $\mathcal{F}_m = \sigma(\{\zeta_n : 1 \leq n \leq m\})$, the minimal sigma-field generated from $\{\zeta_n : 1 \leq n \leq m\}$. Apparently, $\max_{1 \leq m \leq M} E\|\zeta_m\|^4 \leq C$ for some $C < \infty$ under Assumption A.1. In addition,

$$\begin{aligned}
&\frac{1}{M} \sum_{m=1}^{M} c'\zeta_m\zeta_m'c \\
= \; & c' \frac{1}{N_gT} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \tilde{z}_{it}(\gamma_g^0)\tilde{z}_{it}(\gamma_g^0)' c \varepsilon_{it}^2 \\
&+ c' \frac{1}{N_gT} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \varepsilon_{it}^2 \left[2z_{it}(\gamma_g^0) - \frac{1}{T}\sum_{s=1}^{T}\{z_{is}(\gamma_g^0) - E[z_{is}(\gamma_g^0)]\}\right] \left[\frac{1}{T}\sum_{s=1}^{T}\{z_{is}(\gamma_g^0) - E[z_{is}(\gamma_g^0)\}'\right] c \\
\equiv \; & A_{1,1} + A_{1,2}.
\end{aligned}$$

By Assumption A.6, $A_{1,1} \xrightarrow{p} c'\Omega_{g,1}(\gamma_g^0, \gamma_g^0)c$. For $A_{1,2}$, we have by Cauchy-Schwarz and Markov inequalities

$$
\begin{aligned}
|A_{1,2}| \quad &\leq \quad \|c\|^2 \left(\frac{1}{N_g T} \sum_{i \in G_g^0} \sum_{t=1}^{T} \varepsilon_{it}^4 \left\| \left[2z_{it}(\gamma_g^0) - \frac{1}{T} \sum_{s=1}^{T} \{ z_{is}(\gamma_g^0) + E[z_{is}(\gamma_g^0)] \} \right] \right\|^2 \right)^{1/2} \\
&\times \left(\frac{1}{N_g T} \sum_{i \in G_g^0} \sum_{t=1}^{T} \left\| \frac{1}{T} \sum_{s=1}^{T} \{ z_{is}(\gamma_g^0) - E[z_{is}(\gamma_g^0)] \} \right\|^2 \right)^{1/2} = O_P(1)\, O_p\left(T^{-1/2}\right) = o_p(1).
\end{aligned}
$$

Then $A_1 \xrightarrow{d} N(0, \Omega_{g,1}(\gamma_g^0, \gamma_g^0))$ by the Cramér-Wold device and the martingale central limit theorem.

Next, we consider A_2. Note that

$$
\begin{aligned}
A_2 \quad &= \quad \frac{1}{T\sqrt{N_g T}} \sum_{i \in \mathbf{G}_g^0} \sum_{s,t=1}^{T} E\left[z_{it}(\gamma_g^0) \varepsilon_{is} \right] + \frac{1}{T\sqrt{N_g T}} \sum_{i \in \mathbf{G}_g^0} \sum_{s,t=1}^{T} \left\{ \left(z_{it}(\gamma_g^0) - E[z_{it}(\gamma_g^0)] \right) \varepsilon_{is} - E\left[z_{it}(\gamma_g^0) \varepsilon_{is} \right] \right\} \\
&\equiv \quad A_{2,1} + A_{2,2}.
\end{aligned}
$$

For $A_{2,1}$, we have $A_{2,1} = \sqrt{\frac{N_g}{T}} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \sum_{s<t} E\left[z_{it}(\gamma_g^0) \varepsilon_{is} \right] = \sqrt{\frac{N_g}{T}} \mathbb{B}_{g,NT}$. For $A_{2,2}$, we can easily verify that $E(A_{2,2}) = 0$ and

$$
E\|A_{2,2}\|^2 = \frac{1}{N_g T^3} \sum_{i \in \mathbf{G}_g^0} E \left\| \sum_{s,t=1}^{T} \left[\{ z_{it}(\gamma_g^0) - E[z_{it}(\gamma_g^0)] \} \varepsilon_{is} - E\left[z_{it}(\gamma_g^0) \varepsilon_{is} \right] \right] \right\|^2 = O\left(T^{-1}\right)
$$

by using the Davydov inequality for strong mixing processes. Then $A_{2,2} = O_p(T^{-1/2})$ and (i2) follows.

(ii) Now, let $u_i \equiv Z_i(\gamma_g^0)' \mathbb{M}_0 \varepsilon_i / \sqrt{N_g T}$. Then we have u_i independent across i and

$$
\begin{aligned}
E\|u_i\|^2 \quad &= \quad \frac{1}{N_g T} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left(\tilde{z}_{it}(\gamma_g^0) \tilde{z}_{is}(\gamma_g^0)' \varepsilon_{it} \varepsilon_{is} \right) \\
&\leq \quad \frac{1}{N_g T} \sum_{t=1}^{T} \sum_{s=1}^{T} 6\alpha[|t-s|]^{1/2} \left\| \tilde{z}_{it}(\gamma_g^0) \varepsilon_{it} \right\|_4 \left\| \tilde{z}_{is}(\gamma_g^0) \varepsilon_{is} \right\|_4 \\
&= \quad O(1/N_g).
\end{aligned}
$$

By Theorem A of Yang (2016), we have $E \max_i \|u_i\|^{2+\delta} \leq C(N_g)^{-(2+\delta)/2} \max_i \max_{1 \leq t \leq T} \left\| \tilde{z}_{it}(\gamma_g^0) \varepsilon_{it} \right\|_{2+2\delta}$ for some $\delta > 0$ and $C < \infty$. Here $\|\cdot\|_r = \{E\|\cdot\|^r\}^{1/r}$. Then Lindeberg condition holds and we have the desired claim. ■

C Supplementary Lemmas

We first state a technical lemma that is also used in the proof the main results in the paper. Then we study the asymptotic properties of the infeasible estimators.

C.1 A technical lemma

Lemma C.1. Let ξ_t denote a $d_\xi \times 1$ random vector with mean zero and $E\|\xi_t\|^{8+\epsilon} < \infty$ for some $\epsilon > 0$. Suppose that $\{\xi_t, t = 1, ..., T\}$ is strong mixing process with mixing coefficients $\alpha[s] \leq c_\alpha \rho^s$ for some $c_\alpha > 0$

and $\rho \in (0,1)$. Then as $T \to \infty$ and for any $c > 0$ we have

$$\Pr\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_t\right\| > c\right) = o(T^{-4}).$$

Proof of Lemma C.1. The proof is similar to and simpler than that of Lemma B.1(ii) in Wang, Phillips, and Su (2018) and thus omitted. ■

C.2 Asymptotic properties of the infeasible estimators

We present the analysis of infeasible estimator in this section.

Lemma C.2. Suppose Assumptions A.1, A.3(iv) and A.4 hold. For any $g \in \mathcal{G}$, we have that

$$\check{\gamma}_g - \gamma_g^0 = o_p(1), \text{ and } \check{\theta}_g - \theta_g^0 = o_p((NT)^{-\alpha}).$$

Proof of Lemma C.2. First, we show the convergence rate of $\check{\theta}_g(\gamma)$ for any $\gamma \in \Gamma$. Let $Z_i(\gamma) \equiv ([x_{i1}', x_{i1}'d_{i1}(\gamma)]', ..., [x_{iT}', x_{iT}'d_{iT}(\gamma)]')'$, a $T\times 2K$ matrix. Let $X_i(\gamma_1,\gamma_2) \equiv (x_{i1}[d_{i1}(\gamma_1)-d_{i1}(\gamma_2)], ..., x_{iT}[d_{iT}(\gamma_1) -d_{iT}(\gamma_2)])'$, a $T\times K$ matrix. By the definition of $\check{\theta}_g(\gamma)$, we have $\check{\theta}_g(\gamma) = [\sum_{i\in \mathbf{G}_g^0} Z_i(\gamma)'\mathbb{M}_0 Z_i(\gamma)]^{-1}\sum_{i\in\mathbf{G}_g^0} Z_i(\gamma)'\mathbb{M}_0 Y_i$. It follows that

$$\check{\theta}_g(\gamma) - \theta_g^0 = [\Phi_{1g}(\gamma)]^{-1}\frac{1}{N_gT}\sum_{i\in\mathbf{G}_g^0} Z_i(\gamma)'\mathbb{M}_0\varepsilon_i - [\Phi_{1g}(\gamma)]^{-1}\Phi_{2g}(\gamma)\delta_g^0. \tag{C.1}$$

where $\Phi_{1g}(\gamma) \equiv \frac{1}{N_gT}\sum_{i\in\mathbf{G}_g^0} Z_i(\gamma)'\mathbb{M}_0 Z_i(\gamma)$, $\Phi_{2g}(\gamma) = \frac{1}{N_gT}\sum_{i\in\mathbf{G}_g^0} Z_i(\gamma)'\mathbb{M}_0 X_i(\gamma,\gamma_g^0)$. By Assumption A.4(i), $\Phi_{1g}(\gamma) = O_p(1)$ for all $\gamma \in \Gamma$. It is standard to show that $\frac{1}{N_gT}\sum_{i\in\mathbf{G}_g^0} Z_i(\gamma)'\mathbb{M}_0\varepsilon_i = O_p((NT)^{-1/2}+T^{-1})$ and $\Phi_{2g}(\gamma) = O_p(1)$. Then we have $\check{\theta}_g(\gamma)-\theta_g^0 = O_p((NT)^{-\alpha}+T^{-1})$ by exploiting the fact that $\delta_g^0 = O(((NT)^{-\alpha})$. Given the fact that $\alpha < 1/3$ and $N = O(T^2)$, we can conclude from (C.1) that $\check{\theta}_g(\gamma) - \theta_g^0 = O_p((NT)^{-\alpha})$ and

$$\check{\theta}_g(\gamma) - \theta_g^0 = -[\Phi_1(\gamma)]^{-1}\Phi_2(\gamma)\delta_g^0 + o_p((NT)^{-\alpha}). \tag{C.2}$$

Next we show the consistency of $\check{\gamma}_g$. Let $\Phi_{3g}(\gamma) = \frac{1}{N_gT}\sum_{i\in\mathbf{G}_g^0} X_i(\gamma,\gamma_g^0)'\mathbb{M}_0 X_i(\gamma,\gamma_g^0)$. By direct calculations, we can show that

$$\begin{aligned}
& \frac{1}{N_gT}\left(\check{\mathcal{Q}}_g(\check{\theta}_g,\check{\gamma}_g) - \check{\mathcal{Q}}_g(\theta_g^0,\gamma_g^0)\right) \\
= \;& \delta_g^{0\prime}\Phi_{3g}(\check{\gamma}_g)\delta_g^0 + (\check{\theta}_g-\theta_g^0)'\Phi_{1g}(\check{\gamma}_g)(\check{\theta}_g-\theta_g^0) + 2(\check{\theta}_g-\theta_g^0)'\Phi_{2g}(\check{\gamma}_g)\delta_g^0 \\
& -(\check{\theta}_g-\theta_g^0)'\frac{2}{N_gT}\sum_{i\in\mathbf{G}_g^0} Z_i(\check{\gamma}_g)'\mathbb{M}_0\varepsilon_i - \check{\delta}_g'\frac{2}{N_gT}\sum_{i\in\mathbf{G}_g^0} X_i(\check{\gamma}_g,\gamma_g^0)'\mathbb{M}_0\varepsilon_i.
\end{aligned} \tag{C.3}$$

Note that the last two terms on the right hand side (RHS) of the above equation are $o_p((NT)^{-2\alpha})$. This, in conjunction with (C.2) and (C.3) implies that,

$$\frac{1}{N_gT}\left(\check{\mathcal{Q}}_g(\check{\theta}_g,\check{\gamma}_g) - \check{\mathcal{Q}}_g(\theta_g^0,\gamma_g^0)\right) = \delta_g^{0\prime}[\Phi_{3g}(\check{\gamma}_g) - \Phi_{2g}(\check{\gamma}_g)'\Phi_{1g}(\check{\gamma}_g)^{-1}\Phi_{2g}(\check{\gamma}_g)]\delta_g^0 + o_p((NT)^{-2\alpha}).$$

By Assumption A.4(ii), we have that

$$\Phi_3(\check{\gamma}_g) - \Phi_2(\check{\gamma}_g)'\Phi_1(\check{\gamma}_g)^{-1}\Phi_2(\check{\gamma}_g) = \tilde{w}_g(\check{\gamma}_g),$$

which is a $K \times K$ matrix with minimum eigenvalue $\lambda_{\min}[\tilde{w}_g(\check{\gamma}_g)] \geq \tau \min\{1, |\check{\gamma}_g - \gamma_g^0|\}$ w.p.a.1. Hence it follows that

$$(N_g T)^{2\alpha-1} \left(\check{\mathcal{Q}}_g(\check{\theta}_g, \check{\gamma}_g) - \check{\mathcal{Q}}_g(\theta_g^0, \gamma_g^0)\right) \geq \pi_g^{2\alpha} \left\|C_g^0\right\| \tau \min\{1, |\check{\gamma}_g - \gamma_g^0|\} + o_p(1),$$

where we use the fact $\delta_g^0 = (NT)^{-\alpha}$ and $N_g/N \to \pi_g$ by Assumptions A.1(vi) and A.2(iii). On the other hand, we have $\check{\mathcal{Q}}_g(\check{\theta}_g, \check{\gamma}_g) - \check{\mathcal{Q}}_g(\check{\theta}_g, \gamma_g^0) \leq 0$. We can conclude that $\check{\gamma}_g - \gamma_g^0 = o_p(1)$.

Given the consistency of $\check{\gamma}_g$, we can easily show that $\frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\check{\gamma}_g)' \mathbb{M}_0 X_i(\gamma_g^0, \check{\gamma}_g) = o_p(1)$. Then $\check{\theta}_g - \theta_g^0 = o_p((NT)^{-\alpha})$ follows. ■

Lemma C.3. Let $h_{it}(\gamma_1, \gamma_2) = \|x_{it}\varepsilon_{it}\| \, |d_{it}(\gamma_2) - d_{it}(\gamma_1)|$ and $k_{it}(\gamma_1, \gamma_2) = \|x_{it}\| \, |d_{it}(\gamma_2) - d_{it}(\gamma_1)|$. Suppose Assumptions A.1(v) and A.5 hold, there is a constant $C_1 < \infty$ such that for $\underline{\gamma} \leq \gamma_1 < \gamma_2 \leq \overline{\gamma}$ and $r \leq 4$,

$$\max_{i,t} E\left[h_{it}(\gamma_1, \gamma_2)\right]^r \leq C_1 |\gamma_2 - \gamma_1| \text{ and } \max_{i,t} E\left[k_{it}(\gamma_1, \gamma_2)\right]^r \leq C_1 |\gamma_2 - \gamma_1| .$$

Proof of Lemma C.3. For any random variable Z,

$$E\left[Z d_{it}(\gamma)\right] = E(Z \cdot 1\{q_{it} \leq \gamma\}) = E\left[1\{q_{it} \leq \gamma\} E(Z|q_{it})\right] = \int_{-\infty}^{\gamma} E(Z|q_{it}) dF_{it}(q_{it}),$$

where $F_{it}(\cdot)$ is the cumulative distribution function (CDF) of q_{it} with the corresponding PDF $f_{it}(\cdot)$. Taking derivative with respect to γ on both sides yields

$$\frac{d}{d\gamma} E\left[Z d_{it}(\gamma)\right] = E(Z|q_{it} = \gamma) f_{it}(\gamma).$$

Then by the Hölder inequality and Assumptions A.1(v) and A.5

$$\begin{aligned} \frac{d}{d\gamma} E\left[\|x_{it}\varepsilon_{it}\|^r d_{it}(\gamma)\right] &= E(\|x_{it}\varepsilon_{it}\|^r |q_{it} = \gamma) f_{it}(\gamma) \leq [E(\|x_{it}\varepsilon_{it}\|^4 |q_{it} = \gamma)]^{r/4} f_{it}(\gamma) \\ &\leq C c_f \text{ for some } C < \infty \end{aligned}$$

This implies that

$$\max_{i,t} E\left[h_{it}(\gamma_1, \gamma_2)\right]^r \leq C_1 |\gamma_2 - \gamma_1| \text{ with } C_1 = C c_f.$$

Analogously, we have $\max_{i,t} E\left[k_{it}(\gamma_1, \gamma_2)\right]^r \leq C_1 |\gamma_2 - \gamma_1|$. ■

Lemma C.4. Suppose Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold. Then there exists a constant $C_2 < \infty$ such that for all $\underline{\gamma} \leq \gamma_1 < \gamma_2 \leq \overline{\gamma}$ and $g \in \mathcal{G}$

$$\begin{aligned} E\left|\frac{1}{\sqrt{N_g T}} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \left(h_{it}^2(\gamma_1, \gamma_2) - E h_{it}^2(\gamma_1, \gamma_2)\right)\right|^2 &\leq C_2 |\gamma_2 - \gamma_1| , \\ E\left|\frac{1}{\sqrt{N_g T}} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \left(k_{it}^2(\gamma_1, \gamma_2) - E k_{it}^2(\gamma_1, \gamma_2)\right)\right|^2 &\leq C_2 |\gamma_2 - \gamma_1| . \end{aligned}$$

Proof of Lemma C.4. For notational simplicity, let $h_{it}^r(\gamma_1, \gamma_2) = [h_{it}(\gamma_1, \gamma_2)]^r$ for $r \geq 0$. By the independence across i and strong mixing over t for $\{(x_{it}, q_{it}, \varepsilon_{it})\}$, there is a constant $C^{\dagger}$ such that

$$
\begin{aligned}
& E\left|\frac{1}{\sqrt{N_g T}} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \{h_{it}^2(\gamma_1, \gamma_2) - E[h_{it}^2(\gamma_1, \gamma_2)]\}\right|^2 \\
= \; & \frac{1}{N_g} \sum_{i \in \mathbf{G}_g^0} E\left|\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \{h_{it}^2(\gamma_1, \gamma_2) - E[h_{it}^2(\gamma_1, \gamma_2)]\}\right|^2 \\
\leq \; & \frac{C^{\dagger}}{N_g} \sum_{i \in \mathbf{G}_g^0} \frac{1}{T} \sum_{t=1}^{T} E\left\{h_{it}^2(\gamma_1, \gamma_2) - E\left[h_{it}^2(\gamma_1, \gamma_2)\right]\right\}^2 \\
\leq \; & \frac{C^{\dagger}}{N_g T} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} E\left[h_{it}^4(\gamma_1, \gamma_2)\right] \leq C^{\dagger} C_1 \left|\gamma_2 - \gamma_1\right|.
\end{aligned}
$$

The first result follows by setting $C_2 = C^{\dagger} C_1$. Analogously, we can prove the second result in the lemma. ■

Lemma C.5. Let $J_{g,NT}(\gamma) = N_g^{-1/2} T^{-1/2} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} x_{it} e_{it} d_{it}(\gamma)$. Suppose Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold, there are constants K_1 and K_2 such that for all γ_g, $g \in \mathcal{G}$, $\epsilon > 0$, $\eta > 0$ and $\delta \geq (N_g T)^{-1}$, if $\sqrt{N_g T} \geq K_2/\eta$, then

$$
\Pr\left(\sup_{\gamma' \leq \gamma \leq \gamma' + \delta} \left|J_{g,NT}(\gamma) - J_{g,NT}(\gamma')\right| > \eta\right) \leq \frac{K_1 \delta^2}{\eta^4}.
$$

Proof of Lemma C.5. The proof is similar to that of Lemma A.3 in Hansen (2000). ■

Lemma C.6. Suppose Assumptions A.1, A.3(iii)–(iv) and A.4–A.6 hold, we have for $g \in \mathcal{G}$,

$$
J_{g,NT}(\gamma) \Rightarrow J_g(\gamma),
$$

a mean-zero Gaussian process with almost surely continuous sample paths.

Proof of Lemma C.6. The proof is similar to that of Lemma A.4 of Hansen (2000). ■

Lemma C.7. Let $G_{g,NT}(\gamma) = \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} C_g^{0\prime} x_{it} x_{it}' C_g^0 [d_{it}(\gamma) - d_{it}(\gamma_g^0)]$ and $K_{g,NT}(\gamma) = \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \|x_{it}\| \left|d_{it}(\gamma) - d_{it}(\gamma_g^0)\right|$. Under Assumptions A.1, A.3(iii)–(iv) and A.4–A.5, there exist constants $B > 0$, $0 < d < \infty$, such that for all $\eta > 0$ and $\epsilon > 0$, there exists a $\overline{v} < \infty$ such that for all (N, T) and $g \in \mathcal{G}$,

$$
\begin{aligned}
\Pr\left(\inf_{\overline{v}/\alpha_{NT} \leq \left|\gamma - \gamma_g^0\right| \leq B} \frac{G_{g,NT}(\gamma)}{\left|\gamma_g - \gamma_g^0\right|} < (1 - \eta) d\right) & \leq \epsilon, \\
\Pr\left(\sup_{\overline{v}/\alpha_{NT} \leq \left|\gamma - \gamma_g^0\right| \leq B} \frac{K_{g,NT}(\gamma)}{\left|\gamma_g - \gamma_g^0\right|} > (1 + \eta) k\right) & \leq \epsilon.
\end{aligned}
$$

Proof of Lemma C.7. The proof is similar to that of Lemma A.7 of Hansen (2000). ■

Lemma C.8: Under Assumptions A.1, A.3(iii)–(iv) and A.4–A.5, there exists some $\overline{v} < \infty$ such that for any $B < \infty$ and $g = 1, ..., G$,

$$
\Pr\left(\sup_{\overline{v}/\alpha_{NT} \leq \left|\gamma - \gamma_g^0\right| \leq B} \frac{\left|J_{g,NT}(\gamma) - J_{g,NT}(\gamma_g^0)\right|}{\sqrt{\alpha_{NT}} \left|\gamma - \gamma_g^0\right|} > \eta\right) \leq \epsilon.
$$

Proof of Lemma C.8. The proof is similar to that of Lemma A.8 of Hansen (2000). ■

Lemma C.9. Let $\tilde{K}_{g,NT}(\gamma) = N_g^{-1} \sum_{i \in \mathbf{G}_g^0} \left[T^{-1} \sum_{t=1}^{T} \|x_{it}\| \left| d_{it}(\gamma) - d_{it}(\gamma_g^0) \right| \right]^2$ and $\tilde{J}_{g,NT}(\gamma) = N_g^{-1/2} T^{-3/2} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \sum_{s=1}^{T} x_{is} e_{it} (d_{is}(\gamma) - d_{is}(\gamma_g^0))$. Suppose Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold. Then there exists some $\overline{v} < \infty$ and $B > 0$ such that for any $\eta > 0$, $\epsilon > 0$ and $g \in \mathcal{G}$,

$$\Pr\left(\sup_{\overline{v}/\alpha_{NT} \le |\gamma - \gamma_g^0| \le B} \frac{\left| \tilde{J}_{g,NT}(\gamma) \right|}{\sqrt{\alpha_{N_g T}} \left| \gamma - \gamma_g^0 \right|} > \eta \right) \le \epsilon \text{ and } \Pr\left(\sup_{\overline{v}/\alpha_{NT} \le |\gamma - \gamma_g^0| \le B} \frac{\left| \tilde{K}_{g,NT}(\gamma) \right|}{\left| \gamma - \gamma_g^0 \right|} > \eta \right) \le \epsilon.$$

Proof of Lemma C.9. The analysis for the first result is analogous to that of Lemma C.8. For the second result, we consider the case $\gamma > \gamma_g^0$. Letting $k_{it}(\gamma) = k_{it}(\gamma, \gamma_g^0) = \|x_{it}\| \left| d_{it}(\gamma) - d_{it}(\gamma_g^0) \right|$, we have

$$\begin{aligned}
E[\tilde{K}_{g,NT}(\gamma)] &= N_g^{-1} \sum_{i \in \mathbf{G}_g^0} E\left[T^{-1} \sum_{t=1}^{T} \|x_{it}\| \left| d_{it}(\gamma) - d_{it}(\gamma_g^0) \right| \right]^2 = N_g^{-1} \sum_{i \in \mathbf{G}_g^0} E\left[T^{-1} \sum_{t=1}^{T} k_{it}(\gamma) \right]^2 \\
&= N_g^{-1} \sum_{i \in \mathbf{G}_g^0} \mathrm{Var}\left[T^{-1} \sum_{t=1}^{T} k_{it}(\gamma) \right]^2 + N_g^{-1} \sum_{i \in \mathbf{G}_g^0} \left(\frac{1}{T} \sum_{t=1}^{T} E[k_{it}(\gamma)] \right)^2 \\
&\le C^\dagger N_g^{-1} \sum_{i \in \mathbf{G}_g^0} T^{-2} \sum_{t=1}^{T} E\left[k_{it}(\gamma)^2 \right] + N_g^{-1} \sum_{i \in \mathbf{G}_g^0} \left(\frac{1}{T} \sum_{t=1}^{T} E[k_{it}(\gamma)] \right)^2 \\
&\le \frac{C^\dagger C_1}{T} \left| \gamma - \gamma_g^0 \right| + C_1^2 \left| \gamma - \gamma_g^0 \right|^2,
\end{aligned}$$

where the first inequality follow from the fact that $\mathrm{Var}\left[T^{-1} \sum_{t=1}^{T} k_{it}(\gamma) \right]^2 \le C^\dagger T^{-2} \sum_{t=1}^{T} \mathrm{Var}[k_{it}(\gamma)] \le C^\dagger T^{-2} \sum_{t=1}^{T} E[k_{it}(\gamma)]^2$ for some $C^\dagger < \infty$ by using the fact that $\{k_{it}(\gamma), t \ge 1\}$ is also a strong mixing process, and the last inequality follows from Lemma C.2.

First we consider the case $\gamma - \gamma_g^0 > 0$. Choose a $b > 1$, $B < \epsilon(b-1)\eta/(4C_1^2 b^3)$ and $\overline{v}$ such that $\overline{v}/\alpha_{NT} < B$. We set $\gamma_j = \gamma_g^0 + b^{j-1}\overline{v}/\alpha_{NT}$ for $j = 1, ..., n+1$ such that $\gamma_n + 1 \ge B$ and $\gamma_n \le B$. Since $\frac{b^n \overline{v}}{\alpha_{NT}} \le B$, $n \le \log_b(B\alpha_{NT}/\overline{v})$. When (N,T) is large enough,, we can have $\frac{C^\dagger C_1 b}{\eta} \frac{n}{T} \le \epsilon/4$. Then we can calculate

$$\begin{aligned}
\Pr\left(\sup_{1 \le j \le n} \frac{\tilde{K}_{g,NT}(\gamma_{j+1})}{\left| \gamma_j - \gamma_g^0 \right|} > \eta \right) &\le \sum_{j=1}^{n} \frac{E[\tilde{K}_{g,NT}(\gamma_{j+1})]}{\eta \left| \gamma_j - \gamma_g^0 \right|} \\
&\le \sum_{j=1}^{n} \frac{C^\dagger C_1 \left| \gamma_{j+1} - \gamma_g^0 \right| / T}{\eta \left| \gamma_j - \gamma_g^0 \right|} + \sum_{j=1}^{n} \frac{C_1^2 \left| \gamma_{j+1} - \gamma_g^0 \right|^2}{\eta \left| \gamma_j - \gamma_g^0 \right|} \\
&= \frac{C^\dagger C_1 b}{\eta} \frac{n}{T} + \frac{C_1^2 b^2}{\eta} \frac{\overline{v}(b^{n+1} - 1)}{\alpha_{NT}(b-1)} \\
&\le \frac{C^\dagger C_1 b}{\eta} \frac{n}{T} + \frac{C_1^2 b^3}{\eta} \frac{B}{(b-1)} < \epsilon/2.
\end{aligned}$$

For any $\gamma \in [\gamma_g^0 + \overline{v}/\alpha_{NT}, \gamma_g^0 + B]$, there exists a $j \in \{1, ..., n\}$ such that $\gamma_j \le \gamma \le \gamma_{j+1}$. In view of the fact that $\tilde{K}_{g,NT}(\gamma)$ is monotonic in γ, we have $\frac{\tilde{K}_{g,NT}(\gamma)}{|\gamma - \gamma_g^0|} \le \frac{\tilde{K}_{g,NT}(\gamma_{j+1})}{|\gamma_j - \gamma_g^0|}$. It follows that

$$\Pr\left(\sup_{\overline{v}/\alpha_{NT} \le \gamma_g - \gamma_g^0 \le B} \frac{\left| \tilde{K}_{g,NT}(\gamma) \right|}{\left| \gamma - \gamma_g^0 \right|} > \eta \right) \le \Pr\left(\sup_{1 \le j \le n} \frac{\tilde{K}_{g,NT}(\gamma_{j+1})}{\left| \gamma_j - \gamma_g^0 \right|} > \eta \right) \le \epsilon/2.$$

A symmetric argument gives us the proof for the case $-B \leq \gamma-\gamma_g^0 \leq -\overline{v}/\alpha_{NT}$. This completes our proof. ■

Lemma C.10. Suppose that Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold. Then we have $\alpha_{NT}(\check{\gamma}_g - \gamma_g^0) = O_p(1)$ for all $g \in \mathcal{G}$.

Proof of Lemma C.10: Let B, d, k be the coefficients defined in Lemma C.6-C.8 and $c = \left\|C_g^0\right\|$. Pick an η such that $\min\{1, c, k\} > \eta > 0$ and $\pi_g^{2\alpha}(1-\eta)d - 24\pi_g^{\alpha} ck\eta - \pi_g^{2\alpha}(6ck+4c^2)\eta > 0$. Let $\mathbf{E}_{NT}$ be the joint event that, for all $g \in \mathcal{G}$: $\left|\check{\gamma}_g - \gamma_g^0\right| \leq B$, $(NT)^{\alpha}\left\|\check{\beta}_g - \beta_g^0\right\| \leq \eta$, $(NT)^{\alpha}\left\|\check{\delta}_g - \delta_g^0\right\| \leq \eta$,

$$
\begin{aligned}
\inf_{\overline{v}/\alpha_{NT} \leq \left|\gamma-\gamma_g^0\right| \leq B} \frac{G_{g,NT}(\gamma)}{\left|\gamma-\gamma_g^0\right|} &\geq (1-\eta)d, \\
\sup_{\overline{v}/\alpha_{NT} \leq \left|\gamma-\gamma_g^0\right| \leq B} \frac{K_{g,NT}(\gamma)}{\left|\gamma-\gamma_g^0\right|} &\leq (1+\eta)k, \\
\sup_{\overline{v}/\alpha_{NT} \leq \left|\gamma-\gamma_g^0\right| \leq B} \frac{\left|J_{g,NT}(\gamma) - J_{g,NT}(\gamma_g^0)\right|}{\sqrt{\alpha_{N_gT}}\left|\gamma-\gamma_g^0\right|} &\leq \eta, \\
\sup_{\overline{v}/\alpha_{NT} \leq \left|\gamma-\gamma_g^0\right| \leq B} \frac{\left|\tilde{J}_{g,NT}(\gamma)\right|}{\sqrt{\alpha_{N_gT}}\left|\gamma-\gamma_g^0\right|} &\leq \eta, \\
\sup_{\overline{v}/\alpha_{NT} \leq \left|\gamma-\gamma_g^0\right| \leq B} \frac{\tilde{K}_{g,NT}(\gamma)}{\left|\gamma-\gamma_g^0\right|} &\leq \eta.
\end{aligned}
$$

Then by Lemma C.7-C.9.Let $X_i(\gamma, \gamma_g^0) \equiv (x_{i1}[d_{i1}(\gamma) - d_{i1}(\gamma_g^0)], ..., x_{iT}[d_{iT}(\gamma) - d_{iT}(\gamma_g^0)])'$, a $T \times K$ matrix. Let $\Delta\mathbf{X}_{g,\gamma} \equiv \mathbf{X}_g(\gamma, \gamma_g^0) \equiv \left\{X_i(\gamma, \gamma_g^0), i \in \mathbf{G}_g^0\right\}$, which is an $N_gT \times K$ matrix. Let $Z_i(\gamma) \equiv ([x_{i1}', x_{i1}' d_{i1}(\gamma)]', ..., [x_{iT}', x_{iT}' d_{iT}(\gamma)]')'$, a $T \times 2K$ matrix. Let $\mathbf{Z}_g(\gamma) \equiv \left\{Z_i(\gamma), i \in \mathbf{G}_g^0\right\}$, which is an $N_gT \times 2K$ matrix. Let $\Delta\overline{\mathbf{X}}_{g,\gamma} = (I_{N_g} \otimes \mathbb{P}_0)\Delta\mathbf{X}_{g,\gamma}$, and $\overline{\mathbf{Z}}_g(\gamma_g^0) = (I_{N_g} \otimes \mathbb{P}_0)\mathbf{Z}_g(\gamma_g^0)$ where recall that $\mathbb{P}_0 = T^{-1}\iota_T\iota_T'$. Let $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$ and $\boldsymbol{\varepsilon}_g = \left\{\varepsilon_i, i \in \mathbf{G}_g^0\right\}$, an $N_gT \times 1$ vector.

$$
\begin{aligned}
\check{Q}_{g,NT}(\theta,\gamma) - \check{Q}_{g,NT}(\theta,\gamma_g^0) &= \delta'\Delta\mathbf{X}_{g,\gamma}'(I_{N_g} \otimes \mathbb{M}_0)\Delta\mathbf{X}_{g,\gamma}\delta - 2\delta'\Delta\mathbf{X}_{g,\gamma}'(I_{N_g} \otimes \mathbb{M}_0)\mathbf{Z}_g(\gamma_g^0)(\theta-\theta_g^0) \\
&\quad +2\delta'\Delta\mathbf{X}_{g,\gamma}'(I_{N_g} \otimes \mathbb{M}_0)\boldsymbol{\varepsilon}_g \\
&= \delta_g^{0\prime}\Delta\mathbf{X}_{g,\gamma}'\Delta\mathbf{X}_{g,\gamma}^0\delta_g^0 + (\delta-\delta_g^0)'\Delta\mathbf{X}_{g,\gamma}'\Delta\mathbf{X}_{g,\gamma}(\delta+\delta_g^0) - \delta'\Delta\overline{\mathbf{X}}_{g,\gamma}\Delta\overline{\mathbf{X}}_{g,\gamma}\delta \\
&\quad -2\delta'\Delta\mathbf{X}_{g,\gamma}'\mathbf{Z}_g(\gamma_g^0)(\theta-\theta_g^0) + 2\delta'\Delta\overline{\mathbf{X}}_{g,\gamma}'\overline{\mathbf{Z}}_g(\gamma_g^0)(\theta-\theta_g^0) + 2\delta'\Delta\mathbf{X}_{g,\gamma}'\boldsymbol{\varepsilon}_g \\
&\quad -2\delta'\Delta\overline{\mathbf{X}}_{g,\gamma}'\boldsymbol{\varepsilon}_g.
\end{aligned}
$$

Let $\check{\delta}_g = (NT)^{-\alpha}C_g$ for some C_g such that $\left\|C_g - C_g^0\right\| \leq \kappa$, implied by $\mathbf{E}_{NT}$. Suppose that $\mathbf{E}_{NT}$ happens and for $\gamma \in [\gamma_g^0 + \overline{v}/\alpha_{NT}, \gamma_g^0 + B]$, we have

$$
\begin{aligned}
& (N_g T)^{2\alpha-1} \frac{\check{Q}_{g,NT}(\check{\theta}_g, \gamma) - \check{Q}_{g,NT}(\check{\theta}_g, \gamma_g^0)}{\left|\gamma - \gamma_g^0\right|} \\
= \; & \frac{\pi_g^{2\alpha} C_g^{0\prime} \Delta \mathbf{X}_{g,\gamma}' \Delta \mathbf{X}_{g,\gamma} C_g^0}{N_g T \left|\gamma - \gamma_g^0\right|} + \frac{\pi_g^{2\alpha} (C_g + C_g^0)' \Delta \mathbf{X}_{g,\gamma}' \Delta \mathbf{X}_{g,\gamma} (C_g - C_g^0)}{(N_g T)\left|\gamma - \gamma_g^0\right|} - \frac{\pi_g^{2\alpha} C_g' \Delta \overline{\mathbf{X}}_{g,\gamma} \Delta \overline{\mathbf{X}}_{g,\gamma} C_g}{(N_g T)\left|\gamma - \gamma_g^0\right|} \\
& -2 \frac{\pi_g^{\alpha} C_g' \left[\Delta \mathbf{X}_{g,\gamma}' \mathbf{Z}_g(\gamma_g^0) - \Delta \overline{\mathbf{X}}_{g,\gamma}' \overline{\mathbf{Z}}_g(\gamma_g^0)\right] (N_g T)^{\alpha} (\check{\theta}_g - \theta_g^0)}{(N_g T)\left|\gamma - \gamma_g^0\right|} + 2 \left\|\pi_g^{\alpha} C_g\right\| \frac{\Delta \mathbf{X}_{g,\gamma}' \boldsymbol{\varepsilon}_g - \Delta \overline{\mathbf{X}}_{g,\gamma}' \boldsymbol{\varepsilon}_g}{(N_g T)^{1-\alpha}\left|\gamma - \gamma_g^0\right|} \\
\geq \; & \frac{\pi_g^{2\alpha} G_{g,NT}(\gamma)}{\left|\gamma - \gamma_g^0\right|} - \pi_g^{2\alpha}(\left\|C_g^0\right\| + \left\|C_g^0\right\|) \left\|C_g - C_g^0\right\| \frac{K_{g,NT}(\gamma)}{\left|\gamma - \gamma_g^0\right|} - \pi_g^{2\alpha} \left\|C_g\right\|^2 \frac{\tilde{K}_{g,NT}(\gamma)}{\left|\gamma - \gamma_g^0\right|} \\
& -4\pi_g^{\alpha} \left\|C_g\right\| (N_g T)^{\alpha} \left\|\check{\theta}_g - \theta_g^0\right\| \frac{K_{g,NT}(\gamma) + \tilde{K}_{g,NT}(\gamma)}{\left|\gamma - \gamma_g^0\right|} - 2 \left\|\pi_g^{\alpha} C_g\right\| \frac{\left|J_{g,NT}(\gamma) - J_{g,NT}(\gamma_g^0)\right|}{\sqrt{\alpha_{N_g T}}\left|\gamma - \gamma_g^0\right|} \\
& -2 \left\|\pi_g^{\alpha} C_g\right\| \frac{\left|\tilde{J}_{g,NT}(\gamma)\right|}{\sqrt{\alpha_{NT}}\left|\gamma - \gamma_g^0\right|}
\end{aligned}
$$

$$
\begin{aligned}
\geq \; & \pi_g^{2\alpha}(1-\eta)d - \pi_g^{2\alpha}(2c+\eta)\eta(1+\eta)k - \pi_g^{2\alpha}(c+\eta)^2\eta - 4\pi_g^{\alpha}(c+\eta)\eta[(1+\eta)k + \eta] \\
& -4\pi_g^{\alpha}(c+\eta)\eta \\
> \; & \pi_g^{2\alpha}(1-\eta)d - 24\pi_g^{\alpha} ck\eta - \pi_g^{2\alpha}(6ck + 4c^2)\eta \\
> \; & 0,
\end{aligned}
$$

which indicates that $\check{\gamma}_g$ does not belong to $[\gamma_g^0 + \overline{v}/\alpha_{NT}, \gamma_g^0 + B]$. A symmetric argument shows that if $\mathbf{E}_{NT}$ happens $\check{\gamma}_g$ does not belong to $[\gamma_g^0 - B, \gamma_g^0 - \overline{v}/\alpha_{NT},]$. Hence, we have shown $\check{\gamma}_g - \gamma_g^0 = O_p(1/\alpha_{NT})$ for all $g \in \mathcal{G}$. ■

Lemma C.11. Let $G_{g,NT}^*(v) = \alpha_{N_g T} G_{g,NT}(\gamma_g^0 + v/\alpha_{N_g T})$ and $K_{NT}^*(v) = \alpha_{N_g T} K_{g,NT}(\gamma_g^0 + v/\alpha_{N_g T})$. Suppose that Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold. Then we have that uniformly in $v \in \Psi$,

$$
G_{g,NT}^*(v) \xrightarrow{p} w_{g,D} |v|, \text{ and } K_{g,NT}^*(v) \xrightarrow{p} D_g^0 |v|
$$

where $w_{g,D} = C_g^{0\prime} D_g^0 C_g^0$ for $g \in \mathcal{G}$ and Ψ is a compact set.

Proof of Lemma C.11. The proof is similar to that of Lemma A.10 in Hansen (2000). ■

Lemma C.12. Let $R_{g,NT}(v) = \sqrt{\alpha_{N_g T}}\left[J_{g,NT}(\gamma_g^0 + v/\alpha_{N_g T}) - J_{g,NT}(\gamma_g^0)\right]$. Suppose that Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold. Then on any compact set Ψ,

$$
R_{g,NT}(v) \Rightarrow B_g(v)
$$

where $B_g(v)$ is a vector Brownian motion with covariance matrix $E\left[B_g(1)B_g(1)'\right] = V_g^0$.

Proof of Lemma C.12. First, we show the convergence of finite dimensional distribution: $R_{g,NT}(v) \xrightarrow{d} N(0, V_g^0)$. Let $u_{ni}(v) \equiv \frac{1}{\sqrt{N_g T}} \sum_{t=1}^{T} x_{it}\varepsilon_{it}\sqrt{\alpha_{N_g T}}[d_{it}(\gamma_g^0 + v/\alpha_{N_g T}) - d_{it}(\gamma_g^0)]$ and $\mathcal{F}_i = \sigma(\{u_{nj}(v),\, j \leq i\})$. By Assumption A.1(ii) and Liapunov's central limit theorem (e..g., Theorem 23.11 of Davidson (1994, pp.372-373), it suffices to verify that

$$
\sum_{i \in \mathbf{G}_g^0} u_{ni}(v) u_{ni}(v)' \xrightarrow{p} |v|\, V_g^0 \text{ and } \sum_{i \in \mathbf{G}_g^0} \left\|u_{ni}(v)\right\|^4 = o_p(1).
$$

Note that

$$
\begin{aligned}
\sum_{i\in\mathbf{G}_g^0} u_{ni}(v)u_{ni}(v)' &= \frac{\alpha_{N_gT}}{N_gT}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T} x_{it}x_{it}'\varepsilon_{it}^2\left|d_{it}(\gamma_g^0+v/\alpha_{N_gT})-d_{it}(\gamma_g^0)\right| \\
&\quad +\frac{\alpha_{N_gT}}{N_gT}\sum_{i\in\mathbf{G}_g^0}\sum_{1\le s\neq t\le T} x_{it}\varepsilon_{it}x_{is}'\varepsilon_{is}[d_{it}(\gamma_g^0+v/\alpha_{N_gT})-d_{it}(\gamma_g^0)][d_{is}(\gamma_g^0+v/\alpha_{N_gT})-d_{is}(\gamma_g^0)] \\
&\equiv A_{g,NT}+B_{g,NT}.
\end{aligned}
$$

For $A_{g,NT}$, we can conduct similar calculations as used in the proof of Lemma C.3 to obtain

$$
\frac{E\left[x_{it}x_{it}'\varepsilon_{it}^2\left|d_{it}(\gamma_g^0+v/\alpha_{N_gT})-d_{it}(\gamma_g^0)\right|\right]}{v/\alpha_{N_gT}} \to E(x_{it}x_{it}'\varepsilon_{it}^2|q_{it}=\gamma_g^0).
$$

Then we can readily show $A_{g,NT}\to|v|\,V_g^0$ by using the Chebyshev inequality and the fact that $\{(x_{it},q_{it},\varepsilon_{it})\}$ is independent across i and strong mixing along the time dimension. Let $\zeta_{it}=x_{it}\varepsilon_{it}[d_{it}(\gamma_g^0+v/\alpha_{N_gT})-d_{it}(\gamma_g^0)]$. For $B_{g,NT}$, we have for any $K\times 1$ nonrandom vector c with $\|c\|=1$, we have

$$
\begin{aligned}
|E[c'B_{g,NT}c]| &= \frac{\alpha_{N_gT}}{N_gT}\left|\sum_{i\in\mathbf{G}_g^0}\sum_{1\le s\neq t\le T}\mathrm{Cov}(c'\zeta_{it},c'\zeta_{is})\right| \le \frac{\alpha_{N_gT}}{N_gT}\sum_{i\in\mathbf{G}_g^0}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T_0}|\mathrm{Cov}(c'\zeta_{it},c'\zeta_{is})| \\
&= \frac{\alpha_{N_gT}}{N_gT}\sum_{i\in\mathbf{G}_g^0}\sum_{0<|s-t|\le T_0}|\mathrm{Cov}(c'\zeta_{it},c'\zeta_{is})|+\frac{\alpha_{N_gT}}{N_gT}\sum_{i\in\mathbf{G}_g^0}\sum_{|s-t|>T_0}|\mathrm{Cov}(c'\zeta_{it},c'\zeta_{is})| \\
&\le 2T_0\alpha_{N_gT}\max_i\max_{0<|s-t|\le T_0}|\mathrm{Cov}(c_1'\zeta_{it},c_2'\zeta_{is})|+\frac{T\alpha_{N_gT}}{N_g}\sum_{i\in\mathbf{G}_g^0}\{\alpha_i[T_0]\}^{(3+\epsilon_0)/(4+\epsilon_0)}\max_{i,t}\|\zeta_{it}\|_{8+\epsilon_0}^2 \\
&\le T_0\alpha_{N_gT}O((\alpha_{N_gT})^{-2})+CT\alpha_{N_gT}\rho^{T_0(3+\epsilon_0)/(4+\epsilon_0)}=o(1)
\end{aligned}
$$

provided T_0 is chosen such that $T_0=o(\alpha_{N_gT})$ and $T_0/(\ln T)^{c_0}\to\infty$ for some constant $c_0>1$. This implies that $E[B_{g,NT}]=o(1)$. In addition, it is easy to verify that $\mathrm{Var}[c'B_{g,NT}]=o(1)$. Then we have $B_{g,NT}=o_p(1)$. Consequently, $\sum_{i\in\mathbf{G}_g^0}u_{ni}(v)u_{ni}(v)'\xrightarrow{p}|v|\,V_g^0$.

Now, we verify that $\sum_{i\in\mathbf{G}_g^0}\|u_{ni}(v)\|^4=o_p(1)$. Note that

$$
\begin{aligned}
\sum_{i\in\mathbf{G}_g^0}E[c'u_{ni}(v)]^4 &= \frac{\alpha_{N_gT}^2}{(N_gT)^2}\sum_{i\in\mathbf{G}_g^0}E\left|c'\sum_{t=1}^{T}\zeta_{it}\right|^4 \\
&= \frac{\alpha_{N_gT}^2}{(N_gT)^2}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T}E(c'\zeta_{it})^4+o(1) \\
&= O(\alpha_{N_gT}(N_gT)^{-1})+o(1)=o(1),
\end{aligned}
$$

where the second equality follows from the simple application of the Davydov inequality for strong mixing processes and similar arguments as used in the analysis of $B_{g,NT}$. Then $\sum_{i\in\mathbf{G}_g^0}\|u_{ni}(v)\|^4=o_p(1)$ by Markov inequality. Then the pointwise distributional result follows.

For the stochastic equicontinuity, the proof procedure is similar to that in Hansen (2000) and thus omitted. ■

Lemma C.13. Let $\tilde{K}_{g,NT}^*(v)=\alpha_{N_gT}\tilde{K}_{g,NT}(\gamma_g^0+v/\alpha_{N_gT})$ and $\tilde{J}_{g,NT}^*(v)=\alpha_{N_gT}\tilde{J}_{g,NT}(\gamma_g^0+v/\alpha_{N_gT})$. Suppose that Assumptions A.1, A.3(iii)-(iv) and A.4–A.5 hold. Then $\tilde{K}_{g,NT}^*(v)\xrightarrow{p}0$ and $\tilde{J}_{g,NT}^*(v)\xrightarrow{p}0$ uniformly in $v\in\Psi$, where Ψ is a compact set.

Proof of Lemma C.13. By the proof of Lemma C.9, we have

$$E[\tilde{K}^*_{g,NT}(v)] = \alpha_{N_gT} O\left(\frac{1}{T}\left|v/\alpha_{N_gT}\right| + \left|v/\alpha_{N_gT}\right|^2\right) = o(1).$$

Let $\kappa_i(v) = T^{-1}\sum_{t=1}^T \kappa_{it}(v)$, where $\kappa_{it}(v) = \|x_{it}\| \left|d_{it}(\gamma_g^0 + v/\alpha_{N_gT}) - d_{it}(\gamma_g^0)\right|$. Let $\tilde{\kappa}_{it}(v) = \kappa_{it}(v) - E[\kappa_{it}(v)]$. Then

$$\begin{aligned}
\text{Var}(\tilde{K}^*_{g,NT}(v)) &= \alpha_{N_gT}^2 E\left[\left|N_g^{-1}\sum_{i\in G_g^0}\{\kappa_i(v)^2 - E[\kappa_i(v)^2]\}\right|^2\right] = \frac{\alpha_{N_gT}^2}{N_g^2}\sum_{i\in G_g^0} E\{\kappa_i(v)^2 - E[\kappa_i(v)^2]\}^2 \\
&\le \frac{\alpha_{N_gT}^2}{N_g^2}\sum_{i\in \mathbf{G}_g^0} E\left[T^{-1}\sum_{t=1}^T \kappa_{it}(v)\right]^4 \\
&\le \frac{8\alpha_{N_gT}^2}{N_g^2}\sum_{i\in \mathbf{G}_g^0} E\left[T^{-1}\sum_{t=1}^T \tilde{\kappa}_{it}(v)\right]^4 + \frac{8\alpha_{N_gT}^2}{N_g^2}\sum_{i\in G_g^0} E\left[T^{-1}\sum_{t=1}^T E\left[\kappa_{it}(v)\right]\right]^4 \\
&\le \frac{C\alpha_{N_gT}^2}{N_g^2T^4}\sum_{i\in \mathbf{G}_g^0}\left\{\sum_{t=1}^T E\left[\tilde{\kappa}_{it}(v)\right]^4 + \left(\sum_{t=1}^T E\left[\tilde{\kappa}_{it}(v)\right]^2\right)^2\right\} + o(1) + O(N_g^{-1}\alpha_{N_gT}^{-2}) \\
&= O\left(\alpha_{N_gT}N_g^{-1}T^{-3} + N_g^{-1}T^{-2}\right) + o(1) = o(1),
\end{aligned}$$

where the first equality follows from the Jensen inequality, the second inequality follows from the C_r inequality, the third one follows from the repeated application of Davydov inequality and the fact that $\max_{i,t} E\left[\kappa_{it}(v)\right] = O(\alpha_{N_gT}^{-1})$, and the next to last equality holds by the moment calculations. Then $\tilde{K}^*_{g,NT}(v) = o_p(1)$ for each $v \in \Psi$. This result, in conjunction with the monotonicity of $\tilde{K}^*_{g,NT}(v)$ in either the half line $[0,\infty)$ or the half line $(-\infty, 0]$, implies that $\tilde{K}^*_{g,NT}(v) \xrightarrow{p} 0$ uniformly in $v \in \Psi$. See Hansen (2000, p. 598).

For $\tilde{J}^*_{g,NT}(v)$, we can follow the above arguments and show that $\tilde{J}^*_{g,NT}(v) = o_p(1)$ for each $v \in \Psi$. Following Lemma A.11 in Hansen (2000), we can readily show the tightness of the process $\{\tilde{J}^*_{g,NT}(v)\}$. As a result, we have $\tilde{J}^*_{g,NT}(v) \xrightarrow{p} 0$ uniformly in $v \in \Psi$. ■

Lemma C.14. Suppose that Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold. Then on any compact set Ψ,

$$Q^*_{g,NT}(v) \Rightarrow -\pi_g^{2\alpha} w_{g,D}|v| + 2\sqrt{\pi_g^{2\alpha} w_{g,V}}W_g(v) = \frac{w_{g,V}}{w_{g,D}}\left(-\frac{\pi_g^{2\alpha}w_{g,D}^2}{w_{g,V}}|v| + 2W_g(\frac{\pi_g^{2\alpha}w_{g,D}^2}{w_{g,V}}v)\right),$$

where $w_{g,V} = C_g^{0\prime}V_g^0C_g^0$.

Proof of Lemma C.14. Let $X_i(\gamma_g^0 + v/\alpha_{N_gT}, \gamma_g^0) = [x_{i1}[d_{i1}(\gamma_g^0 + v/\alpha_{N_gT}) - d_{i1}(\gamma_g^0)], ..., x_{iT}[d_{iT}(\gamma_g^0 + v/\alpha_{N_gT}) - d_{iT}(\gamma_g^0)]]'$. We have

$$\begin{aligned}
Q^*_{g,NT}(v) &= \check{Q}_g(\check{\theta}_g, \gamma_g^0) - \check{Q}_g(\check{\theta}_g, \gamma_g^0 + v/\alpha_{N_gT}) \\
&= -\sum_{i\in\mathbf{G}_g^0}\delta_g^{0\prime}X_i(\gamma_g^0 + v/\alpha_{N_gT}, \gamma_g^0)'X_i(\gamma_g^0 + v/\alpha_{N_gT}, \gamma_g^0)\delta_g^0 + 2\sum_{i\in\mathbf{G}_g^0}\delta_g^{0\prime}X_i(\gamma_g^0 + v/\alpha_{N_gT}, \gamma_g^0)'\varepsilon_i \\
&\quad + L_{g,NT}(v),
\end{aligned}$$

where

$$\begin{aligned}
L_{g,NT}(v) &= 2(N_gT)^\alpha(\check{\delta}_g-\delta_g^0)'R_{g,NT}^*(v)-2(N_gT)^\alpha\check{\delta}_g'K_{g,NT}^*(v)(N_gT)^\alpha(\check{\beta}_g-\beta_g^0) \\
&\quad -(N_gT)^\alpha(\check{\delta}_g-\delta_g^0)'K_{g,NT}^*(v)(N_gT)^\alpha(\check{\delta}_g+\delta_g^0)+(N_gT)^\alpha\check{\delta}_g'\tilde{K}_{g,NT}^*(v)(N_gT)^\alpha\check{\delta}_g \\
&\quad +2(N_gT)^\alpha\check{\delta}_g'\tilde{J}_{g,NT}^*(v)-2\check{\delta}_g'\sum_{i\in G_g^0}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'\mathbb{M}_0Z_i(\gamma_g^0)(\check{\theta}_g-\theta_g^0) \\
&\equiv L_{1g,NT}(v)+...+L_{6g,NT}(v).
\end{aligned}$$

By Lemma C.10, we have

$$\begin{aligned}
&\sum_{i\in\mathbf{G}_g^0}\delta_g^{0\prime}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)\delta_g^0 \\
=\;& (NT)^{-2\alpha}\sum_{i\in\mathbf{G}_g^0}C_g^{0\prime}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)C_g^0 \\
=\;& (\frac{N_g}{N})^{2\alpha}\frac{\alpha_{N_gT}}{N_gT}\sum_{i\in\mathbf{G}_g^0}C_g^{0\prime}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)C_g^0 \\
=\;& (\frac{N_g}{N})^{2\alpha}G_{g,NT}^*(v)\Rightarrow\pi_g^{2\alpha}w_{g,D}\left|v\right|.
\end{aligned}$$

By Lemma C.11, we have

$$\begin{aligned}
\sum_{i\in\mathbf{G}_g^0}\delta_g^{0\prime}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'\varepsilon_i &= (NT)^{-\alpha}\sum_{i\in\mathbf{G}_g^0}C_g^{0\prime}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'\varepsilon_i \\
&= (\frac{N_g}{N})^{\alpha}(N_gT)^{1/2-\alpha}C_g^{0\prime}\left[J_{g,NT}(\gamma_g^0+v/\alpha_{N_gT})-J_{g,NT}(\gamma_g^0)\right] \\
&= (\frac{N_g}{N})^{\alpha}R_{g,NT}(v) \\
&\Rightarrow \pi_gC_g^{0\prime}B_g(v)=\pi_g\sqrt{w_{g,V}}W_g(v).
\end{aligned}$$

By the fact that $(NT)^\alpha(\check{\theta}_g-\theta_g^0)=o_p(1)$, Assumption A.1(vi), and Lemma C.10, we have $L_{\ell g,NT}(v)=o_p(1)$ uniformly in v for $\ell=1,2,3,4$. By Lemma C.12 we have that $L_{5g,NT}(v)=o_p(1)$ uniformly in v. For $L_{6g,NT}(v)$, we have

$$\begin{aligned}
\left|L_{6g,NT}(v)\right| &\leq 2\left\{(N_gT)^\alpha\left\|\check{\delta}_g\right\|\right\}\left\{(N_gT)^\alpha\left\|\check{\theta}_g-\theta_g^0\right\|\right\}\frac{\alpha_{N_gT}}{N_gT}\left\|\sum_{i\in G_g^0}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'\mathbb{M}_0Z_i(\gamma_g^0)\right\| \\
&= O_p(1)\,o_p(1)\,O_p(1)=o_p(1)\text{ uniformly in }v\in\Psi
\end{aligned}$$

as we can follow the proofs of Lemmas C.10 and C.12 and show that $\frac{\alpha_{N_gT}}{N_gT}||\sum_{i\in G_g^0}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'\mathbb{M}_0Z_i(\gamma_g^0)||$ $=O_p(1)$ uniformly in $v\in\Psi$. Consequently, we have $Q_{g,NT}^*(v)\Rightarrow-\pi_g^{2\alpha}w_{g,D}\left|v\right|+2\sqrt{\pi_g^{2\alpha}w_{g,V}}W_g(v)$ on any compact set Ψ. ■

D Determination of the Number of Groups

Recall that $\hat{\sigma}^2(G)\equiv\frac{1}{NT}\mathcal{Q}(\hat{\Theta}^{(G)},\hat{\mathbf{D}}^{(G)},\hat{\mathbf{G}}^{(G)})$. Let $\bar{\sigma}_{NT}^2\equiv\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\varepsilon_{it}^2$. In the estimation, we require each group to contain at least $\lfloor\nu N\rfloor$ individuals. We denote the index set of members in group g as $\mathbf{G}_g$, where $\mathbf{G}_g \in \mathbb{G}_\nu = \{\mathbf{G}_{\tilde{g}}, |\mathbf{G}_{\tilde{g}}| > \lfloor \nu N \rfloor\}$ for all $g \in \mathcal{G}$. Let $\hat{N}_g = |\mathbf{G}_g|$. We can define five empirical processes that depend on $\mathbf{G}_g$:

$$
\begin{aligned}
J(\mathbf{G}_g, \gamma) &= \frac{1}{\hat{N}_g T} \sum_{i \in \mathbf{G}_g} Z_i(\gamma)' \mathbb{M}_0 \varepsilon_i, \quad \Delta J(\mathbf{G}_g, \gamma, \gamma^*) = \frac{1}{\hat{N}_g T} \sum_{i \in \mathbf{G}_g} X_i(\gamma, \gamma^*)' \mathbb{M}_0 \varepsilon_i, \\
\Phi_1(\mathbf{G}_g, \gamma) &= \frac{1}{\hat{N}_g T} \sum_{i \in \mathbf{G}_g} Z_i(\gamma)' \mathbb{M}_0 Z_i(\gamma), \; \Phi_2(\mathbf{G}_g, \gamma, \gamma^*) = \frac{1}{\hat{N}_g T} \sum_{i \in \mathbf{G}_g} Z_i(\gamma)' \mathbb{M}_0 X_i(\gamma, \gamma^*), \text{ and} \\
\Phi_3(\mathbf{G}_g, \gamma, \gamma^*) &= \frac{1}{\hat{N}_g T} \sum_{i \in \mathbf{G}_g} X_i(\gamma, \gamma^*)' \mathbb{M}_0 X_i(\gamma, \gamma^*).
\end{aligned}
$$

Let $\mathbf{G}^G$ be any possible group structure when the number of groups in $\{1, 2, ..., N\}$ is given by G. We assume the following conditions hold for the empirical processes.

Assumption D.1. (i) $\Pr\left(\inf_{(\mathbf{G}_g, \gamma) \in \mathbb{G}_\nu \times \Gamma} \lambda_{\min}\left[\Phi_1(\mathbf{G}_g, \gamma)\right] \geq c\right) \to 1$ as $(N, T) \to \infty$ for some $c > 0$;

(ii) $\Pr\left(\inf_{\mathbf{G} \in \mathbb{G}_\nu} \inf_{|\gamma - \gamma^*| > \bar{v}/T} \lambda_{\min}\left[\Phi_3(\mathbf{G}_g, \gamma, \gamma^*) - \Phi_2(\mathbf{G}_g, \gamma, \gamma^*)' \Phi_1(\mathbf{G}_g, \gamma)^{-1} \Phi_2(\mathbf{G}_g, \gamma, \gamma^*)\right] / |\gamma - \gamma^*| \geq c\right)$ $\to 1$ as $(N, T) \to \infty$ for some $c > 0$ and $\bar{v} > 0$;

(iii) $\Pr\left(\sup_{\mathbf{G} \in \mathbb{G}_\nu} \sup_{|\gamma - \gamma^*| > \bar{v}/T} \|\Phi_\ell(\mathbf{G}_g, \gamma, \gamma^*)\| / |\gamma - \gamma^*| \leq C\right) \to 1$ for $\ell = 2, 3$ as $(N, T) \to \infty$ for some $C > 0$;

(iv) $\Pr\left(\sup_{(\mathbf{G}_g, \gamma) \in \mathbb{G}_\nu \times \Gamma} \|J(\mathbf{G}_g, \gamma)\| \leq C T^{-1/2}\right) \to 1$ for some $C > 0$;

(v) $\Pr\left(\sup_{\mathbf{G}_g \in \mathbb{G}_\nu} \sup_{|\gamma - \gamma^*| > \bar{v}/T} \|\Delta J(\mathbf{G}_g, \gamma, \gamma^*)\| / |\gamma - \gamma^*| \leq C T^{-1/2}\right) \to 1$ for some $C > 0$ and $\bar{v} > 0$.

Assumption D.2. (i) As $(N, T) \to \infty$, $\min_{1 \leq G < G^0} \min_{\mathbf{G}^G} \hat{\sigma}^2_{\mathbf{G}^G} \xrightarrow{P} \bar{\sigma}^2 > \sigma^2$, where $\sigma^2 \equiv \lim_{(N,T) \to \infty} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T E\left(\varepsilon_{it}^2\right)$.

(ii) $\lambda_{NT} \to 0$ and $T \lambda_{NT} \to \infty$ as $(N, T) \to \infty$.

Assumption D.1(i)-(iii) requires the sample covariance matrices are well behaved for any subset of individuals. Assumption D.1(iv) is the assumption that plays the most important role in our analysis. It requires $\sup_{(\mathbf{G}_g, \gamma) \in \mathbb{G}_\nu \times \Gamma} \|J(\mathbf{G}_g, \gamma)\| = O_p(T^{-1/2})$ for all $(\mathbf{G}_g, \gamma) \in \mathbb{G}_\nu \times \Gamma$. For the true group members $\mathbf{G}_g^0$, we can show that $J(\mathbf{G}_g^0, \gamma) = O_p((NT)^{-1/2})$ under some regularity conditions. However when we are estimating the model with $G > G^0$, it is possible that $\left\|J(\hat{\mathbf{G}}_g, \gamma)\right\| = O_p(T^{-1/2})$. Similar remarks hold for D.1(v). Assumption D.2 specifies the usual condition for the consistency of an information criterion. In particular, Assumption D.2(i) in conjunction with the first part of D.2(ii) helps to eliminate all underfitted models and the second part of D.2(ii) helps to eliminate the overfitted models.

Proposition D.1 *Suppose Assumptions A.1-A.5 in the text and Assumption D.1 hold. The following statement holds:*

$$\hat{\sigma}^2(G) - \bar{\sigma}^2_{NT} = O_p(T^{-1}) \text{ for any } G^0 \leq G \leq G_{\max}.$$

Remark. The probability order $O_p(T^{-1})$ in the above proposition is not a conservative order. To illustrate this point, we consider a simple regression where $y_{it} = \mu + \varepsilon_{it}$ so that there is only one group. If we estimate the model with $G=2$, we have

$$\begin{aligned}
0 \quad \geq \quad & T\left[\hat{\sigma}^2(2)-\bar{\sigma}_{NT}^2\right]=\frac{1}{N}\sum_{g=1}^{2}\sum_{i\in\hat{\mathbf{G}}_g}\sum_{t=1}^{T}\left(y_{it}-\frac{1}{\hat{N}_gT}\sum_{i\in\hat{\mathbf{G}}_g}\sum_{t=1}^{T}y_{it}\right)^2-\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2 \\
= \quad & \frac{1}{N}\sum_{g=1}^{2}\sum_{i\in\hat{\mathbf{G}}_g}\sum_{t=1}^{T}\left(\varepsilon_{it}-\frac{1}{\hat{N}_gT}\sum_{i\in\hat{\mathbf{G}}_g}\sum_{t=1}^{T}\varepsilon_{it}\right)^2-\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\varepsilon_{it}^2 \\
= \quad & -T\sum_{g=1}^{2}\frac{\hat{N}_g}{N}\left(\frac{1}{\hat{N}_g}\sum_{i\in\hat{\mathbf{G}}_g}\bar{\varepsilon}_{i\cdot}\right)^2\leq-\sum_{g=1}^{2}\frac{\tilde{N}_g}{N}\left(\frac{\sqrt{T}}{\tilde{N}_g}\sum_{i\in\tilde{\mathbf{G}}_g}\bar{\varepsilon}_{i\cdot}\right)^2,
\end{aligned}$$

where $\bar{\varepsilon}_{i\cdot}=\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}$, $\tilde{\mathbf{G}}_1=\{i|\bar{\varepsilon}_{i\cdot}\leq 0\}$, $\tilde{\mathbf{G}}_2=\{i|\bar{\varepsilon}_{i\cdot}>0\}$, $\tilde{N}_g=\left|\tilde{\mathbf{G}}_g\right|$, and the last inequality holds by the definitions of $\{\hat{\mathbf{G}}_g\}$ and $\hat{\sigma}^2(2)$. [Note that $\hat{\sigma}^2(2)$ is minimized at $(\hat{\mathbf{G}}_1,\hat{\mathbf{G}}_2)$.] Without loss of generality, we suppose that ε_{it} is i.i.d. $N(0,1)$ over both i and t. Then $\upsilon_{iT}\equiv\sqrt{T}\bar{\varepsilon}_{i\cdot}\backsim N(0,1)$ and by the strong law of large numbers

$$\left|\frac{\sqrt{T}}{\tilde{N}_1}\sum_{i\in\tilde{\mathbf{G}}_1}\bar{\varepsilon}_{i\cdot}\right|=\left|\frac{\sqrt{T}}{\tilde{N}_1}\sum_{i=1}^{N}\bar{\varepsilon}_{i\cdot}\mathbf{1}\left(\bar{\varepsilon}_{i\cdot}\leq 0\right)\right|=\frac{N}{\tilde{N}_1}\left|\frac{1}{N}\sum_{i=1}^{N}\upsilon_{iT}\mathbf{1}\left(\upsilon_{iT}\leq 0\right)\right|\xrightarrow{a.s.}2\left|E\left[Z\mathbf{1}\left(Z\leq 0\right)\right]\right|,$$

where $Z\backsim N(0,1)$ and we use the fact that $\tilde{N}_1/N=\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left(\bar{\varepsilon}_{i\cdot}\leq 0\right)\xrightarrow{a.s.}P\left(Z\leq 0\right)=\frac{1}{2}$. Similarly,

$$\left|\frac{\sqrt{T}}{\tilde{N}_2}\sum_{i\in\tilde{\mathbf{G}}_2}\bar{\varepsilon}_{i\cdot}\right|=\left|\frac{\sqrt{T}}{\tilde{N}_2}\sum_{i=1}^{N}\bar{\varepsilon}_{i\cdot}\mathbf{1}\left(\bar{\varepsilon}_{i\cdot}>0\right)\right|=\frac{N}{\tilde{N}_2}\left|\frac{1}{N}\sum_{i=1}^{N}\upsilon_{iT}\mathbf{1}\left(\upsilon_{iT}>0\right)\right|\xrightarrow{a.s.}2\left|E\left[Z\mathbf{1}\left(Z>0\right)\right]\right|.$$

This calculation indicates that the negative value $T\left[\hat{\sigma}^2(2)-\bar{\sigma}_{NT}^2\right]$ has the probability order $O_p\left(T^{-1}\right)$ that cannot be $o_p\left(T^{-1}\right)$. In other words, the order $O_p(T^{-1})$ is a tight probability order for $\hat{\sigma}^2(2)-\bar{\sigma}_{NT}^2$.

Proof of Proposition D.1. Following similar arguments as used in the proofs of Lemmas A.1-A.3, we can show that individuals from the true group $\mathbf{G}_g^0$ would stay in the same estimated group w.p.a.1, i.e.,

$$\Pr\left[\sup_{1\leq i<j\leq N}\mathbf{1}(\hat{g}_i=\hat{g}_j,\ g_i^0\neq g_j^0)=1\right]\to 0 \text{ as } (N,T)\to\infty.$$

We only consider the case where some true groups are further divided into several groups. For notational simplicity, we only consider the case $G^0=1$ where our true parameters can be rewritten as $(\theta^{0\prime},\gamma^0)'=(\beta^{0\prime},\delta^{0\prime},\gamma^0)'$ without the group-specific subscript. Since we still estimate a PSTR model with $G\geq 1$ groups, the estimators, e.g., $(\hat{\theta}_g^{(G)},\hat{\gamma}_g^{(G)})$, still have the group-specific subscript. But for notational brevity, we will denote $(\hat{\theta}_g^{(G)},\hat{\gamma}_g^{(G)})$ as $(\hat{\theta}_g,\hat{\gamma}_g)$. Then we can write $\mathcal{Q}(\hat{\Theta}^{(G)},\hat{\mathbf{D}}^{(G)},\hat{\mathbf{G}}^{(G)})=\sum_{g=1}^{G}\bar{\mathcal{Q}}_g(\hat{\theta}_g,\hat{\gamma}_g)$, where $\bar{\mathcal{Q}}_g(\cdot,\cdot)$ is defined in Section 4.1. Following the analysis for (C.2) in the proof of Lemma C.2, we have

$$\hat{\theta}_g-\theta^0=\bar{\Phi}_{1,g}(\hat{\gamma}_g)^{-1}\frac{1}{\hat{N}_gT}\sum_{i\in\hat{\mathbf{G}}_g}Z_i(\hat{\gamma}_g)'\mathbb{M}_0\varepsilon_i-\bar{\Phi}_{1,g}(\hat{\gamma}_g)^{-1}\bar{\Phi}_{2,g}(\hat{\gamma}_g)\delta^0, \tag{D.1}$$

where $\bar{\Phi}_{1,g}(\gamma)\equiv\frac{1}{\hat{N}_gT}\sum_{i\in\hat{\mathbf{G}}_g}Z_i(\gamma)'\mathbb{M}_0Z_i(\gamma)=\Phi_1(\hat{\mathbf{G}}_g,\gamma)$ and $\bar{\Phi}_{2,g}(\gamma)\equiv\frac{1}{\hat{N}_gT}\sum_{i\in\hat{\mathbf{G}}_g}Z_i(\gamma)'\mathbb{M}_0X_i(\gamma,\gamma^0)=$

$\Phi_2(\hat{\mathbf{G}}_g, \gamma, \gamma^0)$. Following the similar analysis for (C.3) of Lemma C.2, we have

$$\begin{aligned} & \frac{1}{\hat{N}_g T}\left[\bar{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g) - \bar{\mathcal{Q}}_g(\theta^0, \gamma^0)\right] \\ = \; & \delta^{0\prime}\bar{\Phi}_{3g}(\hat{\gamma}_g)\delta^0 + (\hat{\theta}_g - \theta^0)'\bar{\Phi}_{1g}(\hat{\gamma}_g)(\hat{\theta}_g - \theta^0) + 2(\hat{\theta}_g - \theta^0)'\bar{\Phi}_{2g}(\hat{\gamma}_g)\delta^0 \\ & -(\hat{\theta}_g - \theta^0)'\frac{2}{N_g T}\sum_{i\in\hat{\mathbf{G}}_g} Z_i(\tilde{\gamma}_g)'\mathbb{M}_0\varepsilon_i - \hat{\delta}_g'\frac{2}{N_g T}\sum_{i\in\hat{\mathbf{G}}_g} X_i(\hat{\gamma}_g, \gamma^0)'\mathbb{M}_0\varepsilon_i, \end{aligned}$$

where $\bar{\Phi}_{3g}(\gamma) \equiv \frac{1}{\hat{N}_g T}\sum_{i\in\hat{\mathbf{G}}_g} X_i(\gamma, \gamma^0)'\mathbb{M}_0 X_i(\gamma, \gamma^0) = \Phi_3(\hat{\mathbf{G}}_g, \gamma, \gamma^0)$. Plugging (D.1) into the above equation, we have

$$\begin{aligned} \frac{1}{\hat{N}_g T}\left[\bar{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g) - \bar{\mathcal{Q}}_g(\theta^0, \gamma^0)\right] \;=\; & \delta^{0\prime}\left[\bar{\Phi}_{3g}(\hat{\gamma}_g) - \bar{\Phi}_{2g}(\hat{\gamma}_g)'\bar{\Phi}_{1g}(\hat{\gamma}_g)^{-1}\bar{\Phi}_{2g}(\hat{\gamma}_g)\right]\delta^0 \\ & -J(\hat{\mathbf{G}}_g, \hat{\gamma}_g)'\bar{\Phi}_{1g}(\hat{\gamma}_g)^{-1}J(\hat{\mathbf{G}}_g, \hat{\gamma}_g) + 2\delta^{0\prime}\bar{\Phi}_{2g}(\hat{\gamma}_g)'\bar{\Phi}_{1g}(\hat{\gamma}_g)^{-1}J(\hat{\mathbf{G}}_g, \hat{\gamma}_g) \\ & -2\left(\hat{\delta}_g - \delta^0\right)'\Delta J(\hat{\mathbf{G}}_g, \hat{\gamma}_g, \gamma^0) - 2\delta^{0\prime}\Delta J(\hat{\mathbf{G}}_g, \hat{\gamma}_g, \gamma^0) \\ \equiv\; & \Delta\bar{\mathcal{Q}}_{1,g} + \ldots + \Delta\bar{\mathcal{Q}}_{5,g}. \end{aligned}$$

We discuss two cases: (1) $(NT)^{-\alpha}T^{1/2} = O(1)$ and (2) $(NT)^{-\alpha}T^{1/2} \to \infty$ as $(N,T) \to \infty$.

In Case (1), we have $\delta^0 = O(T^{-1/2})$. By Assumption D.1(iii) and equation (D.1), we can readily show that $\hat{\theta}_g - \theta^0 = O_p(T^{-1/2})$. With this result, then we can show that $\Delta\bar{\mathcal{Q}}_{l,g} = O_p(T^{-1})$ for $l = 1, ..., 5$ by using Assumption D.1.

In Case (2), we have $\hat{\theta}_g - \theta^0 = O_p(T^{-1/2} + (NT)^{-\alpha}\left|\hat{\gamma}_g - \gamma^0\right|)$ by (D.1) and Assumption D.1(iii). Then we can apply Assumption D.1 to show that

$$\begin{aligned} \Delta\bar{\mathcal{Q}}_{1,g} &= O_p((NT)^{-2\alpha}\left|\hat{\gamma}_g - \gamma^0\right|),\ \Delta\bar{\mathcal{Q}}_{2,g} = O_p(T^{-1}), \\ \Delta\bar{\mathcal{Q}}_{3,g} &= O_p(T^{-1/2}(NT)^{-\alpha}\left|\hat{\gamma}_g - \gamma^0\right|), \\ \Delta\bar{\mathcal{Q}}_{4,g} &= O_p(T^{-1/2}(NT)^{-\alpha}\left|\hat{\gamma}_g - \gamma^0\right|^2 + T^{-1}\left|\hat{\gamma}_g - \gamma^0\right|),\ \text{and} \\ \Delta\bar{\mathcal{Q}}_{5,g} &= O_p(T^{-1/2}(NT)^{-\alpha}\left|\hat{\gamma}_g - \gamma^0\right|. \end{aligned}$$

Because $\Delta\bar{\mathcal{Q}}_{1,g} > 0$ by Assumption D.1(ii) and $\frac{1}{\hat{N}_g T}\left[\bar{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g) - \bar{\mathcal{Q}}_g(\theta^0, \gamma^0)\right] < 0$ by the definition of least squares estimation, we can conclude $\Delta\bar{\mathcal{Q}}_{1,g}$ should have at most the same order as $\sum_{l=2}^5 \Delta\bar{\mathcal{Q}}_{l,g}$. By comparison between these orders, we can show that $\left|\hat{\gamma}_g - \gamma^0\right| = O_p(T^{-1}(NT)^{2\alpha})$ and $\sum_{l=1}^5 \Delta\bar{\mathcal{Q}}_{l,g} = O_p(T^{-1})$ follows. Consequently,

$$\begin{aligned} 0 \;\geq\; & \hat{\sigma}^2(G) - \bar{\sigma}^2_{NT} = \frac{1}{NT}\sum_{g=1}^G\left[\bar{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g) - \bar{\mathcal{Q}}_g(\theta^0, \gamma^0)\right] \\ =\; & \sum_{g=1}^G \frac{\hat{N}_g}{N}\frac{1}{\hat{N}_g T}\left[\bar{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g) - \bar{\mathcal{Q}}_g(\theta^0, \gamma^0)\right] \\ \geq\; & \sum_{g=1}^G \frac{1}{\hat{N}_g T}\left[\bar{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g) - \bar{\mathcal{Q}}_g(\theta^0, \gamma^0)\right] = O_p(T^{-1}). \end{aligned}$$

This implies that $\hat{\sigma}^2(G) - \bar{\sigma}^2_{NT} = O_p(T^{-1})$ for any $G^0 \leq G \leq G_{\max}$. ■

E Consistency of group membership estimators in the fixed-threshold-effect framework

In this section, we discuss the asymptotic property of our least squares estimator under the constant threshold effect framework (i.e., $\alpha = 0$). Suppose Assumptions A.1-A.5 hold except that we now let $\alpha = 0$. Then one can follow the arguments as used in the proofs of Lemmas C.2-C.10 to show that $\left|\check{\gamma}_g - \gamma_g^0\right| = O_p((NT)^{-1})$ and $\check{\theta}_g = \check{\theta}(\gamma_g^0) + o_p((NT)^{-1/2})$, where $(\check{\theta}_g, \check{\gamma}_g)$ is infeasible estimator for $g \in \mathcal{G}$.

In the PSTR model, the major difficulty is to show the consistency of the estimator of the latent group structure as in Theorem 3.1. Once we establish a similar result as that of Lemma A.3, we can prove Theorem 3.1. In addition, we can prove Lemmas A.4-A.6 which confirms $\left|\hat{\gamma}_g - \gamma_g^0\right| = O_p((NT)^{-1})$ and $\hat{\theta}_g = \check{\theta}_g + o_p((NT)^{-1/2})$. In the following analysis, we give a sketch of the proof of Theorem 3.1 in the fixed-threshold-effect framework.

To proceed, we add some notations. Define

$$\tilde{M}_{NT}(g, \tilde{g}, \mathbf{G}) \equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(g_i^0 = g)\mathbf{1}(g_i = \tilde{g}) E\left[(x_{it}'\delta_g^0)^2 | \gamma_g^0\right],$$

where $E(\cdot|\gamma) \equiv E(\cdot|q_{it} = \gamma)$. We impose an additional identification condition:

Assumption E.1. As $(N, T) \to \infty$, the following statements hold: (i) For some constants $c > 0$ and $\bar{v} > 0$, we have

$$\begin{aligned} \sup_{1 \le i \le N} \sup_{(\theta\prime, \theta^{*\prime})' \in \mathcal{B}^2} \sup_{|\gamma - \gamma^*| > \bar{v}/T} & \Bigg\{ \Pr\Bigg[\sum_{t=1}^{T} \left[\tilde{z}_{it}(\gamma)'\theta - \tilde{z}_{it}(\gamma^*)'\theta^*\right]^2 \le \\ & c \sum_{t=1}^{T} \left[(\theta - \theta^*)'\tilde{z}_{it}(\gamma^*)\right]^2 + |\gamma - \gamma^*| \, E\left[(x_{it}'\delta)^2 | \gamma\right] \Bigg] \Bigg\} \quad = \quad o(T^{-4}); \end{aligned}$$

(ii) There exists a constant $\underline{c}_{\lambda} > 0$ such that for all $g \in \mathcal{G}$,

$$\Pr\left(\inf_{(\mathbf{G}, \mathbf{D}) \in \mathcal{G}^N \times \Gamma^G} \max_{\tilde{g} \in \mathcal{G}} \{\lambda_{\min}[M_{NT}(g, \tilde{g}, \mathbf{D}, \mathbf{G})] \wedge \tilde{M}_{NT}(g, \tilde{g}, \mathbf{G})\} > \underline{c}_{\lambda} \right) \to 1;$$

(iii) For all $g, \tilde{g} \in \mathcal{G}$, where $g \neq \tilde{g}$, we have $\left\|(\theta_g^{0\prime}, \gamma_g^0)' - (\theta_{\tilde{g}}^{0\prime}, \gamma_{\tilde{g}}^0)'\right\| > \underline{c}_{\beta}$ for some constant $\underline{c}_{\beta} > 0$;
(iv) For any $g \neq \tilde{g}$ and $1 \le i \le N$, we have

$$\max\left(E[\tilde{z}_{it}(\gamma_g^0)'(\theta_{\tilde{g}}^0 - \theta_g^0)]^2, \; \left|\gamma_{\tilde{g}}^0 - \gamma_g^0\right| E\left[(x_{it}'\delta_{\tilde{g}}^0)^2 | \gamma_{\tilde{g}}^0\right] \right) \equiv \underline{\tilde{c}}_{g\tilde{g},i} \ge \underline{\tilde{c}}_{g\tilde{g}},$$

for some constant $\underline{\tilde{c}}_{g\tilde{g}} > 0$.

Assumption E.1 (i) is a non-colinearity condition similar to Assumption A.4(ii) in the main text. However, it requires that the non-colinearity property should hold for each individual. Assumption E.1(ii) is modified from Assumption A.2. Assumption E.1(iii)-(iv) is modified from Assumption A.3(i)-(ii). As remarked in Section 3.1, E.1(iv) is redundant if we assume that $\lambda_{\min}(E[\tilde{z}_{it}(\gamma_g^0)\tilde{z}_{it}(\gamma_g^0)'])$ and $\delta_g^{0\prime} E(x_{it}x_{it}' | \gamma_g^0)\delta_g^0$ are bounded below from zero by a constant $\underline{c}$, say.

Below we prove Theorem 3.1 under Assumptions A.1, A.3(iii)-(iv) and E.1.

Proof of Theorem 3.1. Lemma A.1 still holds under the stated conditions. Lemmas A.2-A.3 are replaced by Lemmas E.1 and E.2 below. Combining Lemmas E.1-E.2 we have the desired claim. ■

Lemma E.1. Suppose that Assumptions A.1, A.3(iii)-(iv) and E.1 hold. Then we have $d_H((\hat{\Theta}, \hat{\mathbf{D}}), (\Theta^0, \mathbf{D}^0)) \xrightarrow{p} 0$, where

$$d_H((\hat{\Theta}, \hat{\mathbf{D}}), (\Theta^0, \mathbf{D}^0)) = \max\left\{\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\hat{\theta}_g - \theta^0_{\tilde{g}}\right\|^2 + \left|\hat{\gamma}_g - \gamma^0_{\tilde{g}}\right|\right),\ \max_{\tilde{g}\in\mathcal{G}}\left(\min_{g\in\mathcal{G}}\left\|\hat{\theta}_g - \theta^0_{\tilde{g}}\right\|^2 + \left|\hat{\gamma}_g - \gamma^0_{\tilde{g}}\right|\right)\right\}.$$

Proof of Lemma E.1. It suffices to show (i) $\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\hat{\theta}_g - \theta^0_{\tilde{g}}\right\| + \left|\hat{\gamma}_g - \gamma^0_{\tilde{g}}\right|\right) = o_p(1)$ and (ii) $\max_{\tilde{g}\in\mathcal{G}}\left(\min_{g\in\mathcal{G}}\left\|\hat{\theta}_g - \theta^0_{\tilde{g}}\right\| + \left|\hat{\gamma}_g - \gamma^0_{\tilde{g}}\right|\right) = o_p(1)$.

We first show (i). By Lemma A.1, we have

$$\begin{aligned}\frac{1}{NT}\tilde{\mathcal{Q}}(\hat{\Theta}, \hat{\mathbf{D}}, \hat{\mathbf{G}}) &= \frac{1}{NT}\mathcal{Q}(\hat{\Theta}, \hat{\mathbf{D}}, \hat{\mathbf{G}}) + o_p(1) \le \frac{1}{NT}\mathcal{Q}(\Theta^0, \mathbf{D}^0, \mathbf{G}^0) + o_p(1)\\ &= \frac{1}{NT}\tilde{\mathcal{Q}}(\Theta^0, \mathbf{D}^0, \mathbf{G}^0) + o_p(1),\end{aligned}$$

where the inequality holds by the definition of least squares estimator. On the other hand, noting that $\tilde{\mathcal{Q}}(\Theta, \mathbf{D}, \mathbf{G})$ is minimized at $(\Theta^0, \mathbf{D}^0, \mathbf{G}^0)$, we have $\frac{1}{NT}[\tilde{\mathcal{Q}}(\hat{\Theta}, \hat{\mathbf{D}}, \hat{\mathbf{G}}) - \tilde{\mathcal{Q}}(\Theta^0, \mathbf{D}^0, \mathbf{G}^0)] \ge 0$. It follows that $\frac{1}{NT}[\tilde{\mathcal{Q}}(\hat{\Theta}, \hat{\mathbf{D}}, \hat{\mathbf{G}}) - \tilde{\mathcal{Q}}(\Theta^0, \mathbf{D}^0, \mathbf{G}^0)] = o_p(1)$. By direct calculation, we have uniformly in $(\Theta, \mathbf{D}, \mathbf{G})$,

$$\begin{aligned}&\frac{1}{NT}\left[\tilde{\mathcal{Q}}(\Theta, \mathbf{D}, \mathbf{G}) - \tilde{\mathcal{Q}}(\Theta^0, \mathbf{D}^0, \mathbf{G}^0_g)\right]\\ =\ & \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left\{\theta^{0\prime}_{g^0_i}\tilde{z}_{it}(\gamma^0_{g^0_i}) - \theta'_{g_i}\tilde{z}_{it}(\gamma_{g_i})\right\}^2\\ \ge\ & \frac{c}{NT}\sum_{i=1}^N\sum_{t=1}^T\left[(\theta_{g_i} - \theta^0_{g^0_i})'\tilde{z}_{it}(\gamma_{g_i})\right]^2 + \frac{c}{NT}\sum_{i=1}^N\sum_{t=1}^T\left|\gamma_{g_i} - \gamma^0_{g^0_i}\right| E\left[(x'_{it}\delta^0_{g^0_i})^2|\gamma^0_{g^0_i}\right] + o_p(1)\\ =\ & \sum_{g=1}^G\sum_{\tilde{g}=1}^G\frac{c}{NT}\sum_{i=1}^N\sum_{t=1}^T\mathbf{1}(g^0_i = g)\mathbf{1}(g_i = \tilde{g})\left\{\left[(\theta^0_g - \theta_{\tilde{g}})'\tilde{z}_{it}(\gamma_{\tilde{g}})\right]^2 + \left|\gamma_{\tilde{g}} - \gamma^0_g\right| E\left[(x'_{it}\delta^0_g)^2|\gamma^0_g\right]\right\} + o_p(1)\\ =\ & c\sum_{g=1}^G\sum_{\tilde{g}=1}^G\left[(\theta^0_g - \theta_{\tilde{g}})'M_{NT}(g, \tilde{g}, \mathbf{D}, \mathbf{G})(\theta^0_g - \theta_{\tilde{g}}) + \left|\gamma_{\tilde{g}} - \gamma^0_g\right|\tilde{M}_{NT}(g, \tilde{g}, \mathbf{G})\right] + o_p(1),\end{aligned}$$

where the inequality holds by Assumption E.1(i) and the last equation is by the definitions of $M_{NT}(g, \tilde{g}, \mathbf{D}, \mathbf{G})$ and $\tilde{M}_{NT}(g, \tilde{g}, \mathbf{G})$. It follows that

$$\begin{aligned}o_p(1) &= c\sum_{g=1}^G\sum_{\tilde{g}=1}^G\left[(\theta^0_g - \theta_{\tilde{g}})'M_{NT}(g, \tilde{g}, \mathbf{D}, \mathbf{G})(\theta^0_g - \theta_{\tilde{g}}) + \left|\gamma_{\tilde{g}} - \gamma^0_g\right|\tilde{M}_{NT}(g, \tilde{g}, \mathbf{G})\right] + o_p(1)\\ &\ge c\sum_{g=1}^G\sum_{\tilde{g}=1}^G\left\{\lambda_{\min}[M_{NT}(g, \tilde{g}, \mathbf{D}, \mathbf{G})] \wedge \tilde{M}_{NT}(g, \tilde{g}, \mathbf{G})\right\}\left(\left\|\theta^0_g - \theta_{\tilde{g}}\right\|^2 + \left|\gamma_{\tilde{g}} - \gamma^0_g\right|\right) + o_p(1)\\ &\ge c\max_{g\in\mathcal{G}}\sum_{\tilde{g}=1}^G\left\{\lambda_{\min}[M_{NT}(g, \tilde{g}, \mathbf{D}, \mathbf{G})] \wedge \tilde{M}_{NT}(g, \tilde{g}, \mathbf{G})\right\}\left(\left\|\theta^0_g - \theta_{\tilde{g}}\right\|^2 + \left|\gamma_{\tilde{g}} - \gamma^0_g\right|\right) + o_p(1)\\ &\ge c\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta^0_g - \hat{\theta}_{\tilde{g}}\right\|^2 + \left|\gamma_{\tilde{g}} - \gamma^0_g\right|\right)\sum_{\tilde{g}=1}^G\left\{\lambda_{\min}[M_{NT}(g, \tilde{g}, \mathbf{D}, \mathbf{G})] \wedge \tilde{M}_{NT}(g, \tilde{g}, \mathbf{G})\right\} + o_p(1)\\ &\ge c\underline{c}_\lambda\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta^0_g - \hat{\theta}_{\tilde{g}}\right\|^2 + \left|\gamma_{\tilde{g}} - \gamma^0_g\right|\right) + o_p(1),\end{aligned}$$

where the last inequality is by Assumption E.1(ii) which says that there exists a group $\tilde{g}^* \in \mathcal{G}$ such that $\left\{\lambda_{\min}[M_{NT}(g,\tilde{g},\mathbf{D},\mathbf{G})] \wedge \tilde{M}_{NT}(g,\tilde{g},\mathbf{G})\right\} > \underline{c}_\lambda > 0$ w.p.a.1. Consequently, we have

$$\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|^2+\left|\gamma_{\tilde{g}}-\gamma_g^0\right|\right)=o_p\left(1\right).$$

To show (ii), we can follow a similar analysis given in the proof of Lemma A.2. The details are omitted here. ■.

Remark. The proof of Lemma E.1 shows that there exists a permutation $\sigma_{\hat{\Theta}}$ such that $\left\|\hat{\theta}_g-\theta^0_{\sigma_{\hat{\Theta}}(g)}\right\|^2+\left|\hat{\gamma}_g-\gamma^0_{\sigma_{\hat{\Theta}}(g)}\right|=o_p(1)$. We can take $\sigma_{\hat{\Theta}}(g)=g$ by relabeling. In the following analysis, we shall write $\left\|\hat{\theta}_g-\theta^0_{\sigma_{\hat{\Theta}}(g)}\right\|^2+\left|\hat{\gamma}_g-\gamma^0_{\sigma_{\hat{\Theta}}(g)}\right|=o_p(1)$ without referring to the relabeling any more.

Lemma E.2. Let $\hat{g}_i(\Theta,\mathbf{D})=\operatorname{argmin}_{g\in\mathcal{G}}\sum_{t=1}^T\left[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_g)'\theta_g\right]^2$. Suppose that Assumptions A.1, A.3(iii)-(iv) and E.1 hold. Then we have that for some $\eta>0$,

$$\Pr\left(\sup_{(\Theta,\mathbf{D})\in\tilde{\mathcal{N}}_\eta}\left[\frac{1}{N}\sum_{i=1}^N\mathbf{1}(\hat{g}_i(\Theta,\mathbf{D})\neq g_i^0)\right]\right)=o(T^{-4}),$$

where $\tilde{\mathcal{N}}_\eta=\left\{(\Theta,\mathbf{D})\in\mathcal{B}^G\times\Gamma^G:\left\|\theta_g-\theta_g^0\right\|^2+\left|\gamma_g-\gamma_g^0\right|<\eta,\ g\in\mathcal{G}\right\}$

Proof of Lemma E.2. The proof is similar to that of Lemma A.3 except the details of bounding $\mathcal{Z}_{ig}(\Theta,\mathbf{D})$, where

$$\mathcal{Z}_{ig}(\Theta,\mathbf{D})\equiv\mathbf{1}(g_i^0\neq g)\mathbf{1}\left(\sum_{t=1}^T[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_g)'\theta_g)]^2\leq\sum_{t=1}^T[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_{g_i^0})'\theta_{g_i^0}]^2\right).$$

For $\mathcal{Z}_{ig}(\Theta,\mathbf{D})$, we have

$$\mathcal{Z}_{ig}(\Theta,\mathbf{D})\leq\max_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\mathbf{1}\left(L_i(g,\tilde{g})\leq 0\right),$$

where

$$L_i(g,\tilde{g})=\sum_{t=1}^T[\tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}-\tilde{z}_{it}(\gamma_g)'\theta_g]\left\{\frac{1}{2}[\tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}-\tilde{z}_{it}(\gamma_g)'\theta_g]+\tilde{\varepsilon}_{it}+\tilde{z}_{it}(\gamma^0_{\tilde{g}})'\theta^0_{\tilde{g}}-\tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}\right\}.$$

Then we can follow the analysis of Lemma A.3 to show that

$$\mathcal{Z}_{ig}(\Theta,\mathbf{D})\leq\max_{\tilde{g}\in\boldsymbol{G}\backslash\{g\}}\mathbf{1}\left\{\sum_{t=1}^T[\tilde{z}_{it}(\gamma^0_{\tilde{g}})'\theta^0_{\tilde{g}}-\tilde{z}^0_{it}(\gamma^0_g)'\theta^0_g]\left\{\frac{1}{2}[\tilde{z}_{it}(\gamma^0_{\tilde{g}})'\theta^0_{\tilde{g}}-\tilde{z}^0_{it}(\gamma^0_g)'\theta^0_g]+\tilde{\varepsilon}_{it}\right\}\leq H_{iT}\right\}\equiv\tilde{\mathcal{Z}}_{ig},$$

where $H_{iT}=C\sqrt{\eta}\sum_{t=1}^T(\|x_{it}\|^2+\varepsilon_{it}^2)$ for some constant $C>0$. Next, we can use the Assumption E.1(i) to show that

$$\begin{aligned}\Pr(\tilde{\mathcal{Z}}_{ig} &= 1)\leq\sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{\frac{c}{2}\sum_{t=1}^T\left[(\theta^0_{\tilde{g}}-\theta^0_g)'\tilde{z}_{it}(\gamma^0_g)\right]^2+\left|\gamma^0_{\tilde{g}}-\gamma^0_g\right|E\left[(x'_{it}\delta^0_{\tilde{g}})^2|\gamma^0_{\tilde{g}}\right]\right.\\ &\left.+\sum_{t=1}^T[\tilde{z}_{it}(\gamma^0_{\tilde{g}})'\theta^0_{\tilde{g}}-\tilde{z}^0_{it}(\gamma^0_g)'\theta^0_g]\tilde{\varepsilon}_{it}\leq H_{iT}\right\}+o(T^{-4}).\end{aligned}$$

Then one can use Assumption E.1(iv) and similar arguments as used in the proof of Lemma A.3 to show that the leading term on the right hand side of the last inequality is $o(T^{-4})$. The result then follows from the Markov inequality as used in the proof of Lemma A.3. ■.

REFERENCES

Davidson, J., 1994. Stochastic Limit Theory: An Introduction for Econometricians. Oxford University Press, Oxford.

Hansen, B. E., 2000. Sample splitting and threshold estimation. Econometrica 68, 575-603.

Wang, W., Phillips, P.C. and Su, L., 2018. Homogeneity pursuit in panel data models: Theory and application. Journal of Applied Econometrics 33, 797-815.

Yang, S.C., 2006. Maximal moment inequality for partial sums of strong mixing sequences and application. Acta Mathematica Sinica 23, 1013-1024.