Econometrica Supplementary Material

SUPPLEMENT TO "IDENTIFYING LATENT STRUCTURES IN PANEL DATA"

(*Econometrica*, Vol. 84, No. 6, November 2016, 2215–2264)

BY LIANGJUN SU, ZHENTAO SHI, AND PETER C. B. PHILLIPS[1]

THIS SUPPLEMENT IS COMPOSED OF FOUR PARTS. Section S1 contains the proofs of some technical lemmas for the proofs of the main results in Section 2. Section S2 gives bias correction formulae in linear panel data models for both PPL and PGMM estimation. Sections S3 and S4 contain some additional simulation and applications results, respectively.

S1. SOME TECHNICAL LEMMAS FOR THE PROOFS OF THE MAIN RESULTS IN SECTION 2 OF THE PAPER

In this appendix, we state and prove some technical lemmas that are used in the proofs of the main results in Section 2. We first state an exponential inequality for strong mixing processes.

LEMMA S1.1: *Let $\{\zeta_t, t = 1, 2, \ldots\}$ be a zero-mean strong mixing process, not necessarily stationary, with the mixing coefficients satisfying $\alpha(\tau) \leq c_\alpha \rho^\tau$ for some $c_\alpha > 0$ and $\rho \in (0, 1)$. If $\sup_{1\leq t\leq T} |\zeta_t| \leq M_T$, then there exists a constant C_0 depending on c_α and ρ such that for any $T \geq 2$ and $\epsilon > 0$,*

$$P\left(\left|\sum_{t=1}^{T} \zeta_t\right| > \epsilon\right) \leq \exp\left(-\frac{C_0\epsilon^2}{v_0^2 T + M_T^2 + \epsilon M_T(\ln T)^2}\right),$$

where $v_0^2 = \sup_{t\geq 1}[\mathrm{Var}(\zeta_t) + 2\sum_{s=t+1}^{\infty} |\mathrm{Cov}(\zeta_t, \zeta_s)|]$.

PROOF: Merlevède, Peilgrad, and Rio (2009, Theorem 2) proved (i) under the condition $\alpha(\tau) \leq \exp(-2c\tau)$ for some $c > 0$. If $c_\alpha = 1$, we can take $\rho = \exp(-2c)$ and apply the theorem to obtain the claim. *Q.E.D.*

The above lemma is used in the proof of the following lemma.

LEMMA S1.2: *Let $\xi(w_{it}; \phi)$ be a $\mathbb{R}^{d_\xi}$-valued function indexed by the parameter $\phi \in \Phi$, where Φ is a convex compact set in $\mathbb{R}^{p+1}$ and $\mathbb{E}[\xi(w_{it}; \phi)] = 0$ for all i,*

[1]Acknowledgments made in the leading footnote of the main paper apply also to this supplement. In particular, Su acknowledges support from the Singapore Ministry of Education for Academic Research Fund (AcRF) under the Tier-2 Grant MOE2012-T2-2-021 and the funding support provided by the Lee Kong Chian Fund for Excellence. Phillips acknowledges NSF support under Grants SES-0956687 and SES-1285258. Address correspondence to: Liangjun Su.

 DOI: 10.3982/ECTA12560

t, and $\phi \in \Phi$. Assume that there exists a function $M(w_{it})$ such that $\|\xi(w_{it};\phi) - \xi(w_{it};\overline{\phi})\| \leq M(w_{it})\|\phi - \overline{\phi}\|$ for all $\phi, \overline{\phi} \in \Phi$ and $\sup_{\phi}\|\xi(w_{it};\phi)\| \leq M(w_{it})$. Assume that $\mathbb{E}|M(w_{it})|^q < \infty$ for some $q \geq 6$ such that $N = O(T^{q/2-1})$. Let $\{\phi_i\}$ be a nonstochastic sequence in Φ. Then

(i) $\max_{1\leq i\leq N}\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi(w_{it};\phi_i)\| = O_P((\ln T)^3)$,

(ii) $\max_{1\leq i\leq N} P(\|\frac{1}{T}\sum_{t=1}^{T}\xi(w_{it};\phi_i)\| \geq c\lambda) = o(N^{-1})$ *for any given* $c > 0$,

(iii) $P(\max_{1\leq i\leq N}\|\frac{1}{T}\sum_{t=1}^{T}\xi(w_{it};\phi_i)\| \geq c\lambda) = o(N^{-1})$ *for any given* $c > 0$ *if* $N^2 = O(T^{q/2-1})$,

where $\lambda = \lambda_{NT}$ *satisfies* $(\ln T)^3 = o(T^{1/2}\lambda)$.

PROOF: (i) Let $\eta_{NT} = T^{1/2}$. Let ι_ξ be an arbitrary $d_\xi \times 1$ nonrandom vector with $\|\iota_\xi\| = 1$. Let $\mathbf{1}_{it} = \mathbf{1}\{\|\xi(w_{it};\phi_i)\| \leq \eta_{NT}\}$ and $\overline{\mathbf{1}}_{it} = 1 - \mathbf{1}_{it}$. Define

$$\xi_1(w_{it};\phi_i) = \iota_\xi'\{\xi(w_{it};\phi_i)\mathbf{1}_{it} - \mathbb{E}[\xi(w_{it};\phi_i)\mathbf{1}_{it}]\},$$

$$\xi_2(w_{it};\phi_i) = \iota_\xi'\xi(w_{it};\phi_i)\overline{\mathbf{1}}_{it}, \quad \text{and} \quad \xi_{3it} = -\iota_\xi'\mathbb{E}[\xi(w_{it};\phi_i)\overline{\mathbf{1}}_{it}].$$

Apparently, $\xi_1(w_{it};\phi_i) + \xi_2(w_{it};\phi_i) + \xi_{3it} = \iota_\xi'\xi(w_{it};\phi_i)$ as $\mathbb{E}[\xi(w_{it};\phi_i)] = 0$. We prove the lemma by showing that (i1) $\max_{1\leq i\leq N}\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_1(w_{it};\phi_i)\| = O_P((\ln T)^3)$, (i2) $P[\max_{1\leq i\leq N}\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_2(w_{it};\phi_i)\| \geq c(\ln T)^3] = o(1)$ for any given $c > 0$, and (i3) $\max_{1\leq i\leq N}\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_{3it}\| = O((\ln T)^3)$.

First, we prove (i3). By the Hölder and Markov inequalities,

$$\begin{aligned}
\max_{1\leq i\leq N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_{3it}\right\| &\leq T^{1/2}\max_{1\leq i\leq N}\max_{1\leq t\leq T}\left\|\mathbb{E}\left[\xi(w_{it};\phi_i)\overline{\mathbf{1}}_{it}\right]\right\| \\
&\leq T^{1/2}\max_{1\leq i\leq N}\max_{1\leq t\leq T}\left\{\mathbb{E}\left\|\xi(w_{it};\phi_i)\right\|^{q/2}\right\}^{2/q} \\
&\quad \times \left\{P\left(\left\|\xi(w_{it};\phi_i)\right\| > T^{1/2}\right)\right\}^{(q-2)/q} \\
&\leq T^{1/2}c_{1q}\max_{1\leq i\leq N}\max_{1\leq t\leq T}\left\{P\left(\left\|\xi(w_{it};\phi_i)\right\| > T^{1/2}\right)\right\}^{(q-2)/q} \\
&\leq T^{1/2}c_{1q}\max_{1\leq i\leq N}\max_{1\leq t\leq T}\left\{T^{-q/2}\mathbb{E}\left(\left\|\xi(w_{it};\phi_i)\right\|^q\right)\right\}^{(q-2)/q} \\
&= c_{1q}c_{2q}T^{(3-q)/2} = o\left((\ln T)^3\right) \quad \text{for any } q \geq 3,
\end{aligned}$$

where

$$c_{1q} \equiv \max_{1\leq i\leq N}\max_{1\leq t\leq T}\left\{\mathbb{E}\left\|\xi(w_{it};\phi_i)\right\|^{q/2}\right\}^{2/q} \quad \text{and}$$

$$c_{2q} \equiv \max_{1\leq i\leq N}\max_{1\leq t\leq T}\left\{\mathbb{E}\left(\left\|\xi(w_{it};\phi_i)\right\|^q\right)\right\}^{(q-2)/q}.$$

Next, we prove (i2). Noting that $\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_2(w_{it};\phi_i)\| \geq (\ln T)^3$ implies that $\max_{1\leq t\leq T}\|\xi(w_{it};\phi_i)\| > \eta_{NT}$, by the Boole and Markov inequalities, the dominated convergence theorem, and the stated conditions, we have

$$\begin{aligned}
&P\left[\max_{1\leq i\leq N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_2(w_{it};\phi_i)\right\| \geq c(\ln T)^3\right] \\
&\quad \leq P\Big[\max_{1\leq i\leq N}\max_{1\leq t\leq T}\|\xi(w_{it};\phi_i)\| > \eta_{NT}\Big] \\
&\quad \leq NT\max_{1\leq i\leq N}\max_{1\leq t\leq T}P\big(M(w_{it}) > \eta_{NT}\big) \\
&\quad \leq \frac{NT}{T^{q/2}}\max_{1\leq i\leq N}\max_{1\leq t\leq T}\mathbb{E}\big[\big|M(w_{it})\big|^q\mathbf{1}\big\{M(w_{it}) > T^{1/2}\big\}\big] \\
&\quad = o\big(NT^{1-q/2}\big) = o(1).
\end{aligned}$$

Now, we prove (i1). We observe that for any $C > 0$,

$$\begin{aligned}
&P\left[\max_{1\leq i\leq N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_1(w_{it};\phi_i)\right\| \geq C(\ln T)^3\right] \\
&\quad \leq \sum_{i=1}^{N}P\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_1(w_{it};\phi_i)\right\| \geq C(\ln T)^3\right].
\end{aligned}$$

We choose $\varepsilon > 0$ and divide Φ into subsets Φ_j, $j = 1, \ldots, n_\varepsilon$ such that $\|\phi - \overline{\phi}\| < \varepsilon/\sqrt{T}$ for all $\phi, \overline{\phi} \in \Phi_j$, where $n_\varepsilon = O(T^{(p+1)/2})$. Then

$$\begin{aligned}
&\sum_{i=1}^{N}P\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_1(w_{it};\phi_i)\right\| \geq C(\ln T)^3\right] \\
&\quad \leq \sum_{i=1}^{N}P\left[\sup_{\phi\in\Phi}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_1(w_{it};\phi)\right\| \geq C(\ln T)^3\right] \\
&\quad \leq \sum_{i=1}^{N}\sum_{j=1}^{n_\varepsilon}P\left[\sup_{\phi\in\Phi_j}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_1(w_{it};\phi)\right\| \geq C(\ln T)^3\right].
\end{aligned}$$

Let $\phi_j \in \Phi_j$. Then for any $\phi \in \Phi_j$, we have

$$\begin{aligned}
&\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_1(w_{it}; \phi) \right\| \\
&\leq \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_1(w_{it}; \phi_j) \right\| + \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\xi_1(w_{it}; \phi_j) - \xi_1(w_{it}; \phi) \right] \right\| \\
&\leq \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_1(w_{it}; \phi_j) \right\| + \frac{2}{\sqrt{T}} \sum_{t=1}^{T} M(w_{it}) \|\phi - \phi_j\| \\
&\leq \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_1(w_{it}; \phi_j) \right\| + \left| \frac{2\varepsilon}{T} \sum_{t=1}^{T} \left\{ M(w_{it}) - \mathbb{E}\left[M(w_{it}) \right] \right\} \right| \\
&\quad + \frac{2\varepsilon}{T} \sum_{t=1}^{T} \mathbb{E}\left[M(w_{it}) \right].
\end{aligned}$$

It follows that

$$\begin{aligned}
&P\left[\sup_{\phi \in \Phi_j} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_1(w_{it}; \phi) \right\| \geq C(\ln T)^3 \right] \\
&\leq P\left[\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_1(w_{it}; \phi_j) \right\| \geq C(\ln T)^3/3 \right] \\
&\quad + P\left[\left| \frac{2\varepsilon}{T} \sum_{t=1}^{T} M(w_{it}) - \mathbb{E}\left[M(w_{it}) \right] \right| \geq C(\ln T)^3/3 \right]
\end{aligned}$$

as $P[\frac{\varepsilon}{T} \sum_{t=1}^{T} \mathbb{E}[M(w_{it})] \geq C(\ln T)^3/3] = 0$. Then

$$\begin{aligned}
&P\left[\max_{1 \leq i \leq N} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi(w_{it}; \phi_i) \right\| \geq C(\ln T)^3 \right] \\
&\leq \sum_{i=1}^{N} \sum_{j=1}^{n_\varepsilon} P\left[\left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \xi_1(w_{it}; \phi_j) \right\| \geq C(\ln T)^3/3 \right] \\
&\quad + \sum_{i=1}^{N} \sum_{j=1}^{n_\varepsilon} P\left[\left| \frac{2\varepsilon}{T} \sum_{t=1}^{T} M(w_{it}) - \mathbb{E}\left[M(w_{it}) \right] \right| \geq C(\ln T)^3/3 \right].
\end{aligned}$$

For the first term, we have, by Lemma S1.1,

$$\begin{aligned}
&\sum_{i=1}^{N}\sum_{j=1}^{n_\varepsilon} P\left[\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_1(w_{it};\phi_j)\right\| \geq C(\ln T)^3/3\right] \\
&\quad \leq CNn_\varepsilon \exp\left(-\frac{C^2C_0T(\ln T)^6/9}{\overline{v}^2T + 4\eta_{NT}^2 + \frac{2C}{3}\eta_{NT}T^{1/2}(\ln T)^{3+2}}\right) \\
&\quad = C\exp\left(-\frac{C^2C_0T(\ln T)^6/9}{\overline{v}^2T + 4T + \frac{2C}{3}T(\ln T)^5} + \ln N + \ln n_\varepsilon\right) \\
&\quad \to 0 \quad \text{for sufficiently large } C.
\end{aligned}$$

Similarly, we can show that $\sum_{i=1}^{N}\sum_{j=1}^{n_\varepsilon} P[\|\frac{2\varepsilon}{T}\sum_{t=1}^{T} M(w_{it}) - \mathbb{E}[M(w_{it})]\| \geq C(\ln T)^3/3] = o(1)$. Then (i1) follows. This completes the proof of (i).

(ii) Let ξ_1, ξ_2, and ξ_{3it} be as defined in (i). Noting that ξ_{3it} is nonrandom, it suffices to show that for any given $c > 0$, we have (ii1) $N\max_{1\leq i\leq N} P(\|\frac{1}{T} \times \sum_{t=1}^{T}\xi_1(w_{it};\phi_i)\| \geq c\lambda) = o(1)$, (ii2) $N\max_{1\leq i\leq N} P(\|\frac{1}{T}\sum_{t=1}^{T}\xi_2(w_{it};\phi_i)\| \geq c\lambda) = o(1)$, and (ii3) $\max_{1\leq i\leq N}\|\frac{1}{T}\sum_{t=1}^{T}\xi_{3it}\| = o(\lambda)$. Following the analysis of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_{3it}$ in (i), we have

$$\max_{1\leq i\leq N}\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_{3it}\right\| \leq c_{1q}c_{2q}T^{(2-q)/2} = o(\lambda),$$

where we use the fact that $\lambda \gg T^{-1/2}(\ln T)^3$ and $q \geq 3$ by the stated conditions. Thus, (ii3) follows. Following the analysis of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_3(w_{it};\phi_i)$ in (i2), we have

$$\begin{aligned}
&N\max_{1\leq i\leq N} P\left(\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_2(w_{it};\phi_i)\right\| \geq c\lambda\right) \\
&\quad \leq N\max_{1\leq i\leq N} P\Big(\max_{1\leq t\leq T}\big\|\xi(w_{it};\phi_i)\big\| > \eta_{NT}\Big) \\
&\quad \leq NT\max_{1\leq i\leq N}\max_{1\leq t\leq T} P\big(M(w_{it}) > \eta_{NT}\big) \\
&\quad = o\big(NT^{1-q/2}\big) = o(1).
\end{aligned}$$

That is, (ii2) follows. For (ii1), the analysis is similar to that of $\max_{1\leq i\leq N}\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_1(w_{it};\phi_i)\|$ in (i1) with $(\ln T)^3$ replaced by $T^{1/2}\lambda$. We now require $T^{1/2}\lambda/(\ln T)^3 \to \infty$ as $(N,T)\to\infty$. This completes the proof of (ii).

(iii) Let ξ_1, ξ_2, and ξ_{3it} be as defined in (i). Noting that ξ_{3it} is nonrandom, it suffices to show that for any given $c > 0$, we have (iii1) $N \cdot P(\max_{1\le i\le N} \|\frac{1}{T}\sum_{t=1}^{T} \xi_1(w_{it};\phi_i)\| \ge c\lambda) = o(1)$, (iii2) $N \cdot P(\max_{1\le i\le N} \|\frac{1}{T} \times \sum_{t=1}^{T} \xi_2(w_{it};\phi_i)\| \ge c\lambda) = o(1)$, and (iii3) $\max_{1\le i\le N} \|\frac{1}{T}\sum_{t=1}^{T} \xi_{3it}\| = o(\lambda)$. Following the analysis of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \xi_{3it}$ in (i), we have

$$\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_{3it}\right\| \le c_{1q}c_{2q}T^{(2-q)/2} = o(\lambda),$$

where we use the fact that $\lambda \gg T^{-1/2}(\ln T)^3$ and $q \ge 6$ by the stated conditions. Thus, (iii3) follows. Following the analysis of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \xi_3(w_{it};\phi_i)$ in (i2), we have

$$\begin{aligned} &N \cdot P\left(\max_{1\le i\le N}\left\|\frac{1}{T}\sum_{t=1}^{T}\xi_2(w_{it};\phi_i)\right\| \ge c\lambda\right) \\ &\quad\le N \cdot P\Big(\max_{1\le i\le N}\max_{1\le t\le T}\big\|\xi(w_{it};\phi_i)\big\| > \eta_{NT}\Big) \\ &\quad\le N^2 T \max_{1\le i\le N}\max_{1\le t\le T} P\big(M(w_{it}) > \eta_{NT}\big) \\ &\quad= o\big(N^2 T^{1-q/2}\big) = o(1). \end{aligned}$$

That is, (iii2) follows. For (iii1), the analysis is similar to that of $\max_{1\le i\le N} \|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \xi_1(w_{it};\phi_i)\|$ in (i1) with $(\ln T)^3$ replaced by $T^{1/2}\lambda$. We now require $T^{1/2}\lambda/(\ln T)^3 \to \infty$ as $(N, T) \to \infty$. This completes the proof of (ii). *Q.E.D.*

Recall that $\hat{\Psi}_i(\beta,\mu) = \frac{1}{T}\sum_{t=1}^{T} \psi(w_{it};\beta,\mu)$ and $\Psi_i(\beta,\mu) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\psi(w_{it};\beta,\mu)]$. Recall that $\hat{\mu}_i(\beta_i) = \arg\min_{\mu_i} \frac{1}{T}\sum_{t=1}^{T} \psi(w_{it};\beta_i,\mu_i)$. The following three lemmas study the properties of $\hat{\Psi}_i(\beta,\mu)$ and $\hat{\mu}_i(\beta_i)$.

LEMMA S1.3: *For any* $\eta > 0$, *we have* $P[\max_{1\le i\le N}\sup_{(\beta,\mu)} |\hat{\Psi}_i(\beta,\mu) - \Psi_i(\beta,\mu)| \ge \eta] = o(N^{-1})$.

PROOF: The proof is analogous to that of Lemma S1.2(iii). *Q.E.D.*

LEMMA S1.4: *For any* $\eta > 0$, *we have* $P[\max_{1\le i\le N} |\hat{\mu}_i(\beta_i) - \mu_i(\beta_i)| \ge \eta] = o(N^{-1})$.

PROOF: Let $\varepsilon = \min_i[\inf_{\mu_i:|\mu_i-\mu_i(\beta_i)|>\eta} \Psi_i(\beta_i,\mu_i) - \Psi_i(\beta_i,\mu_i(\beta_i))]$. Then $\varepsilon > 0$ by Assumption A1(ii) and (v). Then, conditional on the event $A \equiv$

$\{\max_{1\le i\le N}\sup_{(\beta,\mu)}|\hat{\Psi}_i(\beta,\mu)-\Psi_i(\beta,\mu)|\le\frac{1}{3}\varepsilon\}$, we have

$$\begin{aligned}\inf_{|\mu_i-\mu_i(\beta_i)|>\eta}\hat{\Psi}_i(\beta_i,\mu_i) &\ge \inf_{|\mu_i-\mu_i(\beta_i)|>\eta}\Psi_i(\beta_i,\mu_i)-\frac{1}{3}\varepsilon\\ &\ge \Psi_i\big(\beta_i,\mu_i(\beta_i)\big)+\frac{2}{3}\varepsilon\\ &\ge \hat{\Psi}_i\big(\beta_i,\mu_i(\beta_i)\big)+\frac{1}{3}\varepsilon.\end{aligned}$$

On the other hand, $\hat{\Psi}_i(\beta_i,\hat{\mu}_i(\beta_i))\le\hat{\Psi}_i(\beta_i,\mu_i(\beta_i))$. It follows that $P(\max_{1\le i\le N}|\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)|\le\eta)\le P(A)=o(N^{-1})$ by Lemma S1.3. *Q.E.D.*

LEMMA S1.5: (i) $\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)=O_P(T^{-1/2})$ *for each* i,
(ii) $\max_{1\le i\le N}|\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)|=O_P(T^{-1/2}(\ln T)^3)$,
(iii) $\max_{1\le i\le N}|\Psi_i(\beta_i,\hat{\mu}_i(\beta_i))-\Psi_i(\beta_i,\mu_i(\beta_i))|=O_P(T^{-1/2}(\ln T)^3)$,
(iv) $P(\max_{1\le i\le N}|\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)|\ge CT^{-1/2}(\ln T)^{3+\nu})=o(N^{-1})$ *for any* $\nu>0$ *and* $C>0$,
(v) $P(\max_{1\le i\le N}|\Psi_i(\beta_i,\hat{\mu}_i(\beta_i))-\Psi_i(\beta_i,\mu_i(\beta_i))|\ge CT^{-1/2}(\ln T)^{3+\nu})=o(N^{-1})$ *for any* $\nu>0$ *and* $C>0$.

PROOF: (i)–(ii) Noting that $\hat{\mu}_i(\beta_i)=\arg\min_{\mu_i}\frac{1}{T}\sum_{t=1}^T\psi(w_{it};\beta_i,\mu_i)$, we have

$$\begin{aligned}0&=\frac{1}{T}\sum_{t=1}^T V_i\big(w_{it};\beta_i,\hat{\mu}_i(\beta_i)\big)\\ &=\frac{1}{T}\sum_{t=1}^T V_i\big(w_{it};\beta_i,\mu_i(\beta_i)\big)\\ &\quad+\frac{1}{T}\sum_{t=1}^T V_i^{\mu_i}\big(w_{it};\beta_i,\breve{\mu}_i(\beta_i)\big)\big[\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)\big],\end{aligned}$$

where $\breve{\mu}_i(\beta_i)$ lies between $\hat{\mu}_i(\beta_i)$ and $\mu_i(\beta_i)$ for each i. It follows that

$$\begin{aligned}&\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)\\ &\quad=-\left[\frac{1}{T}\sum_{t=1}^T V_i^{\mu_i}\big(w_{it};\beta_i,\breve{\mu}_i(\beta_i)\big)\right]^{-1}\frac{1}{T}\sum_{t=1}^T V_i\big(w_{it};\beta_i,\mu_i(\beta_i)\big)\end{aligned}\tag{S1}$$

provided $\frac{1}{T}\sum_{t=1}^{T} V_i^{\mu_i}(w_{it}; \beta_i, \breve{\mu}_i(\beta_i))$ is asymptotically nonvanishing. Let $V_{it}(\beta_i) = V_i(w_{it}; \beta_i, \mu_i(\beta_i))$. Noting that $\mathbb{E}[V_{it}(\beta_i)] = 0$ and

$$\begin{aligned}
\operatorname{Var}\left(\frac{1}{T}\sum_{t=1}^{T} V_{it}(\beta_i)\right) &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\operatorname{Cov}\bigl(V_{it}(\beta_i), V_{is}(\beta_i)\bigr) \\
&\leq 8\max_{i,t}\bigl\{\mathbb{E}\bigl|V_{it}(\beta_i)\bigr|^q\bigr\}^{2/q}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\alpha\bigl(|t-s|\bigr)^{1-2/q} \\
&\leq 8\max_{i,t}\bigl\{\mathbb{E}\bigl|V_{it}(\beta_i)\bigr|^q\bigr\}^{2/q}\frac{1}{T}\sum_{\tau=1}^{\infty}\alpha(\tau)^{1-2/q} = O\left(\frac{1}{T}\right)
\end{aligned}$$

by the Davydov inequality (e.g., Corollary A.2 in Hall and Heyde (1980)), we have $\frac{1}{T}\sum_{t=1}^{T} V_{it}(\beta_i) = O_P(T^{-1/2})$ by the Chebyshev inequality. In addition, by a simple application of Lemma S1.2(i), we can show that $\max_{1\leq i\leq N}|\frac{1}{T}\times\sum_{t=1}^{T} V_{it}(\beta_i)| = O_P(T^{-1/2}(\ln T)^3)$.

For $\frac{1}{T}\sum_{t=1}^{T} V_i^{\mu_i}(w_{it}; \beta_i, \breve{\mu}_i(\beta_i))$, we make the following decomposition:

$$\begin{aligned}
\text{(S2)}\qquad &\frac{1}{T}\sum_{t=1}^{T} V_i^{\mu_i}\bigl(w_{it}; \beta_i, \breve{\mu}_i(\beta_i)\bigr) \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\bigl[V_i^{\mu_i}\bigl(w_{it}; \beta_i, \mu_i(\beta_i)\bigr)\bigr] \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\bigl\{V_i^{\mu_i}\bigl(w_{it}; \beta_i, \mu_i(\beta_i)\bigr) - \mathbb{E}\bigl[V_i^{\mu_i}\bigl(w_{it}; \beta_i, \mu_i(\beta_i)\bigr)\bigr]\bigr\} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\bigl\{V_i^{\mu_i}\bigl(w_{it}; \beta_i, \breve{\mu}_i(\beta_i)\bigr) - V_i^{\mu_i}\bigl(w_{it}; \beta_i, \mu_i(\beta_i)\bigr)\bigr\}.
\end{aligned}$$

By Assumption A1(v), $\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[V_i^{\mu_i}(w_{it}; \beta_i, \mu_i(\beta_i))] = H_{i\mu\mu}(\beta_i) \geq c_H > 0$ uniformly in i. By a simple application of Lemma S1.2(i), we have

$$\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\bigl\{V_i^{\mu_i}\bigl(w_{it}; \beta_i, \mu_i(\beta_i)\bigr) - \mathbb{E}\bigl[V_i^{\mu_i}\bigl(w_{it}; \beta_i, \mu_i(\beta_i)\bigr)\bigr]\bigr\}\right| = o_P(1).$$

Next, by Assumption A1, and Lemmas S1.2(i) and S1.4, we have

$$\text{(S3)}\qquad \max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\left[V_i^{\mu_i}\big(w_{it};\beta_i,\breve{\mu}_i(\beta_i)\big)-V_i^{\mu_i}\big(w_{it};\beta_i,\mu_i(\beta_i)\big)\right]\right|$$

$$\le \max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}M(w_{it})\left|\breve{\mu}_i(\beta_i)-\mu_i(\beta_i)\right|$$

$$\le \left\{\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[M(w_{it})\big]+\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\big\{M(w_{it})-\mathbb{E}\big[M(w_{it})\big]\big\}\right|\right\}$$

$$\times \max_{1\le i\le N}\left|\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)\right|$$

$$\le \left[c_M^{1/q}+o_P(1)\right]o_P(1)=o_P(1).$$

It follows that $\frac{1}{T}\sum_{t=1}^{T}V_i^{\mu_i}(w_{it};\beta_i,\breve{\mu}_i(\beta_i)) = H_{i\mu\mu}(\beta_i)+o_P(1)$ uniformly in i, $\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)=O_P(T^{-1/2})$ for each i, and $\max_{1\le i\le N}|\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)| = O_P(T^{-1/2}(\ln T)^3)$.

(iii) In view of the definition that $\Psi_i(\beta_i,\mu_i)=\mathbb{E}[\psi(w_{it};\beta_i,\mu_i)]$, we have $\max_{1\le i\le N}|\Psi_i(\beta_i,\hat{\mu}_i(\beta_i))-\Psi_i(\beta_i,\ \mu_i(\beta_i))| = \max_{i,t}\mathbb{E}|M(w_{it})||\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)| = O_P(T^{-1/2}(\ln T)^3)$.

(iv) We define the following events:

$$A_1\equiv\left\{\max_{1\le i\le N}\left|\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)\right|\le c_H/\big(6c_M^{1/q}\big)\right\},$$

$$A_2\equiv\left\{\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\big\{M(w_{it})-\mathbb{E}\big[M(w_{it})\big]\big\}\right|\le c_M^{1/q}/2\right\},$$

$$A_3\equiv\left\{\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}\left[V_i^{\mu_i}\big(w_{it};\beta_i,\breve{\mu}_i(\beta_i)\big)-V_i^{\mu_i}\big(w_{it};\beta_i,\mu_i(\beta_i)\big)\right]\right|\right.$$

$$\left.\le c_H/4\right\},$$

$$A_4\equiv\left\{\min_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}V_i^{\mu_i}\big(w_{it};\beta_i,\mu_i(\beta_i)\big)\right|\ge c_H/2\right\},$$

$$A_5\equiv\left\{\min_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T}V_i^{\mu_i}\big(w_{it};\beta_i,\breve{\mu}_i(\beta_i)\big)\right|\ge c_H/4\right\}.$$

Let A_j^c denote the complement of A_j for $j = 1, 2, 3, 4, 5$. Let $\delta_i = \hat{\mu}_i(\beta_i) - \mu_i(\beta_i)$. By Lemmas S1.4 and S1.2(iii), $P(A_1^c) = o(N^{-1})$ and $P(A_2^c) = o(N^{-1})$. Then by (S3),

$$\begin{aligned}
&P\big(A_3^c\big) \\
&\quad \le P\Bigg(\Bigg\{\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[M(w_{it})\big] \\
&\qquad + \max_{1\le i\le N}\Bigg|\frac{1}{T}\sum_{t=1}^{T}\big\{M(w_{it}) - \mathbb{E}\big[M(w_{it})\big]\big\}\Bigg|\Bigg\}\max_{1\le i\le N}|\delta_i| \ge c_H/4\Bigg) \\
&\quad \le P\Bigg(\Bigg\{\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big[M(w_{it})\big] \\
&\qquad + \max_{1\le i\le N}\Bigg|\frac{1}{T}\sum_{t=1}^{T}\big\{M(w_{it}) - \mathbb{E}\big[M(w_{it})\big]\big\}\Bigg|\Bigg\}\max_{1\le i\le N}|\delta_i| \ge c_H/4,\, A_2\Bigg) \\
&\qquad + P\big(A_2^c\big) \\
&\quad \le P\Big(3c_M^{1/q}\max_{1\le i\le N}|\delta_i| \ge c_H/2\Big) + P\big(A_2^c\big) \\
&\quad \le P\big(A_1^c\big) + P\big(A_2^c\big) = o\big(N^{-1}\big).
\end{aligned}$$

Let $V_{it}^{\mu_i}(\beta_i) \equiv V_i^{\mu_i}(w_{it}; \beta_i, \mu_i(\beta_i))$. Noting that $\frac{1}{T}\sum_{t=1}^{T} V_{it}^{\mu_i}(\beta_i) = \frac{1}{T} \times \sum_{t=1}^{T}\mathbb{E}[V_{it}^{\mu_i}(\beta_i)] + \frac{1}{T}\sum_{t=1}^{T}\{V_{it}^{\mu_i}(\beta_i) - \mathbb{E}[V_{it}^{\mu_i}(\beta_i)]\}$, $\min_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[V_{it}^{\mu_i}(\beta_i)] \ge c_H$ by Assumption A1(v), and

$$\begin{aligned}
&P\Bigg(\max_{1\le i\le N}\Bigg|\frac{1}{T}\sum_{t=1}^{T}\big\{V_{it}^{\mu_i}(\beta_i) - \mathbb{E}\big[V_{it}^{\mu_i}(\beta_i)\big]\big\}\Bigg| \ge c_H/2\Bigg) \\
&\quad = o\big(N^{-1}\big) \quad \text{by Lemma S1.2(iii)},
\end{aligned}$$

we have $P(A_4) = P(\min_{1\le i\le N}|\frac{1}{T}\sum_{t=1}^{T} V_i^{\mu_i}(w_{it}; \beta_i, \mu_i(\beta_i))| \ge c_H/2) = 1 - o(N^{-1})$. It follows that

$$P(A_5) \ge P(A_3 \cap A_4) \ge 1 - P\big(A_3^c\big) - P\big(A_4^c\big) = 1 - o\big(N^{-1}\big).$$

Consequently, we have that by Lemma S1.2(iii),

$$\begin{aligned}
&N \cdot P\Big(\max_{1\le i\le N}\big|\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)\big| \ge T^{-1/2}(\ln T)^{3+\nu}\Big)\\
&\quad = N\cdot P\left(\left(\frac{c_H}{4}\right)^{-1}\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T} V_i\big(w_{it};\beta_i,\mu_i(\beta_i)\big)\right|\right.\\
&\qquad \left.\ge T^{-1/2}(\ln T)^{3+\nu}, A_5\right) + N\cdot P\big(A_5^c\big)\\
&\quad \le N\cdot P\left(\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^{T} V_i\big(w_{it};\beta_i,\mu_i(\beta_i)\big)\right| \ge c_H T^{-1/2}(\ln T)^{3+\nu}/4\right)\\
&\qquad + N\cdot P\big(A_5^c\big)\\
&\quad = o(1)+o(1)=o(1).
\end{aligned}$$

(v) Noting that $|\Psi_i(\beta_i,\hat{\mu}_i(\beta_i)) - \Psi_i(\beta_i,\mu_i(\beta_i))| \le \mathbb{E}[M(w_{it})]|\hat{\mu}_i(\beta_i) - \mu_i(\beta_i)| \le c_M^{1/q}|\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)|$ by Assumption A1(iv), the result follows directly from (iv). *Q.E.D.*

Recall that $\hat{S}_i \equiv \frac{1}{T}\sum_{t=1}^{T} U_i(w_{it};\beta_i^0,\hat{\mu}_i(\beta_i^0))$. Let $S_i \equiv \frac{1}{T}\sum_{t=1}^{T} U_i(w_{it};\beta_i^0,\mu_i^0)$. The next lemma studies the asymptotic properties of $\hat{S}_i$, S_i, and their difference.

LEMMA S1.6: *(i)* $S_i = O_P(T^{-1/2})$ *and* $\hat{S}_i - S_i = O_P(T^{-1/2})$ *for each* i,
(ii) $\max_{1\le i\le N}\|S_i\| = O_P(T^{-1/2}(\ln T)^3)$ *and* $\max_{1\le i\le N}\|\hat{S}_i - S_i\| = O_P(T^{-1/2}(\ln T)^3)$,
(iii) $\frac{1}{N}\sum_{i=1}^{N}\|\hat{S}_i\|^2 = O_P(T^{-1})$,
(iv) $\max_{1\le i\le N} P(\|S_i\|\ge \eta\lambda_1) = o(N^{-1})$ *for any given constant* $\eta>0$,
(v) $P(\max_{1\le i\le N}\|\hat{S}_i - S_i\| \ge CT^{-1/2}(\ln T)^{3+\nu}) = o(N^{-1})$ *for any* $\nu>0$ *and* $C>0$.

PROOF: (i) Let ι_p be an arbitrary $p\times 1$ nonrandom vector with $\|\iota_p\|=1$. Recall that $U_{it} = U_i(w_{it};\beta_i^0,\mu_i^0)$. Note that $\mathbb{E}(U_{it})=0$ and

$$\begin{aligned}
\operatorname{Var}\big(\iota_p' S_i\big) &= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\operatorname{Cov}\big(\iota_p' U_{it}, U_{is}'\iota_p\big)\\
&\le 8\max_{i,t}\big\{\mathbb{E}\big|\iota_p' U_{it}\big|^q\big\}^{2/q}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\alpha\big(|t-s|\big)^{1-2/q}
\end{aligned}$$

$$\leq 8 \max_{i,t} \{\mathbb{E}|\iota_p' U_{it}|^q\}^{2/q} \frac{1}{T} \sum_{\tau=1}^{\infty} \alpha(\tau)^{1-2/q} = O\left(\frac{1}{T}\right)$$

by the Davydov inequality (e.g., Corollary A.2 in Hall and Heyde (1980)). Then $S_i = \frac{1}{T}\sum_{t=1}^{T} U_i(w_{it}; \beta_i^0, \mu_i^0) = O_P(T^{-1/2})$ by the Chebyshev inequality. By second-order Taylor expansion,

$$\begin{aligned}
\hat{S}_i - S_i &= \frac{1}{T}\sum_{t=1}^{T}\left[U_i\left(w_{it}; \beta_i^0, \hat{\mu}_i(\beta_i^0)\right) - U_i\left(w_{it}; \beta_i^0, \mu_i^0\right)\right] \\
&= \frac{1}{T}\sum_{t=1}^{T} U_i^{\mu_i}\left(w_{it}; \beta_i^0, \mu_i(\beta_i^0)\right)\left[\hat{\mu}_i(\beta_i^0) - \mu_i(\beta_i^0)\right] \\
&\quad + \frac{1}{2T}\sum_{t=1}^{T} U_i^{\mu_i\mu_i}\left(w_{it}; \beta_i^0, \check{\mu}_i(\beta_i^0)\right)\left[\hat{\mu}_i(\beta_i^0) - \mu_i(\beta_i^0)\right]^2,
\end{aligned}$$

where $\check{\mu}_i(\beta_i^0)$ lies between $\hat{\mu}_i(\beta_i^0)$ and $\mu_i(\beta_i^0)$. By Assumption A1, Lemma S1.5, and the Markov inequality, one can readily show that the first term is $O_P(T^{-1/2})$ and the second is $O_P(T^{-1})$. It follows that $\hat{S}_i - S_i = O_P(T^{-1/2})$.

(ii) By a simple application of Lemma S1.2(i), $\max_{1\leq i\leq N} \|S_i\| = O_P(T^{-1/2}(\ln T)^3)$. Next,

$$\begin{aligned}
\max_{1\leq i\leq N} \|\hat{S}_i - S_i\| &\leq \max_{1\leq i\leq N} \frac{1}{T}\sum_{t=1}^{T} M(w_{it})\left|\hat{\mu}_i(\beta_i^0) - \mu_i(\beta_i^0)\right| \\
&\leq \left\{\max_{1\leq i\leq N} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[M(w_{it})\right]\right. \\
&\quad \left. + \max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}\left\{M(w_{it}) - \mathbb{E}\left[M(w_{it})\right]\right\}\right|\right\} \\
&\quad \times \max_{1\leq i\leq N}\left|\hat{\mu}_i(\beta_i^0) - \mu_i(\beta_i^0)\right| \\
&= \left\{O(1) + o_P(1)\right\} O_P\left(T^{-1/2}(\ln T)^3\right) \\
&= O_P\left(T^{-1/2}(\ln T)^3\right).
\end{aligned}$$

(iii) By the Cauchy–Schwarz inequality, $\frac{1}{N}\sum_{i=1}^{N} \|\hat{S}_i\|^2 \leq \frac{2}{N}\sum_{i=1}^{N} \|S_i\|^2 + \frac{2}{N}\sum_{i=1}^{N} \|\hat{S}_i - S_i\|^2$. The first term is $O_P(T^{-1})$ by the Markov inequality and the calculation in (i). Using the decomposition of $\hat{S}_i - S_i$ in (i), we can readily show that the second term is $O_P(T^{-1})$. Then $\frac{1}{N}\sum_{i=1}^{N} \|\hat{S}_i\|^2 = O_P(T^{-1})$.

(iv) The result follows by a simple application of Lemma S1.2(ii) and Assumption A2.

(v) The proof is similar to that of (ii) but we now apply Lemmas S1.2(iii) and S1.5(iv). *Q.E.D.*

The next lemma establishes the uniform consistency of $\hat{\beta}_i$.

LEMMA S1.7: *For any $\eta > 0$, we have $P(\max_{1\le i\le N}\|\hat{\beta}_i - \beta_i^0\|) > \eta) = o(N^{-1})$.*

PROOF: Recall that $Q_{1NT,\lambda_1}^{(K_0)}(\boldsymbol{\beta},\boldsymbol{\alpha}) = Q_{1,NT}(\boldsymbol{\beta}) + \frac{\lambda_1}{N}\sum_{i=1}^N\prod_{k=1}^{K_0}\|\beta_i - \alpha_k\|$, where $Q_{1,NT}(\boldsymbol{\beta}) = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\psi(w_{it};\beta_i,\hat{\mu}_i(\beta_i)) = \frac{1}{N}\sum_{i=1}^N\hat{\Psi}_i(\beta_i,\hat{\mu}_i(\beta_i))$. Noting that $(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}}) = \arg\min_{(\boldsymbol{\beta},\boldsymbol{\alpha})} Q_{1NT,\lambda_1}^{(K_0)}(\boldsymbol{\beta},\boldsymbol{\alpha})$, we have $Q_{1NT,\lambda_1}^{(K_0)}(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{\alpha}}) \le Q_{1NT,\lambda_1}^{(K_0)}(\boldsymbol{\beta}^0,\hat{\boldsymbol{\alpha}})$ and

$$\hat{\Psi}_i\big(\hat{\beta}_i,\hat{\mu}_i(\hat{\beta}_i)\big) + \lambda_1\prod_{k=1}^{K_0}\|\hat{\beta}_i - \hat{\alpha}_k\| \le \hat{\Psi}_i\big(\beta_i^0,\hat{\mu}_i\big(\beta_i^0\big)\big) + \lambda_1\prod_{k=1}^{K_0}\big\|\beta_i^0 - \hat{\alpha}_k\big\| \quad \text{for} \quad i = 1,\dots,N.$$

Let $\varepsilon = \min_i[\inf_{\beta_i:\|\beta_i-\beta_i^0\|>\eta}\Psi_i(\beta_i,\mu_i(\beta_i)) - \Psi_i(\beta_i^0,\mu_i(\beta_i^0))] > 0$. Define three events $A_1 \equiv \{\max_{1\le i\le N}\sup_{(\beta,\mu)}|\hat{\Psi}_i(\beta,\mu) - \Psi_i(\beta,\mu)| \le \frac{1}{6}\varepsilon\}$ and $A_2 \equiv \{\max_{1\le i\le N}\sup_{\beta_i}|\Psi_i(\beta_i,\hat{\mu}_i(\beta_i)) - \Psi_i(\beta_i,\mu_i(\beta_i))| \le \frac{1}{6}\varepsilon\}$ and $A_3 \equiv \{\lambda_1\max_{\beta_i:\beta_i\in\mathcal{B}}\prod_{k=1}^{K_0}\|\beta_i - \hat{\alpha}_k\| \le \frac{1}{6}\varepsilon\}$. By Lemmas S1.3, S1.5(v), and Assumption A2(i), $P(A_1\cap A_2\cap A_3) \ge 1 - P(A_1^c) - P(A_2^c) - P(A_3^c) = 1 - o(N^{-1})$. Then conditional on $A_1\cap A_2\cap A_3$, we have, uniformly in i,

$$\begin{aligned}
&\inf_{\beta_i:\|\beta_i-\beta_i^0\|>\eta}\hat{\Psi}_i\big(\beta_i,\hat{\mu}_i(\beta_i)\big) + \lambda_1\prod_{k=1}^{K_0}\|\beta_i - \hat{\alpha}_k\| \\
&\quad\ge \inf_{\beta_i:\|\beta_i-\beta_i^0\|>\eta}\Psi_i\big(\beta_i,\hat{\mu}_i(\beta_i)\big) - \frac{1}{6}\varepsilon + 0 \\
&\quad\ge \inf_{\beta_i:\|\beta_i-\beta_i^0\|>\eta}\Psi_i\big(\beta_i,\mu_i(\beta_i)\big) - \frac{1}{6}\varepsilon - \frac{1}{6}\varepsilon \\
&\quad\ge \Psi_i\big(\beta_i^0,\mu_i\big(\beta_i^0\big)\big) + \varepsilon - \frac{1}{6}\varepsilon - \frac{1}{6}\varepsilon \\
&\quad\ge \Psi_i\big(\beta_i^0,\hat{\mu}_i\big(\beta_i^0\big)\big) - \frac{1}{6}\varepsilon + \varepsilon - \frac{1}{6}\varepsilon - \frac{1}{6}\varepsilon
\end{aligned}$$

$$\begin{aligned}
&\geq \hat{\Psi}_i\big(\beta_i^0, \hat{\mu}_i\big(\beta_i^0\big)\big) - \frac{1}{6}\varepsilon - \frac{1}{6}\varepsilon + \varepsilon - \frac{1}{6}\varepsilon - \frac{1}{6}\varepsilon \\
&= \hat{\Psi}_i\big(\beta_i^0, \hat{\mu}_i\big(\beta_i^0\big)\big) + \frac{1}{3}\varepsilon \\
&\geq \hat{\Psi}_i\big(\beta_i^0, \hat{\mu}_i\big(\beta_i^0\big)\big) + \lambda_1 \prod_{k=1}^{K_0} \big\| \beta_i^0 - \hat{\alpha}_k \big\| + \frac{1}{6}\varepsilon.
\end{aligned}$$

On the other hand, $\hat{\Psi}_i(\hat{\beta}_i, \hat{\mu}_i(\hat{\beta}_i)) + \lambda_1 \prod_{k=1}^{K_0} \|\hat{\beta}_i - \hat{\alpha}_k\| \leq \hat{\Psi}_i(\beta_i^0, \hat{\mu}_i(\beta_i^0)) + \lambda_1 \prod_{k=1}^{K_0} \|\beta_i^0 - \hat{\alpha}_k\|$. It follows that $P(\max_{1\leq i\leq N} \|\hat{\beta}_i - \beta_i^0\|) > \eta) = o(N^{-1})$. *Q.E.D.*

To state and prove the next lemma, we follow Hahn and Newey (2004) and introduce some notation. Let F_i and $\hat{F}_i$ denote the cumulative and empirical distribution functions of w_{it}, respectively. Let $F_i(\epsilon) \equiv F_i + \epsilon\sqrt{T}(\hat{F}_i - F_i)$ for $\epsilon \in [0, T^{-1/2}]$. For fixed β_i and ϵ, let $\mu_i(\beta_i, F_i(\epsilon)) \equiv \arg\min_{\mu_i} \int \psi(\cdot; \beta_i, \mu_i)\, dF_i(\epsilon)$, which is the solution to the estimating equation

$$0 = \int V_i\big(\cdot; \beta_i, \mu_i\big(\beta_i, F_i(\epsilon)\big)\big)\, dF_i(\epsilon). \tag{S4}$$

Define $\mu_i^{\beta_i}(\epsilon) = \partial\mu_i(\beta_i, F_i(\epsilon))/\partial\beta_i$. Apparently, $F_i(0) = F_i$, $F_i(T^{-1/2}) = \hat{F}_i$,

$$\begin{aligned}
&\mu_i(\beta_i) = \mu_i\big(\beta_i, F_i(0)\big), \\
&\hat{\mu}_i(\beta_i) = \mu_i\big(\beta_i, F_i\big(T^{-1/2}\big)\big), \\
&\frac{\partial\mu_i(\beta_i)}{\partial\beta_i} = \frac{\partial\mu_i\big(\beta_i, F_i(0)\big)}{\partial\beta_i} = \mu_i^{\beta_i}(0), \quad \text{and} \\
&\frac{\partial\hat{\mu}_i(\beta_i)}{\partial\beta_i} = \frac{\partial\mu_i\big(\beta_i, F_i\big(T^{-1/2}\big)\big)}{\partial\beta_i} = \mu_i^{\beta_i}\big(T^{-1/2}\big).
\end{aligned}$$

We study the properties of $\mu_i(\beta_i, F_i(\epsilon))$ and $\mu_i^{\beta_i}(\epsilon)$ in the next two lemmas.

LEMMA S1.8: (i) $P(\max_{1\leq i\leq N} \max_{0\leq\epsilon\leq T^{-1/2}} |\mu_i(\beta_i, F_i(\epsilon)) - \mu_i(\beta_i)| \geq \eta) = o_P(N^{-1})$ *for any* $\eta > 0$,

(ii) $\max_{1\leq i\leq N, \max\|\beta_i - \beta_i^0\| = o(1)} |\mu_i(\beta_i) - \mu_i(\beta_i^0)| = o(1)$,

(iii) $P(\max_{1\leq i\leq N, \max\|\beta_i - \beta_i^0\| = o(1)} |\hat{\mu}_i(\beta_i) - \hat{\mu}_i(\beta_i^0)| \geq \eta) = o(N^{-1})$ *for any* $\eta > 0$.

PROOF: (i) Let $\varepsilon = \min_i[\inf_{\mu_i: |\mu_i - \mu_i(\beta_i)| > \eta} \Psi_i(\beta_i, \mu_i) - \Psi_i(\beta_i, \mu_i(\beta_i))] > 0$. Noting that

$$\int \psi(\cdot; \beta_i, \mu_i)\, dF_i(\epsilon) = (1 - \epsilon\sqrt{T})\Psi_i(\beta_i, \mu_i) + \epsilon\sqrt{T}\hat{\Psi}_i(\beta_i, \mu_i),$$

we have

$$\left|\int \psi(\cdot;\beta_i,\mu_i)\,dF_i(\epsilon)-\Psi_i(\beta_i,\mu_i)\right| \le \epsilon\sqrt{T}\left|\hat{\Psi}_i(\beta_i,\mu_i)-\Psi_i(\beta_i,\mu_i)\right| \le \left|\hat{\Psi}_i(\beta_i,\mu_i)-\Psi_i(\beta_i,\mu_i)\right|.$$

By Lemma S1.3, we have $P[A]=o(N^{-1})$,where

$$A\equiv\left\{\max_{0\le\epsilon\le T^{-1/2}}\max_{1\le i\le N}\left|\int \psi(\cdot;\beta_i,\mu_i)\,dF_i(\epsilon)-\Psi_i(\beta_i,\mu_i)\right|\ge\varepsilon/3\right\}.$$

Therefore, for every $\epsilon\in[0,T^{-1/2}]$ and conditional on the event A, we have

$$\begin{aligned}\inf_{\mu_i:|\mu_i-\mu_i(\beta_i)|>\eta}\int \psi(\cdot;\beta_i,\mu_i)\,dF_i(\epsilon) &\ge \inf_{\mu_i:|\mu_i-\mu_i(\beta_i)|>\eta}\Psi_i(\beta_i,\mu_i)-\frac{1}{3}\varepsilon\\ &\ge \Psi_i\big(\beta_i,\mu_i(\beta_i)\big)+\frac{2}{3}\varepsilon\\ &\ge \int \Psi_i\big(\beta_i,\mu_i(\beta_i)\big)\,dF_i(\epsilon)+\frac{1}{3}\varepsilon.\end{aligned}$$

On the other hand, we have $\int\psi(\cdot;\beta_i,\mu_i(\beta_i,F_i(\epsilon)))\,dF_i(\epsilon)\le\int\psi(\cdot;\beta_i,\mu_i(\beta_i))\,dF_i(\epsilon)$ by the definition of $\mu_i(\beta_i,F_i(\epsilon))$. It follows that $P(\max_{1\le i\le N}\max_{0\le\epsilon\le T^{-1/2}}|\mu_i(\beta_i,F_i(\epsilon))-\mu_i(\beta_i)|\ge\eta)=o(N^{-1})$.

(ii) Let $\eta>0$ be given. Let $\varepsilon=\min_i[\inf_{\mu_i:|\mu_i-\mu_i(\beta_i^0)|>\eta}\Psi_i(\beta_i,\mu_i)-\Psi_i(\beta_i^0,\mu_i(\beta_i^0))]>0$ and $\overline{M}=\max_{i,t}\mathbb{E}[M(w_{it})]$. Note that $\max_i|\int[\psi(\cdot;\beta_i,\mu_i(\beta_i^0))-\psi(\cdot;\beta_i^0,\mu_i(\beta_i^0))]\,dF_i|\le\overline{M}\max_i\|\beta_i-\beta_i^0\|=o(1)$, implying that $|\int[\psi(\cdot;\beta_i,\mu_i(\beta_i^0))-\psi(\cdot;\beta_i^0,\mu_i(\beta_i^0))]\,dF_i|\le\varepsilon/3$ when $\max_i\|\beta_i-\beta_i^0\|\le\varepsilon/(3\overline{M})$. Then, for all β_i with $\max\|\beta_i-\beta_i^0\|\le\varepsilon/(3\overline{M})$, we have

$$\begin{aligned}\inf_{\mu_i:|\mu_i-\mu_i(\beta_i^0)|>\eta}\int\psi(\cdot;\beta_i,\mu_i)\,dF_i &\ge \Psi_i\big(\beta_i^0,\mu_i\big(\beta_i^0\big)\big)+\varepsilon\\ &\ge \Psi_i\big(\beta_i,\mu_i\big(\beta_i^0\big)\big)+\frac{2}{3}\varepsilon\\ &= \int\psi_i\big(\cdot;\beta_i,\mu_i\big(\beta_i^0\big)\big)\,dF_i+\frac{2}{3}\varepsilon.\end{aligned}$$

On the other hand, we have $\int\psi(\cdot;\beta_i,\mu_i(\beta_i))\,dF_i\le\int\psi(\cdot;\beta_i,\mu_i(\beta_i^0))\,dF_i$ by the definition of $\mu_i(\beta_i)$. It follows that $\max_{1\le i\le N,\max\|\beta_i-\beta_i^0\|=o(1)}|\mu_i(\beta_i)-\mu_i(\beta_i^0)|=o(1)$.

(iii) By the triangle inequality,

$$\max_{1\le i\le N}\big|\hat{\mu}_i(\beta_i)-\hat{\mu}_i\big(\beta_i^0\big)\big|$$
$$\le \max_{1\le i\le N}\big|\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)\big|+\max_{1\le i\le N}\big|\hat{\mu}_i\big(\beta_i^0\big)-\mu_i\big(\beta_i^0\big)\big|$$
$$+\max_{1\le i\le N}\big|\mu_i(\beta_i)-\mu_i\big(\beta_i^0\big)\big|.$$

By Lemma S1.5(iv), $P(\max_{1\le i\le N}|\hat{\mu}_i(\beta_i)-\mu_i(\beta_i)|\ge\eta/3)=o(N^{-1})$. The last term in the above displayed equation is $o(1)$ uniformly in the set $\max_i\|\beta_i-\beta_i^0\|=o(1)$ by (ii). It follows that $P(\max_{1\le i\le N,\max\|\beta_i-\beta_i^0\|=o(1)}|\hat{\mu}_i(\beta_i)-\hat{\mu}_i(\beta_i^0)|\ge\eta)=o(N^{-1})$ for any $\eta>0$. *Q.E.D.*

LEMMA S1.9: (i) $P(\max_{1\le i\le N}\max_{0\le\epsilon\le T^{-1/2}}\|\frac{\partial\mu_i(\beta_i,F_i(\epsilon))}{\partial\beta_i}-\frac{\partial\mu_i(\beta_i)}{\partial\beta_i}\|\ge\eta)=o(N^{-1})$ *for any* $\eta>0$,

(ii) $\max_{1\le i\le N,\max\|\beta_i-\beta_i^0\|=o(1)}\|\frac{\partial\mu_i(\beta_i)}{\partial\beta_i}-\frac{\partial\mu_i(\beta_i^0)}{\partial\beta_i}\|=o(1)$,

(iii) $P(\max_{1\le i\le N,\max\|\beta_i-\beta_i^0\|=o(1)}\|\frac{\partial\hat{\mu}_i(\beta_i)}{\partial\beta_i}-\frac{\partial\hat{\mu}_i(\beta_i^0)}{\partial\beta_i}\|)=o(N^{-1})$ *for any* $\eta>0$.

PROOF: (i) Differentiating both sides of (S4) with respect to β_i yields

$$0=\int V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i\big(\beta_i,F_i(\epsilon)\big)\big)\,dF_i(\epsilon)$$
$$+\int V_i^{\mu_i}\big(\cdot;\beta_i,\mu_i\big(\beta_i,F_i(\epsilon)\big)\big)\,dF_i(\epsilon)\frac{\partial\mu_i\big(\beta_i,F_i(\epsilon)\big)}{\partial\beta_i}.$$

It follows that

$$\mu_i^{\beta_i}(\epsilon)\equiv\frac{\partial\mu_i\big(\beta_i,F_i(\epsilon)\big)}{\partial\beta_i}=-\frac{\int V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i\big(\beta_i,F_i(\epsilon)\big)\big)\,dF_i(\epsilon)}{\int V_i^{\mu_i}\big(\cdot;\beta_i,\mu_i\big(\beta_i,F_i(\epsilon)\big)\big)\,dF_i(\epsilon)}.\tag{S5}$$

Noting that $\int V_i^{\mu_i}(\cdot;\beta_i,\mu_i(\beta_i))\,dF_i=H_{i\mu\mu}(\beta_i)>c_H>0$ uniformly in i by Assumption A1(v), it suffices to show that

$$P\bigg(\max_{1\le i\le N}\max_{0\le\epsilon\le T^{-1/2}}\bigg\|\int V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i\big(\beta_i,F_i(\epsilon)\big)\big)\,dF_i(\epsilon)$$
$$-\int V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\,dF_i\bigg\|\ge\eta/2\bigg)=o\big(N^{-1}\big),\tag{S6}$$

and

$$
\text{(S7)} \qquad P\bigg(\max_{1\le i\le N}\max_{0\le\epsilon\le T^{-1/2}}\bigg|\int V_i^{\mu_i}\big(\cdot;\beta_i,\mu_i\big(\beta_i,F_i(\epsilon)\big)\big)\,dF_i(\epsilon) - \int V_i^{\mu_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\,dF_i\bigg| \ge \eta/2\bigg) = o\big(N^{-1}\big).
$$

By the triangle inequality,

$$
\begin{aligned}
&\bigg\|\int V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i\big(\beta_i,F_i(\epsilon)\big)\big)\,dF_i(\epsilon) - \int V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\,dF_i\bigg\| \\
&\quad\le \bigg\|\int \big[V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i\big(\beta_i,F_i(\epsilon)\big)\big) - V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\big]\,dF_i(\epsilon)\bigg\| \\
&\qquad + \bigg\|\int V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\,d\big[F_i(\epsilon)-F_i\big]\bigg\| \\
&\quad= \bigg\|\int \big[V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i\big(\beta_i,F_i(\epsilon)\big)\big) - V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\big]\,dF_i(\epsilon)\bigg\| \\
&\qquad + \epsilon\sqrt{T}\bigg\|\int V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\,d[\hat{F}_i - F_i]\bigg\|.
\end{aligned}
$$

Using Lemma S1.2(iii), we have

$$
P\bigg(\max_{1\le i\le N}\max_{0\le\epsilon\le T^{-1/2}}\epsilon\sqrt{T}\bigg\|\int V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\,d[\hat{F}_i - F_i]\bigg\| \ge \eta/4\bigg) = o\big(N^{-1}\big).
$$

In addition, by Lemma S1.8(i),

$$
\begin{aligned}
&P\bigg(\max_{1\le i\le N}\max_{0\le\epsilon\le T^{-1/2}}\bigg\|\int \big[V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i\big(\beta_i,F_i(\epsilon)\big)\big) - V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\big]\,dF_i(\epsilon)\bigg\| \ge \eta/4\bigg) \\
&\quad\le P\bigg(\max_{1\le i\le N}\int M(\cdot)\,dF_i(\epsilon)\max_{1\le i\le N}\max_{0\le\epsilon\le T^{-1/2}}\big|\mu_i\big(\beta_i,F_i(\epsilon)\big) - \mu_i(\beta_i)\big| \ge \eta/4\bigg) = o\big(N^{-1}\big).
\end{aligned}
$$

Then (S6) follows. Analogously, we can prove (S7).

(ii) Recall that

$$\frac{\partial \mu_i(\beta_i)}{\partial \beta_i} = -\frac{\int V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\,dF_i}{\int V_i^{\mu_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\,dF_i}. \tag{S8}$$

To prove (ii), it suffices to show that

$$\max_{1\le i\le N,\max\|\beta_i-\beta_i^0\|=o(1)}\bigg\|\int V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\,dF_i - \int V_i^{\beta_i}\big(\cdot;\beta_i^0,\mu_i\big(\beta_i^0\big)\big)\,dF_i\bigg\| = o(1),$$

and

$$\max_{1\le i\le N,\max\|\beta_i-\beta_i^0\|=o(1)}\bigg\|\int V_i^{\mu_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\,dF_i - \int V_i^{\mu_i}\big(\cdot;\beta_i^0,\mu_i\big(\beta_i^0\big)\big)\,dF_i\bigg\| = o(1).$$

We only show the first result, as the proof of the second one is similar. By Assumption A1(iv) and Lemma S1.8(ii),

$$\begin{aligned}
&\max_{1\le i\le N,\max\|\beta_i-\beta_i^0\|=o(1)}\bigg\|\int V_i^{\beta_i}\big(\cdot;\beta_i,\mu_i(\beta_i)\big)\,dF_i \\
&\qquad - \int V_i^{\beta_i}\big(\cdot;\beta_i^0,\mu_i\big(\beta_i^0\big)\big)\,dF_i\bigg\| \\
&\quad\le \max_{i,t}\mathbb{E}\big[M(w_{it})\big] \\
&\qquad\times \max_{1\le i\le N,\max\|\beta_i-\beta_i^0\|=o(1)}\big\{\big\|\beta_i-\beta_i^0\big\| + \big|\mu_i(\beta_i)-\mu_i\big(\beta_i^0\big)\big|\big\} \\
&\quad= o(1).
\end{aligned}$$

(iii) By the triangle inequality,

$$\begin{aligned}
&\max_{1\leq i\leq N}\left\|\frac{\partial\hat{\mu}_i(\beta_i)}{\partial\beta_i}-\frac{\partial\hat{\mu}_i(\beta_i^0)}{\partial\beta_i}\right\| \\
&\quad\leq \max_{1\leq i\leq N}\left\|\frac{\partial\hat{\mu}_i(\beta_i)}{\partial\beta_i}-\frac{\partial\mu_i(\beta_i)}{\partial\beta_i}\right\|+\max_{1\leq i\leq N}\left\|\frac{\partial\hat{\mu}_i(\beta_i^0)}{\partial\beta_i}-\frac{\partial\mu_i(\beta_i^0)}{\partial\beta_i}\right\| \\
&\qquad+\max_{1\leq i\leq N}\left\|\frac{\partial\mu_i(\beta_i)}{\partial\beta_i}-\frac{\partial\mu_i(\beta_i^0)}{\partial\beta_i}\right\|.
\end{aligned}$$

Noting that $P(\max_{1\leq i\leq N}\|\frac{\partial\hat{\mu}_i(\beta_i)}{\partial\beta_i}-\frac{\partial\mu_i(\beta_i)}{\partial\beta_i}\|\geq\eta/3)=o(N^{-1})$ by (i) and the last term in the above displayed equation is $o(1)$ uniformly in the set $\max_i\|\beta_i-\beta_i^0\|=o_P(1)$ by (ii), we have $P(\max_{1\leq i\leq N,\max\|\beta_i-\beta_i^0\|=o(1)}\|\frac{\partial\hat{\mu}_i(\beta_i)}{\partial\beta_i}-\frac{\partial\hat{\mu}_i(\beta_i^0)}{\partial\beta_i}\|\geq\eta)=o(N^{-1})$ for any $\eta>0$. *Q.E.D.*

Recall from (A.2) that

$$\begin{aligned}
&\hat{H}_{i\beta\beta}(\beta_i) \\
&\quad=\frac{1}{T}\sum_{t=1}^{T}\left[U_i^{\beta_i}\left(w_{it};\beta_i,\hat{\mu}_i(\beta_i)\right)+U_i^{\mu_i}\left(w_{it};\beta_i,\hat{\mu}_i(\beta_i)\right)\frac{\partial\hat{\mu}_i(\beta_i)}{\partial\beta_i'}\right].
\end{aligned}$$

Let $\check{H}_{i\beta\beta}(\beta_i)=\frac{1}{T}\sum_{t=1}^{T}[U_i^{\beta_i}(w_{it};\beta_i,\mu_i(\beta_i))+U_i^{\mu_i}(w_{it};\beta_i,\mu_i(\beta_i))\frac{\partial\mu_i(\beta_i)}{\partial\beta_i'}]$. Note that $H_{i\beta\beta}(\beta_i)=\mathbb{E}[\check{H}_{i\beta\beta}(\beta_i)]$, where $H_{i\beta\beta}(\cdot)$ is defined in Section 2.3. The next lemma study the asymptotics of $\hat{H}_{i\beta\beta}(\beta_i)$.

LEMMA S1.10: (i) $P(\max_{1\leq i\leq N}\|\hat{H}_{i\beta\beta}(\breve{\beta}_i)-H_{i\beta\beta}(\beta_i^0)\|\geq\eta)=o(N^{-1})$.
(ii) $c_{\hat{H}}\equiv\min_{1\leq i\leq N}\mu_{\min}(\hat{H}_{i\beta\beta}(\breve{\beta}_i))\geq c_H-o_P(1)$.

PROOF: (i) By the triangle inequality,

$$\begin{aligned}
&\max_{1\leq i\leq N}\left\|\hat{H}_{i\beta\beta}(\breve{\beta}_i)-H_{i\beta\beta}\left(\beta_i^0\right)\right\| \\
&\quad\leq\max_{1\leq i\leq N}\left\|\hat{H}_{i\beta\beta}(\breve{\beta}_i)-\hat{H}_{i\beta\beta}\left(\beta_i^0\right)\right\|+\max_{1\leq i\leq N}\left\|\hat{H}_{i\beta\beta}\left(\beta_i^0\right)-\check{H}_{i\beta\beta}\left(\beta_i^0\right)\right\| \\
&\qquad+\max_{1\leq i\leq N}\left\|\check{H}_{i\beta\beta}\left(\beta_i^0\right)-H_{i\beta\beta}\left(\beta_i^0\right)\right\|.
\end{aligned}$$

We prove (i) by showing that (i1) $P(\max_{1\leq i\leq N}\|\hat{H}_{i\beta\beta}(\breve{\beta}_i)-\hat{H}_{i\beta\beta}(\beta_i^0)\|\geq\eta/3)=o(N^{-1})$, (i2) $P(\max_{1\leq i\leq N}\|\hat{H}_{i\beta\beta}(\beta_i^0)-\check{H}_{i\beta\beta}(\beta_i^0)\|\geq\eta/3)=o(N^{-1})$, and (i3) $P(\max_{1\le i\le N}\|\check{H}_{i\beta\beta}(\beta_i^0)-H_{i\beta\beta}(\beta_i^0)\|\ge\eta/3)=o(N^{-1})$. For (i1), we make the following decomposition:

$$\begin{aligned}
&\hat{H}_{i\beta\beta}(\breve{\beta}_i)-\hat{H}_{i\beta\beta}\big(\beta_i^0\big)\\
&\quad=\frac{1}{T}\sum_{t=1}^{T}\big[U_i^{\beta_i}\big(w_{it};\breve{\beta}_i,\hat{\mu}_i(\breve{\beta}_i)\big)-U_i^{\beta_i}\big(w_{it};\beta_i^0,\hat{\mu}_i\big(\beta_i^0\big)\big)\big]\\
&\qquad+\frac{1}{T}\sum_{t=1}^{T}\bigg[U_i^{\mu_i}\big(w_{it};\breve{\beta}_i,\hat{\mu}_i(\breve{\beta}_i)\big)\frac{\partial\hat{\mu}_i(\breve{\beta}_i)}{\partial\beta_i'}\\
&\qquad-U_i^{\mu_i}\big(w_{it};\beta_i^0,\hat{\mu}_i\big(\beta_i^0\big)\big)\frac{\partial\hat{\mu}_i\big(\beta_i^0\big)}{\partial\beta_i'}\bigg]\\
&\quad\equiv H_{11i}+H_{12i},\quad\text{say.}
\end{aligned}$$

For H_{11i}, we have

$$\max_{1\le i\le N}\|H_{11i}\|\le\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}M(w_{it})\big\{\big\|\breve{\beta}_i-\beta_i^0\big\|+\big\|\hat{\mu}_i(\breve{\beta}_i)-\hat{\mu}_i\big(\beta_i^0\big)\big\|\big\}.$$

Using the arguments as used in the proof of Lemma S1.5(iv), we can show that

$$P\left(\max_{1\le i\le N}\frac{1}{T}\sum_{t=1}^{T}M(w_{it})\le 2c_M^{1/q}\right)=1-o\big(N^{-1}\big).$$

Then, by Lemmas S1.7 and S1.8(iii), we can readily show that $P(\max_{1\le i\le N}\|H_{11i}\|\ge\eta/6)=o(N^{-1})$. For H_{12i}, we make the following decomposition:

$$\begin{aligned}
H_{12i}&=\frac{1}{T}\sum_{t=1}^{T}\big[U_i^{\mu_i}\big(w_{it};\breve{\beta}_i,\hat{\mu}_i(\breve{\beta}_i)\big)-U_i^{\mu_i}\big(w_{it};\beta_i^0,\hat{\mu}_i\big(\beta_i^0\big)\big)\big]\frac{\partial\hat{\mu}_i(\breve{\beta}_i)}{\partial\beta_i'}\\
&\quad+\frac{1}{T}\sum_{t=1}^{T}U_i^{\mu_i}\big(w_{it};\beta_i^0,\hat{\mu}_i\big(\beta_i^0\big)\big)\left[\frac{\partial\hat{\mu}_i(\breve{\beta}_i)}{\partial\beta_i'}-\frac{\partial\hat{\mu}_i\big(\beta_i^0\big)}{\partial\beta_i'}\right]\\
&\equiv H_{12i,1}+H_{12i,2},\quad\text{say.}
\end{aligned}$$

Following the analysis of H_{11i} and applying Lemma S1.8(i) and (iii) and Lemma S1.9(i) and (iii), we can readily show that $P(\max_{1\le i\le N}\|H_{12i,s}\|\ge\eta/12)=o(N^{-1})$ for $s=1,2$. Then $P(\max_{1\le i\le N}\|H_{12i}\|\ge\eta/6)=o(N^{-1})$. Consequently, we have $P(\max_{1\le i\le N}\|\hat{H}_{i\beta\beta}(\breve{\beta}_i)-\hat{H}_{i\beta\beta}(\beta_i^0)\|\ge\eta/3)=o(N^{-1})$.

To prove (i2), we make the following decomposition:

$$\begin{aligned}
&\hat{H}_{i\beta\beta}\big(\beta_i^0\big) - \check{H}_{i\beta\beta}\big(\beta_i^0\big) \\
&\quad = \frac{1}{T}\sum_{t=1}^{T}\big[U_i^{\beta_i}\big(w_{it}; \beta_i^0, \hat{\mu}_i\big(\beta_i^0\big)\big) - U_i^{\beta_i}\big(w_{it}; \beta_i^0, \mu_i\big(\beta_i^0\big)\big)\big] \\
&\qquad + \frac{1}{T}\sum_{t=1}^{T}\bigg[U_i^{\mu_i}\big(w_{it}; \beta_i^0, \hat{\mu}_i\big(\beta_i^0\big)\big)\frac{\partial \hat{\mu}_i\big(\beta_i^0\big)}{\partial \beta_i'} \\
&\qquad - U_i^{\mu_i}\big(w_{it}; \beta_i^0, \mu_i\big(\beta_i^0\big)\big)\frac{\partial \mu_i\big(\beta_i^0\big)}{\partial \beta_i'}\bigg] \\
&\quad \equiv H_{21i} + H_{22i}.
\end{aligned}$$

Following the analysis of $\max_{1\le i\le N}\|\hat{H}_{i\beta\beta}(\breve{\beta}_i) - \hat{H}_{i\beta\beta}(\beta_i^0)\|$ and using Lemmas S1.2, S1.7, S1.8, and S1.9 and Assumption A1, we can show $P(\max_{1\le i\le N}\|H_{2si}\| \ge \eta/6) = o(N^{-1})$ for $s = 1, 2$. Then (i2) holds.

Next,

$$\begin{aligned}
&\check{H}_{i\beta\beta}\big(\beta_i^0\big) - H_{i\beta\beta}\big(\beta_i^0\big) \\
&\quad = \frac{1}{T}\sum_{t=1}^{T}\big\{U_i^{\beta_i}\big(w_{it}; \beta_i^0, \mu_i\big(\beta_i^0\big)\big) - \mathbb{E}\big[U_i^{\beta_i}\big(w_{it}; \beta_i^0, \mu_i\big(\beta_i^0\big)\big)\big]\big\} \\
&\qquad + \frac{1}{T}\sum_{t=1}^{T}\big\{U_i^{\mu_i}\big(w_{it}; \beta_i^0, \mu_i\big(\beta_i^0\big)\big) - \mathbb{E}\big[U_i^{\mu_i}\big(w_{it}; \beta_i^0, \mu_i\big(\beta_i^0\big)\big)\big]\big\} \\
&\qquad \times \frac{\partial \mu_i\big(\beta_i^0\big)}{\partial \beta_i'} \\
&\quad \equiv H_{31i} + H_{32i}, \quad \text{say.}
\end{aligned}$$

Using Lemma S1.2, we can show $P(\max_{1\le i\le N}\|H_{3si}\| \ge \eta/6) = o(N^{-1})$ for $s = 1, 2$. Then (i3) holds. This completes the proof of (i).

(ii) By the Weyl inequality and the fact that $|\mu_{\max}(A)| \le \|A\|$ for any symmetric matrix A, we have $\mu_{\min}(\hat{H}_{i\beta\beta}(\breve{\beta}_i)) \ge \mu_{\min}(H_{i0}(\beta_i^0)) - \|\hat{H}_{i\beta\beta}(\breve{\beta}_i) - H_{i\beta\beta}(\beta_i^0)\|$. Then by (i) and Assumption A1(v), $c_{\hat{H}} \equiv \min_{1\le i\le N}\mu_{\min}(\hat{H}_{i\beta\beta}(\breve{\beta}_i)) \ge \min_{1\le i\le N}\mu_{\min}(H_{i\beta\beta}(\beta_i^0)) - \max_{1\le i\le N}\|\hat{H}_{i\beta\beta}(\breve{\beta}_i) - H_{i\beta\beta}(\beta_i^0)\| \ge c_H - o_P(1)$. *Q.E.D.*

LEMMA S1.11: *Recall that $\overline{H}_{i\beta\beta} \equiv \hat{H}_{i\beta\beta}(\overline{\beta}_i)$, where $\overline{\beta}_i$ lies between $\hat{\beta}_i$ and β_i^0 elementwise. Then*

(i) $P(\max_{1\le i\le N}\|\overline{H}_{i\beta\beta}-H_{i\beta\beta}(\beta_i^0)\|\ge\eta)=o(N^{-1})$ *for any* $\eta>0$,
(ii) $\max_{1\le i\le N}\|\overline{H}_{i\beta\beta}\|=O_P(1)$.

PROOF: (i) The proof is identical to that of Lemma S1.10(i) with $\breve{\beta}_i$ replaced by $\overline{\beta}_i$.

(ii) By (i) and the triangle inequality,

$$\max_{1\le i\le N}\|\overline{H}_{i\beta\beta}\|\le\max_{1\le i\le N}\big\|H_{i\beta\beta}\big(\beta_i^0\big)\big\|+\max_{1\le i\le N}\big\|\overline{H}_{i\beta\beta}-H_{i\beta\beta}\big(\beta_i^0\big)\big\|$$
$$=O(1)+o_P(1)=O_P(1).\qquad Q.E.D.$$

Recall that $U_{it}=U_i(w_{it};\beta_i^0,\mu_i^0)$, $U_{it}^{\mu_i}=U_i^{\mu_i}(w_{it};\beta_i^0,\mu_i^0)$, $U_{it}^{\mu_i\mu_i}=U_i^{\mu_i\mu_i}(w_{it};\beta_i^0,\mu_i^0)$, and similarly for V_{it} and $V_{it}^{\mu_i}$. Recall that $m_{iU}=\frac{1}{T}\sum_{t=1}^T\mathbb{E}(U_{it}^{\mu_i})$, $m_{iV}=\frac{1}{T}\sum_{t=1}^T\mathbb{E}(V_{it}^{\mu_i})$, $m_{iU2}=\frac{1}{T}\sum_{t=1}^T\mathbb{E}(U_{it}^{\mu_i\mu_i})$, $m_{iV2}=\frac{1}{T}\sum_{t=1}^T\mathbb{E}(V_{it}^{\mu_i\mu_i})$, and $\mathbb{U}_{it}=U_{it}-\frac{m_{iU}}{m_{iV}}V_{it}$. The next two lemmas are essential to establish the asymptotic distribution of the C-Lasso and post-Lasso estimators.

LEMMA S1.12: *Let* $\hat{S}_{\hat{G}_k}\equiv\frac{1}{\sqrt{N_kT}}\sum_{i\in\hat{G}_k}\sum_{t=1}^TU_i(w_{it};\alpha_k^0,\hat{\mu}_i(\alpha_k^0))$. *Then* $\hat{S}_{\hat{G}_k}+\mathbb{B}_{kNT}\xrightarrow{D}N(0,\Omega_k)$.

PROOF: Let $\hat{S}_{G_k^0}\equiv\frac{1}{\sqrt{N_kT}}\sum_{i\in G_k^0}\sum_{t=1}^TU_i(w_{it};\alpha_k^0,\hat{\mu}_i(\alpha_k^0))$. Using the fact that $\mathbf{1}\{i\in\hat{G}_k\}=\mathbf{1}\{i\in G_k^0\}+\mathbf{1}\{i\in\hat{G}_k\setminus G_k^0\}-\mathbf{1}\{i\in G_k^0\setminus\hat{G}_k\}$, we have

$$\hat{S}_{\hat{G}_k}-\hat{S}_{G_k^0}=\frac{1}{\sqrt{N_kT}}\sum_{i\in\hat{G}_k\setminus G_k^0}\sum_{t=1}^TU_i\big(w_{it};\alpha_k^0,\hat{\mu}_i\big(\alpha_k^0\big)\big)$$
$$-\frac{1}{\sqrt{N_kT}}\sum_{i\in G_k^0\setminus\hat{G}_k}\sum_{t=1}^TU_i\big(w_{it};\alpha_k^0,\hat{\mu}_i\big(\alpha_k^0\big)\big)$$
$$\equiv\hat{S}_{k,1}-\hat{S}_{k,2},\quad\text{say.}$$

Let $\epsilon>0$ be an arbitrary constant. By Theorem 2.2, $P(\|\hat{S}_{k,1}\|\ge\epsilon)\le P(\hat{F}_{kNT})\to0$, and $P(\|\hat{S}_{k,2}\|\ge\epsilon)\le P(\hat{E}_{kNT})\to0$. Thus $\hat{S}_{\hat{G}_k}=\hat{S}_{G_k^0}+o_P(1)$ and it suffices to prove the lemma by showing that (i) $\hat{S}_{G_k^0}+\mathbb{B}_{kNT}=\frac{1}{\sqrt{N_kT}}\sum_{i\in G_k^0}\sum_{t=1}^T\mathbb{U}_{it}+o_P(1)$, and (ii) $\frac{1}{\sqrt{N_kT}}\sum_{i\in G_k^0}\sum_{t=1}^T\mathbb{U}_{it}\xrightarrow{D}N(0,\Omega_k)$.

Part (i): We prove $\hat{S}_{G_k^0} + \mathbb{B}_{kNT} = \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^T \mathbb{U}_{it} + o_P(1)$. By second-order Taylor expansion,

$$\begin{aligned}
\text{(S9)} \qquad \hat{S}_{G_k^0} &= \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^T U_{it} + \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^T U_{it}^{\mu_i} \big[\hat{\mu}_i(\alpha_k^0) - \mu_i^0\big] \\
&\quad + \frac{1}{2\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^T U_i^{\mu_i \mu_i}\big(w_{it}; \alpha_k^0, \mu_i^*\big)\big[\hat{\mu}_i(\alpha_k^0) - \mu_i^0\big]^2 \\
&\equiv S_{k,1} + S_{k,2} + S_{k,3}, \quad \text{say},
\end{aligned}$$

where μ_i^* lies between $\hat{\mu}_i(\alpha_k^0)$ and μ_i^0. We will show that $S_{k,1}$ contributes to the asymptotic variance of $\hat{S}_{G_k^0}$, $S_{k,3}$ contributes to the asymptotic bias, and $S_{k,2}$ contributes to both. We analyze $S_{k,3}$ first. Let $S_{k,3}^0 = \frac{1}{2\sqrt{N_k T}} \times \sum_{i \in G_k^0} \sum_{t=1}^T U_{it}^{\mu_i \mu_i} [\hat{\mu}_i(\alpha_k^0) - \mu_i^0]^2$. By Assumption A1, the Markov inequality, and Lemma S1.5(ii), we have

$$\begin{aligned}
\big\|S_{k,3} - S_{k,3}^0\big\| &= \frac{1}{2\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^T \big\| U_i^{\mu_i \mu_i}\big(w_{it}; \alpha_k^0, \mu_i^*\big) \\
&\quad - U_i^{\mu_i \mu_i}\big(w_{it}; \alpha_k^0, \mu_i^0\big)\big\| \big[\hat{\mu}_i(\alpha_k^0) - \mu_i^0\big]^2 \\
&\le \left\{ \frac{1}{2 N_k T} \sum_{i \in G_k^0} \sum_{t=1}^T M(w_{it}) \right\} \sqrt{N_k T} \big|\hat{\mu}_i(\alpha_k^0) - \mu_i^0\big|^3 \\
&= O_P(1) \sqrt{N_k T} O_P\big(T^{-3/2} (\ln T)^9\big) \\
&= O_P\big(N_k^{1/2} T^{-1} (\ln T)^9\big) = o_P(1).
\end{aligned}$$

By (S1) in the proof of Lemma S1.5,

$$S_{k,3}^0 = \frac{1}{2\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^T U_{it}^{\mu_i \mu_i} \left(\frac{\frac{1}{T} \sum_{t=1}^T V_{it}}{\frac{1}{T} \sum_{t=1}^T V_i^{\mu_i}\big(w_{it}; \alpha_k^0, \breve{\mu}_i(\alpha_k^0)\big)} \right)^2.$$

As in the analysis of $S_{k,3} - S_{k,3}^0$, by Lemmas S1.5(ii) and S1.2(i) and the fact that $\max_{1 \le i \le N} |\frac{1}{T} \sum_{t=1}^T V_{it}| = O_P(T^{-1/2} (\ln T)^3)$, we can readily show that $S_{k,3}^0 =$

$S_{k,3}^{00}+O_P(N_k^{1/2}T^{-1}(\ln T)^9)=S_{k,3}^{00}+o_P(1)$, where

$$S_{k,3}^{00}\equiv\frac{1}{2\sqrt{N_kT}}\sum_{i\in G_k^0}\sum_{t=1}^{T}U_{it}^{\mu_i\mu_i}\left(\frac{\frac{1}{T}\sum_{t=1}^{T}V_{it}}{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[V_{it}^{\mu_i}]}\right)^2$$
$$=\frac{1}{2\sqrt{N_kT}}\sum_{i\in G_k^0}m_{iU2}m_{iV}^{-2}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}V_{it}\right)^2+O_P\big(N_k^{1/2}T^{-1}(\ln T)^9\big),$$

where we also use the fact that $\max_{1\le i\le N}\|\frac{1}{T}\sum_{t=1}^{T}U_{it}^{\mu_i\mu_i}-m_{iU2}\|=O_P(T^{-1/2}(\ln T)^3)$ by Lemma S1.2(i). Thus, we have

$$S_{k,3}=\frac{1}{2\sqrt{N_kT}}\sum_{i\in G_k^0}m_{iU2}m_{iV}^{-2}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}V_{it}\right)^2+o_P(1). \tag{S10}$$

Now, we study $S_{k,2}$. By Lemma S1.5(ii), (S1) in its proof, and the fact that $\max_{1\le i\le N}|\frac{1}{T}\sum_{t=1}^{T}V_{it}|=O_P(T^{-1/2}(\ln T)^3)$ and $\max_{1\le i\le N}|\frac{1}{T}\sum_{t=1}^{T}V_{it}^{\mu_i}-m_{iV}|=O_P(T^{-1/2}(\ln T)^3)$, we have

$$\hat{\mu}_i(\alpha_k^0)-\mu_i^0=-\frac{\frac{1}{T}\sum_{t=1}^{T}V_{it}}{\frac{1}{T}\sum_{t=1}^{T}V_i^{\mu_i}\big(w_{it};\alpha_k^0,\breve{\mu}_i(\alpha_k^0)\big)}$$
$$=-\frac{\frac{1}{T}\sum_{t=1}^{T}V_{it}}{\frac{1}{T}\sum_{t=1}^{T}V_{it}^{\mu_i}}+O_P\big(T^{-1}(\ln T)^6\big)$$
$$=-m_{iV}^{-1}\frac{1}{T}\sum_{t=1}^{T}V_{it}+O_P\big(T^{-1}(\ln T)^6\big)\quad\text{uniformly in } i\in G_k^0.$$

But the above expansion is not sufficient to study $S_{k,2}$ and we need to get better control on the remainder term. Noting that $\hat{\mu}_i(\beta_i^0)=\arg\min_{\mu_i}\frac{1}{T}\sum_{t=1}^{T}\psi(w_{it};$

$\beta_i^0, \mu_i)$, we have

$$\begin{aligned} 0 &= \frac{1}{T}\sum_{t=1}^{T} V_i\big(w_{it}; \beta_i^0, \hat{\mu}_i\big(\beta_i^0\big)\big) \\ &= \frac{1}{T}\sum_{t=1}^{T} V_{it} + \frac{1}{T}\sum_{t=1}^{T} V_{it}^{\mu_i}\big[\hat{\mu}_i\big(\beta_i^0\big) - \mu_i\big(\beta_i^0\big)\big] \\ &\quad + \frac{1}{2T}\sum_{t=1}^{T} V_i^{\mu_i\mu_i}\big(w_{it}; \beta_i^0, \check{\mu}_i\big(\beta_i^0\big)\big)\big[\hat{\mu}_i\big(\beta_i^0\big) - \mu_i\big(\beta_i^0\big)\big]^2, \end{aligned}$$

where $\check{\mu}_i(\beta_i^0)$ lies between $\hat{\mu}_i(\beta_i^0)$ and $\mu_i(\beta_i^0)$ for each i. It follows that

$$\begin{aligned} &\hat{\mu}_i\big(\beta_i^0\big) - \mu_{i0}\big(\beta_i^0\big) \\ &\quad = -\left[\frac{1}{T}\sum_{t=1}^{T} V_{it}^{\mu_i}\right]^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T} V_{it} \right.\\ &\qquad \left. + \frac{1}{2T}\sum_{t=1}^{T} V_i^{\mu_i\mu_i}\big(w_{it}; \beta_i^0, \check{\mu}_i\big(\beta_i^0\big)\big)\big[\hat{\mu}_i\big(\beta_i^0\big) - \mu_i\big(\beta_i^0\big)\big]^2\right\} \\ &\quad = -\left[\frac{1}{T}\sum_{t=1}^{T} V_{it}^{\mu_i}\right]^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T} V_{it} + \frac{1}{2T}\sum_{t=1}^{T} V_{it}^{\mu_i\mu_i}\big[\hat{\mu}_i\big(\beta_i^0\big) - \mu_i\big(\beta_i^0\big)\big]^2\right\} \\ &\qquad + O_P\big(T^{-3}(\ln T)^9\big) \\ &\quad = -\left[\frac{1}{T}\sum_{t=1}^{T} V_{it}^{\mu_i}\right]^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T} V_{it} + \frac{1}{2} m_{iV}^{-2}\frac{1}{T}\sum_{t=1}^{T} V_{it}^{\mu_i\mu_i}\left(\frac{1}{T}\sum_{t=1}^{T} V_{it}\right)^2\right\} \\ &\qquad + O_P\big(T^{-3}(\ln T)^9\big) \\ &\quad = -\left[\frac{1}{T}\sum_{t=1}^{T} V_{it}^{\mu_i}\right]^{-1}\left\{\frac{1}{T}\sum_{t=1}^{T} V_{it} + \frac{1}{2} m_{iV}^{-2} m_{iV2}\left(\frac{1}{T}\sum_{t=1}^{T} V_{it}\right)^2\right\} \\ &\qquad + O_P\big(T^{-3}(\ln T)^9\big), \end{aligned} \tag{S11}$$

where we use the fact $\max_{1\le i\le N} |\frac{1}{T}\sum_{t=1}^{T}[V_i^{\mu_i\mu_i}(w_{it}; \beta_i, \check{\mu}_i(\beta_i^0)) - V_{it}^{\mu_i\mu_i}]| \le \max_{1\le i\le N} \frac{1}{T}\sum_{t=1}^{T} M(w_{it}) \times \max_{1\le i\le N} |\check{\mu}_i(\beta_i^0) - \mu_i(\beta_i^0)| = O_P(T^{-1/2}(\ln T)^3)$ and $\max_{1\le i\le N} |\frac{1}{T}\sum_{t=1}^{T} V_{it}^{\mu_i\mu_i} - m_{iV2}| = O_P(T^{-1/2}(\ln T)^3)$ by Lemma S1.2(i). It follows that

$$
\begin{aligned}
S_{k,2} &= \frac{-1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} U_{it}^{\mu_i} \left\{ \left[\frac{1}{T} \sum_{t=1}^{T} V_{it}^{\mu_i} \right]^{-1} \right. \\
&\quad \times \left. \left\{ \frac{1}{T} \sum_{t=1}^{T} V_{it} + \frac{1}{2} m_{iV}^{-2} m_{iV2} \left(\frac{1}{T} \sum_{t=1}^{T} V_{it} \right)^2 \right\} + O_P\big(T^{-3} (\ln T)^9\big) \right\} \\
&= -\frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} V_{it} \frac{\frac{1}{T} \sum_{t=1}^{T} U_{it}^{\mu_i}}{\frac{1}{T} \sum_{t=1}^{T} V_{it}^{\mu_i}} \\
&\quad - \frac{1}{2\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} U_{it}^{\mu_i} \frac{m_{iV}^{-2} m_{iV2} \left(\frac{1}{T} \sum_{t=1}^{T} V_{it} \right)^2}{\frac{1}{T} \sum_{t=1}^{T} V_{it}^{\mu_i}} + o_P(1) \\
&\equiv -S_{k,21} - S_{k,22} + o_P(1).
\end{aligned} \tag{S12}
$$

For $S_{k,21}$, we make the following decomposition:

$$
\begin{aligned}
S_{k,21} &= \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} V_{it} \frac{m_{iU}}{m_{iV}} + \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} V_{it} \frac{\frac{1}{T} \sum_{t=1}^{T} (U_{it}^{\mu_i} - m_{iU})}{m_{iV}} \\
&\quad + \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} V_{it} m_{iU} \left\{ \frac{1}{\frac{1}{T} \sum_{t=1}^{T} V_{it}^{\mu_i}} - \frac{1}{m_{iV}} \right\} \\
&\quad + \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} V_{it} \frac{1}{T} \sum_{t=1}^{T} (U_{it}^{\mu_i} - m_{iU}) \left\{ \frac{1}{\frac{1}{T} \sum_{t=1}^{T} V_{it}^{\mu_i}} - \frac{1}{m_{iV}} \right\} \\
&\equiv S_{k,21a} + S_{k,21b} + S_{k,21c} + S_{k,21d}.
\end{aligned} \tag{S13}
$$

Apparently, $S_{k,21b} = \frac{1}{\sqrt{N_k T^3}} \sum_{i \in G_k^0} m_{iV}^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} V_{is}[U_{it}^{\mu_i} - \mathbb{E}(U_{it}^{\mu_i})]$. For $S_{k,21c}$, we can use the fact that $\max_{1\le i\le N} |\frac{1}{T}\sum_{t=1}^{T} V_{it}| = O_P(T^{-1/2}(\ln T)^3)$ and $\max_{1\le i\le N} |\frac{1}{T}\sum_{t=1}^{T} V_{it}^{\mu_i} - m_{iV}| = O_P(T^{-1/2}(\ln T)^3)$ to show that

$$S_{k,21c} = \frac{-1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} V_{it} m_{iU} \frac{\frac{1}{T}\sum_{t=1}^{T}(V_{it}^{\mu_i} - m_{iV})}{m_{iV}\frac{1}{T}\sum_{t=1}^{T} V_{it}^{\mu_i}}$$

$$= \frac{-1}{\sqrt{N_k T^3}} \sum_{i \in G_k^0} m_{iU} m_{iV}^{-2} \sum_{s=1}^{T} \sum_{t=1}^{T} V_{is}(V_{it}^{\mu_i} - m_{iV}) + o_P(1).$$

It follows that

$$S_{k,21b} + S_{k,21c} = \frac{1}{\sqrt{N_k T^3}} \sum_{i \in G_k^0} m_{iV}^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} V_{is} \mathbb{U}_{it}^{\mu_i} + o_P(1), \tag{S14}$$

where we use the definition of $\mathbb{U}_{it}^{\mu_i} (\equiv U_{it}^{\mu_i} - \frac{m_{iU}}{m_{iV}} V_{it}^{\mu_i})$ and the fact that $\mathbb{E}[\mathbb{U}_{it}^{\mu_i}] = 0$. For $S_{k,21d}$, we can bound it directly:

$$|S_{k,21d}| \le \sqrt{N_k T} \max_{1\le i\le N} \left| \frac{1}{T}\sum_{t=1}^{T} V_{it} \right| \max_{1\le i\le N} \left\| \frac{1}{T}\sum_{t=1}^{T} (U_{it}^{\mu_i} - m_{iU}) \right\| \times \max_{1\le i\le N} \left| \frac{1}{\frac{1}{T}\sum_{t=1}^{T} V_{it}^{\mu_i}} - \frac{1}{m_{iV}} \right| \tag{S15}$$

$$= \sqrt{N_k T} O_P(T^{-3/2}(\ln T)^9) = o_P(1).$$

Combining (S13)–(S15) yields

$$S_{k,21} = \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} V_{it} \frac{m_{iU}}{m_{iV}} + \frac{1}{\sqrt{N_k T^3}} \sum_{i \in G_k^0} m_{iV}^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} V_{is} \mathbb{U}_{it}^{\mu_i} + o_P(1). \tag{S16}$$

In addition, for $S_{k,22}$ we have

$$S_{k,22} = \frac{1}{2\sqrt{N_k T}} \sum_{i \in G_k^0} \left(\frac{1}{T} \sum_{t=1}^{T} U_{it}^{\mu_i} \right) m_{iV}^{-3} m_{iV2} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right)^2 \tag{S17}$$

$$= \frac{1}{2\sqrt{N_k T}} \sum_{i \in G_k^0} m_{iU} m_{iV}^{-3} m_{iV2} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right)^2 + o_P(1).$$

Combining (S12)–(S17) yields

$$\begin{aligned} S_{k,2} = &- \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} V_{it} \frac{m_{iU}}{m_{iV}} \\ &- \left\{ \frac{1}{\sqrt{N_k T^3}} \sum_{i \in G_k^0} m_{iV}^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} V_{is} \mathbb{U}_{it}^{\mu_i} \right. \\ &+ \left. \frac{1}{2\sqrt{N_k T}} \sum_{i \in G_k^0} m_{iU} m_{iV}^{-3} m_{iV2} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right)^2 \right\} + o_P(1). \end{aligned} \tag{S18}$$

Then by (S9), (S10), and (S18),

$$\begin{aligned} \hat{S}_{G_k^0} = &\frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} \left(U_{it} - \frac{m_{iU}}{m_{iV}} V_{it} \right) \\ &- \left\{ \frac{1}{\sqrt{N_k T^3}} \sum_{i \in G_k^0} m_{iV}^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} V_{is} \mathbb{U}_{it}^{\mu_i} \right. \\ &- \left. \frac{1}{2\sqrt{N_k T}} \sum_{i \in G_k^0} m_{iV}^{-2} \left[m_{iU2} - \frac{m_{iV2}}{m_{iV}} m_{iU} \right] \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right)^2 \right\} \\ &+ o_P(1) \\ = &\frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^{T} \mathbb{U}_{it} - \mathbb{B}_{kNT} + o_P(1). \end{aligned}$$

This completes the proof of (i).

Part (ii): We prove $\mathbb{Z}_{NT} \equiv \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^T \mathbb{U}_{it} \xrightarrow{D} N(0, \Omega_k)$. Let $Z_{iT} \equiv \frac{1}{\sqrt{N_k T}} \sum_{t=1}^T \iota_p' \mathbb{U}_{it}$, where ι_p is an arbitrary $p \times 1$ nonrandom vector with $\|\iota_p\| = 1$. Then $\iota_p' \mathbb{Z}_{NT} = \sum_{i \in G_k^0} Z_{iT}$. Noting that $\mathbb{E}(\mathbb{U}_{it}) = 0$, we have, by Assumption A3(i),

$$\operatorname{Var}(\mathbb{Z}_{NT}) = \frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{t=1}^T \sum_{s=1}^T \mathbb{E}(\mathbb{U}_{it} \mathbb{U}_{is}') = \frac{1}{N_k} \sum_{i \in G_k^0} \Omega_{iT} \to \Omega_k > 0,$$

where $\Omega_{iT} = \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \mathbb{E}(\mathbb{U}_{it} \mathbb{U}_{is}')$. By the Lindeberg–Feller central limit theorem (e.g., Theorem 5.6 in White (2001)), it suffices to verify the Lindeberg condition:

$$S_{NT} \equiv \sum_{i \in G_k^0} \mathbb{E}\big[Z_{iT}^2 \mathbf{1}\{|Z_{iT}| > \varepsilon\}\big] \to 0 \quad \text{for any given } \varepsilon > 0.$$

By the Cauchy–Schwarz and Markov inequalities,

$$\begin{aligned} \sum_{i \in G_k^0} \mathbb{E}\big[Z_{iT}^2 \mathbf{1}\{|Z_{iT}| > \varepsilon\}\big] &\le \sum_{i \in G_k^0} \big\{\mathbb{E}(Z_{iT}^4)\big\}^{1/2} \big\{P(|Z_{iT}| > \varepsilon)\big\}^{1/2} \\ &\le \frac{1}{\varepsilon^2} \sum_{i \in G_k^0} \mathbb{E}(Z_{iT}^4). \end{aligned}$$

By straightforward moment conditions and properties of strong mixing processes, we can readily show that

$$\mathbb{E}(Z_{iT}^4) = \frac{1}{N_k^2 T^2} \left(\sum_{t=1}^T \iota_p' \mathbb{U}_{it} \right)^4 = O(N_k^{-2}) \quad \text{uniformly in } i.$$

It follows that $S_{NT} \to 0$ for any $\varepsilon > 0$ and $\mathbb{Z}_{NT} \equiv \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \sum_{t=1}^T \mathbb{U}_{it} \xrightarrow{D} N(0, \Omega_k)$. *Q.E.D.*

REMARK: Note that $\mathbb{U}_i(w_{it}; \beta_i, \mu_i) \equiv U_i(w_{it}; \beta_i, \mu_i) - \frac{m_{iU}}{m_{iV}} V_i(w_{it}; \beta_i, \mu_i)$ and $\mathbb{U}_{it}$ correspond to $U_i(x_{it}; \theta, \gamma_i)$ and U_{it} in Hahn and Kuersteiner (2011, HK hereafter), respectively. Let $\mathbb{U}_i^{\mu_i}$ and $\mathbb{U}_i^{\mu_i \mu_i}$ denote the first and second derivatives of $\mathbb{U}_i$ with respect to μ_i. Let $\mathbb{U}_{it}^{\mu_i} = \mathbb{U}_i^{\mu_i}(w_{it}; \beta_i^0, \mu_i^0)$ and $\mathbb{U}_{it}^{\mu_i \mu_i} = \mathbb{U}_i^{\mu_i \mu_i}(w_{it}; \beta_i^0, \mu_i^0)$. Following HK, the asymptotic bias term of $\hat{S}_{\hat{G}_k}$ takes the form

$$\mathbb{B}_{kNT}^{HK} = \frac{1}{\sqrt{N_k T}} \sum_{i \in G_k^0} \left[m_{iV}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right] \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(\mathbb{U}_{it}^{\mu_i} - \frac{m_{i\mathbb{U}2}}{2m_{iV}} V_{it} \right) \right],$$

where $m_{i\mathbb{U}2} \equiv \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(\mathbb{U}_{it}^{\mu_i \mu_i})$. Note that

$$\begin{aligned}
\mathbb{B}_{kNT}^{HK} &= \frac{1}{\sqrt{N_k T^3}} \sum_{i \in G_k^0} m_{iV}^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} V_{is} \mathbb{U}_{it}^{\mu_i} \\
&\quad - \frac{1}{2\sqrt{N_k T}} \sum_{i \in G_k^0} m_{iV}^{-2} m_{i\mathbb{U}2} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right)^2 \\
&\equiv \mathbb{B}_{1kNT}^{HK} - \mathbb{B}_{2kNT}^{HK}, \quad \text{say.}
\end{aligned}$$

Let $\mathbb{B}_{1kNT}$ and $\mathbb{B}_{2kNT}$ be as defined in Theorem 2.4. Apparently, $\mathbb{B}_{1kNT}^{HK} = \mathbb{B}_{1kNT}$. Noting that $m_{i\mathbb{U}2} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(U_{it}^{\mu_i \mu_i} - \frac{m_{iU}}{m_{iV}} V_i^{\mu_i \mu_i}) = m_{iU2} - \frac{m_{iU}}{m_{iV}} m_{iV2}$ with $m_{iU2} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}(U_{it}^{\mu_i \mu_i})$, we have

$$\mathbb{B}_{2kNT}^{HK} = \frac{1}{2\sqrt{N_k T}} \sum_{i \in G_k^0} m_{iV}^{-2} \left(m_{iU2} - \frac{m_{iU}}{m_{iV}} m_{iV2} \right) \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} V_{it} \right)^2 = \mathbb{B}_{2kNT}.$$

It follows that $\mathbb{B}_{kNT}^{HK} = \mathbb{B}_{kNT}$.

LEMMA S1.13: *Let* $\hat{H}_{(k)} \equiv \frac{1}{N_k T} \sum_{i \in \hat{G}_k} \sum_{t=1}^{T} [U_i^{\beta_i}(w_{it}; \breve{\alpha}_k, \hat{\mu}_i(\breve{\alpha}_k)) + U_i^{\mu_i}(w_{it}; \breve{\alpha}_k^0, \hat{\mu}_i(\breve{\alpha}_k)) \frac{\partial \hat{\mu}_i(\breve{\alpha}_k)}{\partial \alpha_k'}]$ *and* $\breve{\alpha}_k$ *lying between* $\hat{\alpha}_k$ *and* α_k^0 *elementwise. Then* $\hat{H}_{(k)} = \mathbb{H}_{kNT} + o_P(\nu_{NT})$, *where* $\nu_{NT} = \min(1, \sqrt{T/N_k})$.

PROOF: As in the proof of Lemma S1.12, we can readily show that $\hat{H}_{(k)} = \hat{H}_{G_k^0} + o_P(1)$, where $\hat{H}_{G_k^0} = \frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{t=1}^{T} [U_i^{\beta_i}(w_{it}; \breve{\alpha}_k, \hat{\mu}_i(\breve{\alpha}_k)) + U_i^{\mu_i}(w_{it}; \breve{\alpha}_k^0, \hat{\mu}_i(\breve{\alpha}_k)) \frac{\partial \hat{\mu}_i(\breve{\alpha}_k)}{\partial \alpha_k'}]$. For $\hat{H}_{G_k^0}$, we make the following decomposition:

$$\begin{aligned}
\hat{H}_{G_k^0} &\equiv \frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{t=1}^{T} \Big[U_i^{\beta_i}\big(w_{it}; \alpha_k^0, \mu_i\big(\alpha_k^0\big)\big) \\
&\quad + U_i^{\mu_i}(w_{it}; \alpha_k^0, \mu_i\big(\alpha_k^0\big) \frac{\partial \mu_i\big(\alpha_k^0\big)}{\partial \alpha_k'} \Big]
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^{T}\big[U_i^{\beta_i}\big(w_{it};\breve{\alpha}_k,\hat{\mu}_i(\breve{\alpha}_k)\big)-U_i^{\alpha_k}\big(w_{it};\alpha_k^0,\mu_i\big(\alpha_k^0\big)\big)\big]\\
&+\frac{1}{NT}\sum_{i\in\hat{G}_k}\sum_{t=1}^{T}\bigg[U_i^{\mu_i}\big(w_{it};\alpha_k^0,\mu_i\big(\alpha_k^0\big)\big)\frac{\partial\mu_i\big(\alpha_k^0\big)}{\partial\alpha_k'}\\
&-U_i^{\mu_i}\big(w_{it};\breve{\alpha}_k^0,\hat{\mu}_i(\breve{\alpha}_k)\big)\frac{\partial\hat{\mu}_i(\breve{\alpha}_k)}{\partial\alpha_k'}\bigg]\\
\equiv{}& H_{G_k^0,1}+H_{G_k^0,2}+H_{G_k^0,3}.
\end{aligned}
$$

Using the arguments in the proof of Lemma S1.10, we can readily show that $H_{G_k^0,s}=o_P(\nu_{NT})$ for $s=2,3$. In addition, by the Chebyshev inequality we can show that $H_{G_k^0,1}=\overline{H}_{G_k^0,1}+o_P(\nu_{NT})$, where $\overline{H}_{G_k^0,1}=\mathbb{E}[\overline{H}_{G_k^0,1}]$. Then by (S5) with $\epsilon=0$,

$$
\begin{aligned}
\overline{H}_{G_k^0,1}&=\frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^{T}\mathbb{E}\bigg[U_i^{\beta_i}\big(w_{it};\alpha_k^0,\mu_i\big(\alpha_k^0\big)\big)\\
&\qquad+U_i^{\mu_i}\big(w_{it};\alpha_k^0,\mu_i\big(\alpha_k^0\big)\big)\frac{\partial\mu_i\big(\alpha_k^0\big)}{\partial\alpha_k'}\bigg]\\
&=\frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^{T}\mathbb{E}\left[U_{it}^{\beta_i}-U_{it}^{\mu_i}\frac{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big(V_{it}^{\beta_i}\big)'}{m_{iV}}\right]\\
&=\frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^{T}\mathbb{E}\left[U_{it}^{\beta_i}-\frac{m_{iU}}{m_{iV}}V_{it}^{\beta_i\prime}\right]=\frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^{T}\mathbb{E}\big[\mathbb{U}_{it}^{\beta_i}\big]\\
&=\mathbb{H}_{kNT}.
\end{aligned}
$$

It follows that $\hat{H}_{(k)}=\mathbb{H}_{kNT}+o_P(\nu_{NT})$. *Q.E.D.*

REMARK: When $\{\mathbb{U}_{it},t\geq 1\}$ are serially uncorrelated, we have

$$
\Omega_k=\lim_{(N_k,T)\to\infty}\frac{1}{N_k}\sum_{i\in G_k^0}\Omega_{iT}=\lim_{(N_k,T)\to\infty}\frac{1}{N_k}\sum_{i\in G_k^0}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big(\mathbb{U}_{it}\mathbb{U}_{it}'\big).
$$

When the likelihood function is correctly specified, we can apply the second Bartlett identity (i.e., the information matrix equality) to obtain

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[U_{it}U_{it}'] = -\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[U_{it}^{\beta_i}],$$

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[U_{it}V_{it}] = -\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[U_{it}^{\mu_i}] = -m_{iU} = -\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[V_{it}^{\beta_i}],$$

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[V_{it}^2] = -\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[V_{it}^{\mu_i}] = -m_{iV},$$

when $i \in G_k^0$. Then

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\mathbb{U}_{it}\mathbb{U}_{it}') \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left(U_{it} - \frac{m_{iU}}{m_{iV}}V_{it}\right)\left(U_{it} - \frac{m_{iU}}{m_{iV}}V_{it}\right)'\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[U_{it}U_{it}' + \frac{m_{iU}m_{iU}'}{m_{iV}^2}V_{it}^2 - U_{it}V_{it}\frac{m_{iU}'}{m_{iV}} - \frac{m_{iU}}{m_{iV}}U_{it}'V_{it}\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[-U_{it}^{\beta_i} - \frac{m_{iU}m_{iU}'}{m_{iV}} + \frac{m_{iU}m_{iU}'}{m_{iV}} + \frac{m_{iU}m_{iU}'}{m_{iV}}\right] \\
&= -\frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^{T}\mathbb{E}\left[U_{it}^{\beta_i} - \frac{m_{iU}}{m_{iV}}V_{it}^{\beta_i'}\right].
\end{aligned}$$

It follows that

$$\Omega_k = -\lim_{(N_k,T)\to\infty}\frac{1}{N_k}\sum_{i\in G_k^0}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\mathbb{U}_{it}\mathbb{U}_{it}') = -\mathbb{H}_k.$$

LEMMA S1.14: *Suppose the conditions in Theorem* 2.6 *hold*. *Recall that* $\hat{\sigma}^2_{\hat{G}(K,\lambda_1)} = \frac{2}{NT}\sum_{k=1}^{K}\sum_{i\in\hat{G}_k(K,\lambda_1)}\sum_{t=1}^{T}\psi(w_{it}; \hat{\alpha}_{\hat{G}_k(K,\lambda_1)}, \hat{\mu}_i(\hat{\alpha}_{\hat{G}_k(K,\lambda_1)}))$. *Let* $\overline{\sigma}^2_{G^0} = \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\psi(w_{it}; \beta_i^0, \mu_i^0)$. *Then* $\max_{K_0\le K\le K_{\max}}|\hat{\sigma}^2_{\hat{G}(K,\lambda_1)} - \overline{\sigma}^2_{G^0}| = O_P(T^{-1})$.

PROOF: When $K \geq K_0$, following the proof of Theorem 2.1, we can show that $\|\hat{\beta}_i - \beta_i^0\| = O_P(T^{-1/2} + \lambda_1)$ for each i, and

$$\frac{1}{N}\sum_{i=1}^{N}\prod_{k=1}^{K}\|\beta_i^0 - \hat{\alpha}_k\| = \frac{N_1}{N}\prod_{k=1}^{K}\|\hat{\alpha}_k - \alpha_1^0\| + \cdots + \frac{N_{K_0}}{N}\prod_{k=1}^{K}\|\hat{\alpha}_k - \alpha_{K_0}^0\|$$
$$= O_P(T^{-1/2}).$$

Then by Assumption A1(vii), $\prod_{k=1}^{K}\|\hat{\alpha}_k - \alpha_l^0\| = O_P(T^{-1/2})$ for $l = 1, \ldots, K_0$. It follows that the collection $\mathcal{C} \equiv \{\hat{\alpha}_k, k = 1, \ldots, K\}$ contains at least K_0 distinct vectors, say, $\hat{\alpha}_1, \ldots, \hat{\alpha}_{K_0}$, possibly after relabeling the vectors, such that

$$\|\hat{\alpha}_k - \alpha_k^0\|\prod_{l=K_0+1}^{K}\|\hat{\alpha}_l - \alpha_k^0\| = O_P(T^{-1/2}) \quad \text{for} \quad k = 1, \ldots, K_0.$$

As before, we classify $i \in \hat{G}_k(K, \lambda_1)$ if $\|\hat{\beta}_i - \hat{\alpha}_k\| = 0$ for $k = 1, \ldots, K$, and $i \in \hat{G}_0(K, \lambda_1)$ otherwise. Suppose that $i \in \hat{G}_k(K, \lambda_1)$ for $k \in \{K_0 + 1, \ldots, K\}$. Then by the pointwise consistency of $\hat{\beta}_i$, we know that the probability limit of $\hat{\alpha}_k$ must be given by one of the columns in $\boldsymbol{\alpha}^0 = (\alpha_1^0, \ldots, \alpha_{K_0}^0)$ and it converges in probability to the true value at the rate $T^{-1/2} + \lambda_1$. Apparently, if $\mathcal{C}$ contains n_k elements with probability limit given by α_k^0, we can derive that $\|\hat{\alpha}_k - \alpha_k^0\| = O_P(\min(T^{-1/(2n_k)}, T^{-1/2} + \lambda_1))$ for $k = 1, \ldots, K_0$. Without loss of generality, assume that if $n_k > 1$ for $k \in \{1, \ldots, K_0\}$, $\hat{G}_k(K, \lambda_1)$ contains the maximum number of elements among the subsets $\hat{G}_l(K, \lambda_1)$ with $\operatorname{plim}_{(N,T)\to\infty}\hat{\alpha}_l = \alpha_k^0$.

Using arguments like those in the proof of Theorem 2.2, we can show that

$$\text{(S19)} \qquad \sum_{i \in G_k^0} P(\hat{E}_{kNT,i}) = o(1) \quad \text{for} \quad k = 1, \ldots, K_0 \quad \text{and}$$
$$\sum_{i \in \hat{G}_k(K,\lambda_1)} P(\hat{F}_{kNT,i}) = o(1) \quad \text{for} \quad k = 1, \ldots, K_0.$$

The first part implies that $\sum_{i=1}^{N} P(i \in \hat{G}_0(K, \lambda_1) \cup \hat{G}_{K_0+1}(K, \lambda_1) \cup \cdots \cup \hat{G}_K(K, \lambda_1)) = o(1)$.

Let $\hat{\psi}_{it}(k) = 2\psi(w_{it}; \hat{\alpha}_{\hat{G}_k(K,\lambda_1)}(K, \lambda_1), \hat{\mu}_i(\hat{\alpha}_{\hat{G}_k(K,\lambda_1)}))$. Using the fact that $\mathbf{1}\{i \in \hat{G}_k\} = \mathbf{1}\{i \in G_k^0\} + \mathbf{1}\{i \in \hat{G}_k \setminus G_k^0\} - \mathbf{1}\{i \in G_k^0 \setminus \hat{G}_k\}$, we have

$$\hat{\sigma}^2_{\hat{G}(K,\lambda_1)} = \frac{1}{NT}\sum_{k=1}^{K}\sum_{i \in \hat{G}_k(K,\lambda_1)}\sum_{t=1}^{T}\hat{\psi}_{it}(k) = D_{1NT} + D_{2NT} - D_{3NT} + D_{4NT},$$

where

$$D_{1NT} = \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in G_k^0} \sum_{t=1}^{T} \hat{\psi}_{it}(k),$$

$$D_{2NT} = \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in \hat{G}_k(K,\lambda_1) \setminus G_k^0} \sum_{t=1}^{T} \hat{\psi}_{it}(k),$$

$$D_{3NT} = \frac{1}{NT} \sum_{k=1}^{K_0} \sum_{i \in G_k^0 \setminus \hat{G}_k(K,\lambda_1)} \sum_{t=1}^{T} \hat{\psi}_{it}(k), \quad \text{and}$$

$$D_{4NT} = \frac{1}{NT} \sum_{k=K_0+1}^{K} \sum_{i \in \hat{G}_k(K,\lambda_1)} \sum_{t=1}^{T} \hat{\psi}_{it}(k).$$

Let $\delta_{NT} = \min(\sqrt{NT}, T)$. By (S19), we have that for any $\epsilon > 0$, $P(D_{2NT} \geq \delta_{NT}^{-2}\epsilon) \leq \sum_{i=1}^{K_0} P(\hat{F}_{kNT}) \to 0$, $P(D_{3NT} \geq \delta_{NT}^{-2}\epsilon) \leq \sum_{i=1}^{K_0} P(\hat{E}_{kNT}) \to 0$, and $P(D_{4NT} \geq \delta_{NT}^{-2}\epsilon) \leq \sum_{i=1}^{N} P(i \in \bigcup_{K_0+1 \leq k \leq K} \hat{G}_k(K, \lambda_1)) \to 0$. It follows that $\hat{\sigma}^2_{\hat{G}(K,\lambda_1)} = D_{1NT} + o_P(\delta_{NT}^{-2})$ for all $K_0 \leq K \leq K_{\max}$.

Following the proof of Theorem 2.5, we can show that $\hat{\alpha}_{\hat{G}_k(K,\lambda_1)} - \alpha_k^0 = O_P(\delta_{NT}^{-1})$ for $k = 1, \ldots, K_0$. Then by Taylor expansion, we can readily show that

$$\begin{aligned}
D_{1NT} &= \frac{2}{NT} \sum_{k=1}^{K_0} \sum_{i \in G_k^0} \sum_{t=1}^{T} \psi\bigl(w_{it}; \hat{\alpha}_{\hat{G}_k(K,\lambda_1)}(K, \lambda_1), \hat{\mu}_i(\hat{\alpha}_{\hat{G}_k(K,\lambda_1)})\bigr) \\
&= \frac{2}{NT} \sum_{k=1}^{K_0} \sum_{i \in G_k^0} \sum_{t=1}^{T} \psi\bigl(w_{it}; \alpha_k^0, \hat{\mu}_i\bigl(\alpha_k^0\bigr)\bigr) \\
&\quad + \frac{2}{NT} \sum_{k=1}^{K_0} \sum_{i \in G_k^0} \sum_{t=1}^{T} U_i\bigl(w_{it}; \alpha_k^0, \hat{\mu}_i\bigl(\alpha_k^0\bigr)\bigr)\bigl[\hat{\alpha}_{\hat{G}_k(K,\lambda_1)} - \alpha_k^0\bigr] \\
&\quad + \frac{2}{NT} \sum_{k=1}^{K_0} \sum_{i \in G_k^0} \sum_{t=1}^{T} V_i\bigl(w_{it}; \alpha_k^0, \hat{\mu}_i\bigl(\alpha_k^0\bigr)\bigr) \frac{\partial \hat{\mu}_i\bigl(\alpha_k^0\bigr)}{\partial \alpha_k} \bigl[\hat{\alpha}_{\hat{G}_k(K,\lambda_1)} - \alpha_k^0\bigr] \\
&\quad + O_P\bigl(\delta_{NT}^{-2}\bigr) \\
&\equiv D_{1NT,1} + D_{1NT,2} + D_{1NT,3} + O_P\bigl(\delta_{NT}^{-2}\bigr).
\end{aligned}$$

By Lemma S1.12 and the fact that $\mathbb{B}_k = O_P((N/T)^{1/2})$, we can show that $D_{1NT,2} = O_P(\delta_{NT}^{-2})$. Let

$$\overline{D}_{1NT,3} \equiv \frac{1}{NT}\sum_{k=1}^{K_0}\sum_{i\in G_k^0}\sum_{t=1}^{T} V_i\big(w_{it};\alpha_k^0,\hat{\mu}_i\big(\alpha_k^0\big)\big)\frac{\partial\hat{\mu}_i\big(\alpha_k^0\big)}{\partial\alpha_k}.$$

Then

$$\begin{aligned}\overline{D}_{1NT,3} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} V_{it}\frac{\partial\mu_i\big(\beta_i^0\big)}{\partial\beta_i} \\ &\quad + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} V_{it}\left(\frac{\partial\hat{\mu}_i\big(\beta_i^0\big)}{\partial\beta_i} - \frac{\partial\mu_i\big(\beta_i^0\big)}{\partial\beta_i}\right) \\ &\quad + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} V_{it}^{\mu_i}\big(\hat{\mu}_i\big(\beta_i^0\big)-\mu_i^0\big)\frac{\partial\hat{\mu}_i\big(\alpha_k^0\big)}{\partial\alpha_k} + O_P\big(T^{-1}\big) \\ &\equiv \overline{D}_{1NT,31} + \overline{D}_{1NT,32} + \overline{D}_{1NT,33} + O_P\big(T^{-1}\big).\end{aligned}$$

By the Chebyshev and Davydov inequalities, we can readily show that $\overline{D}_{1NT,31} = O_P((NT)^{-1/2})$. By (S5),

$$\begin{aligned}\text{(S20)}\quad &\frac{\partial\hat{\mu}_i\big(\beta_i^0\big)}{\partial\beta_i} - \frac{\partial\mu_i\big(\beta_i^0\big)}{\partial\beta_i} \\ &= \frac{\partial\mu_i\big(\beta_i^0,F_i\big(T^{-1/2}\big)\big)}{\partial\beta_i} - \frac{\partial\mu_i\big(\beta_i^0,F_i(0)\big)}{\partial\beta_i} \\ &= \frac{\int V_i^{\beta_i}\big(\cdot;\beta_i^0,\mu_i^0\big)\,dF_i}{\int V_i^{\mu_i}\big(\cdot;\beta_i^0,\mu_i^0\big)\,dF_i} - \frac{\int V_i^{\beta_i}\big(\cdot;\beta_i^0,\hat{\mu}_i\big(\beta_i^0\big)\big)\,d\hat{F}_i}{\int V_i^{\mu_i}\big(\cdot;\beta_i^0,\hat{\mu}_i\big(\beta_i^0\big)\big)\,d\hat{F}_i} \\ &= \frac{n_{iV}}{m_{iV}} - \frac{\hat{n}_{iV}}{\hat{m}_{iV}} = \frac{n_{iV}\hat{m}_{iV} - \hat{n}_{iV}n_{iV}}{m_{iV}\hat{m}_{iV}} \\ &= \frac{n_{iV}(\hat{m}_{iV}-m_{iV}) + (n_{iV}-\hat{n}_{iV})m_{iV}}{m_{iV}\hat{m}_{iV}},\end{aligned}$$

where $n_{iV} \equiv \int V_i^{\mu_i}(\cdot;\beta_i^0,\mu_i^0)\,dF_i$, $\hat{n}_{iV} \equiv \int V_i^{\beta_i}(\cdot;\beta_i^0,\hat{\mu}_i(\beta_i^0))\,d\hat{F}_i$, $\hat{m}_{iV} \equiv \int V_i^{\mu_i}(\cdot;\beta_i^0,\hat{\mu}_i(\beta_i^0))\,d\hat{F}_i$, and recall $m_{iV} \equiv \int V_i^{\mu_i}(\cdot;\beta_i^0,\mu_i^0)\,dF_i$. Then by (S11) and

Lemma S1.2(i), we can show that

$$\overline{D}_{1NT,32} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} V_{it} m_{iV}^{-2} n_{iV} (\hat{m}_{iV} - m_{iV}) - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} V_{it} m_{iV}^{-1} (\hat{n}_{iV} - n_{iV}) + O_P(\delta_{NT}^{-1}),$$

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N}(\hat{m}_{iV} - m_{iV})^2 &= \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{T=1}^{T}\left[V_i^{\mu_i}\left(w_{it}; \beta_i^0, \hat{\mu}_i\left(\beta_i^0\right)\right) - m_{iV}\right]\right\}^2 \\
&\leq \frac{2}{N}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{T=1}^{T}\left[V_i^{\mu_i}\left(w_{it}; \beta_i^0, \hat{\mu}_i\left(\beta_i^0\right)\right) - V_{it}^{\mu_i}\right]\right\}^2 \\
&\quad + \frac{2}{N}\sum_{i=1}^{N}\left\{\frac{1}{T}\sum_{T=1}^{T}\left(V_{it}^{\mu_i} - m_{iV}\right)\right\}^2 \\
&= O_P\left(T^{-1}\right) + O_P\left(T^{-1}\right) = O_P\left(T^{-1}\right),
\end{aligned}$$

and similarly $\frac{1}{N}\sum_{i=1}^{N}(\hat{n}_{iV} - n_{iV})^2 = O_P(T^{-1})$. Then

$$\begin{aligned}
\|\overline{D}_{1NT,32}\| &\leq \left\{\frac{1}{N}\sum_{i=1}^{N}\left\|m_{iV}^{-2} n_{iV}\right\|\left(\frac{1}{T}\sum_{t=1}^{T} V_{it}\right)^2\right\}^{1/2} \\
&\quad \times \left\{\frac{1}{N}\sum_{i=1}^{N}(\hat{m}_{iV} - m_{iV})^2\right\}^{1/2} \\
&\quad + \left\{\frac{1}{N}\sum_{i=1}^{N}\left|m_{iV}^{-1}\right|\left(\frac{1}{T}\sum_{t=1}^{T} V_{it}\right)^2\right\}^{1/2} \\
&\quad \times \left\{\frac{1}{N}\sum_{i=1}^{N}\|\hat{n}_{iV} - n_{iV}\|^2\right\}^{1/2} \\
&\quad + O_P\left(\delta_{NT}^{-1}\right) \\
&= O_P\left(T^{-1}\right) + O_P\left(T^{-1}\right) + O_P\left(\delta_{NT}^{-1}\right) = O_P\left(\delta_{NT}^{-1}\right).
\end{aligned}$$

For $\overline{D}_{1NT,33}$, using (S11), (S20), and Lemma S1.2(i), we can readily show that

$$\begin{aligned}\overline{D}_{1NT,33} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}V_{it}^{\mu_i}\big(\hat{\mu}_i\big(\beta_i^0\big)-\mu_i^0\big)\frac{\partial\mu_i\big(\alpha_k^0\big)}{\partial\alpha_k}+O_P\big(\delta_{NT}^{-1}\big)\\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}V_{it}\frac{\partial\mu_i\big(\alpha_k^0\big)}{\partial\alpha_k}+O_P\big(\delta_{NT}^{-1}\big)\\ &= O_P\big((NT)^{-1/2}\big)+O_P\big(\delta_{NT}^{-1}\big)=O_P\big(\delta_{NT}^{-1}\big).\end{aligned}$$

Then $\overline{D}_{1NT,3}=O_P(\delta_{NT}^{-1})$ and $D_{1NT,3}=O_P(\delta_{NT}^{-2})$.

By Taylor expansion,

$$\begin{aligned}D_{1NT,1}-\overline{\sigma}_{G^0}^2 &= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\psi\big(w_{it};\beta_i^0,\hat{\mu}_i\big(\beta_i^0\big)\big)-\overline{\sigma}_{G^0}^2\\ &= \frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}V_i\big(w_{it};\beta_i^0,\mu_i^0\big)\big[\hat{\mu}_i\big(\beta_i^0\big)-\mu_i^0\big]\\ &\quad+\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}V_i^{\mu_i}\big(w_{it};\beta_i^0,\check{\mu}_i\big(\beta_i^0\big)\big)\big[\hat{\mu}_i\big(\beta_i^0\big)-\mu_i^0\big]^2\\ &\equiv D_{1NT,11}+D_{1NT,12}.\end{aligned}$$

Using (S11), we can readily show that $D_{1NT,11}=O_P(T^{-1})$ and $D_{1NT,12}=O_P(T^{-1})$. Then $D_{1NT}=\overline{\sigma}_{G^0}^2+O_P(T^{-1})$. It follows that $\hat{\sigma}_{\hat{G}(K,\lambda_1)}^2-\overline{\sigma}_{G^0}^2=O_P(T^{-1})$ for each $K_0\le K\le K_{\max}$. *Q.E.D.*

S2. BIAS CORRECTION IN LINEAR PANEL DATA MODELS

S2.1. *Bias Correction for the PPL C-Lasso Estimator*

For the linear models considered in Section 2.6, the bias of the Lasso and post-Lasso estimator takes the form

$$b_{kNT}=\mathbb{H}_{kNT}^{-1}\mathbb{B}_{kNT}=\mathbb{H}_{kNT}^{-1}\mathbb{B}_{1kNT},$$

where $\mathbb{H}_{kNT}\equiv\frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^{T}\mathbb{E}\{[x_{it}-\mathbb{E}(\overline{x}_{i\cdot})][x_{it}-\mathbb{E}(\overline{x}_{i\cdot})]'\}$ and $\mathbb{B}_{1kNT}=-\frac{1}{N_k^{1/2}T^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\varepsilon_{it}[x_{is}-\mathbb{E}(\overline{x}_{i\cdot})]$. Let $\hat{\varepsilon}_{is}=y_{it}-x_{it}'\hat{\alpha}_{\hat{G}_k}-\hat{\mu}_i$ and $\hat{\mu}_i=$

$\frac{1}{T}\sum_{t=1}^{T}(y_{it}-x_{it}'\hat{\alpha}_{\hat{G}_k})$ for all $i\in\hat{G}_k$.[2] We propose to estimate b_{kNT} by

$$\hat{b}_{kNT}=\widehat{\mathbb{H}}_{kNT}^{-1}\hat{\mathbb{B}}_{1kNT},$$

where $\widehat{\mathbb{H}}_{kNT}=\frac{1}{\hat{N}_kT}\sum_{i\in\hat{G}_k}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}'$ and

$$\hat{\mathbb{B}}_{kNT}=-\frac{1}{N_k^{1/2}T^{3/2}}\sum_{i\in\hat{G}_k}\sum_{s=1}^{T}\sum_{t=1}^{T}k_{M_T}(t,s)x_{is}\hat{\varepsilon}_{it}.$$

Here $k_{M_T}(t,s)=k_{M_T}^0(|t-s|)$ and $k_{M_T}^0(u)$ denotes the Bartlett kernel:

$$k_{M_T}^0(u)=\big(1-|u|/M_T\big)\mathbf{1}\big\{|u|\le M_T\big\}.$$

Dynamic misspecification is permitted here. If the model is dynamically correctly specified in the sense that $\mathbb{E}(\varepsilon_{it}|\mathcal{F}_{i,t-1})=0$ where $\mathcal{F}_{i,t-1}=\sigma(u_{i,t-1},u_{i,t-2},\ldots;x_{it},x_{it-1},\ldots)$, a one-sided kernel can be used with $k_{M_T}(t,s)=k_{M_T}^1(s-t)$, where

$$k_{M_T}^1(u)=(1-u/M_T)\mathbf{1}\{0\le u\le M_T\}.$$

Other choices of kernels are possible. So the bias-corrected PLS C-Lasso estimator is given by

$$\hat{\alpha}_k^{(c)}=\hat{\alpha}_k-\frac{1}{\sqrt{\hat{N}_kT}}\widehat{\mathbb{H}}_{kNT}^{-1}\hat{\mathbb{B}}_{1kNT}.$$

Similarly, we can obtain the bias-corrected estimator for the post-Lasso estimator $\hat{\alpha}_{\hat{G}_k}$.

Let $x_i\equiv(x_{i1},\ldots,x_{iT})'$ and $\varepsilon_i\equiv(\varepsilon_{i1},\ldots,\varepsilon_{iT})'$. Let $\|A\|_a=\{\mathbb{E}\|A\|^a\}^{1/a}$ for any $a\ge 1$. Let C denote a generic positive constant that does not depend on N and T. We add the following assumption.

ASSUMPTION D1: (i) *For each* $i=1,\ldots,N$, $\{(x_{it},\varepsilon_{it}):t=1,2,\ldots\}$ *is strong mixing with mixing coefficients* $\{\alpha_i(\cdot)\}$ *such that* $\alpha_i(\tau)\le c_{\alpha,i}\rho^{\tau}$ *for some* $c_{\alpha,i}<\infty$*and* $\rho\in(0,1)$. $\frac{1}{N_k}\sum_{i\in G_k^0}c_{\alpha,i}^{(2q-1)/q}=O(1)$.

(ii) (x_i,ε_i) *are independent across* $i\in G_k^0$, *where* $k=1,\ldots,K_0$.

(iii) $\max_{i,t}\mathbb{E}\|x_{it}\|^{4q}<C<\infty$ *and* $\max_{i,t}\mathbb{E}\|\varepsilon_{it}\|^{4q}<C<\infty$ *for some* $q\ge 1$.

[2] Observing that $\hat{\alpha}_k-\alpha_k^0=O_P((N_kT)^{-1/2}+T^{-1})$ and $\hat{\alpha}_{\hat{G}_k}-\alpha_k^0=O_P((N_kT)^{-1/2}+T^{-1})$, one can use either estimator in the definition of the residuals. We recommend using the post-Lasso estimator $\hat{\alpha}_{\hat{G}_k}$ because of its better finite-sample performance.

(iv) *As* $(N,T)\to\infty$, $M_T\to\infty$, $M_T^2/T\to 0$, $M_T^2N_k/T^3\to 0$, *and* $N_k^{-1/2}T^{1/2}\sum_{i\in G_k^0}\alpha_i(M_T)^{\frac{2q-1}{2q}}\to 0$ *for each* $k=1,\dots,K_0$.

Assumption D1(i) assumes the usual mixing condition. Assumption D1(ii) assumes cross-sectional independence to simplify the proof which can be relaxed at the cost of lengthy arguments. Assumption D1(iii) assumes moment conditions. The last condition in Assumption D1(iv) can be easily ensured under Assumption D1(i) because for any $M_T\gg -\frac{2q}{(2q-1)\ln q}\ln(N^{1/2}T^{1/2})$ (e.g., $M_T=(\ln(N^{1/2}T^{1/2}))^{1+\epsilon}$ for some $\epsilon>0$), we have

$$\begin{aligned} &N_k^{-1/2}T^{1/2}\sum_{i\in G_k^0}\alpha_i(M_T)^{(2q-1)/(2q)}\\ &\quad\le\left(N_k^{-1}\sum_{i\in G_k^0}c_{\alpha,i}^{(2q-1)/(2q)}\right)N_k^{1/2}T^{1/2}\rho^{M_T(2q-1)/(2q)}\\ &\quad=O(1)\exp\left(\ln\left(N_k^{1/2}T^{1/2}\right)+\frac{(2q-1)M_T}{2q}\ln\rho\right)\to 0. \end{aligned}$$

The first three requirements in Assumption D1(iv) can be easily satisfied, too. For example, if $N_k\propto T^a$ for some $a<3$, it suffices to set $M_T\propto T^{1/b}$ for some $b>\max\{2,2/(3-a)\}$.

PROPOSITION S2.1: *Suppose Assumption* D1 *holds*. *Then* $\widehat{\mathbb{H}}_{kNT}^{-1}\hat{\mathbb{B}}_{1kNT}-\mathbb{H}_{kNT}^{-1}\mathbb{B}_{1kNT}=o_P(1)$.

PROOF: Noting that $\widehat{\mathbb{H}}_{kNT}^{-1}\hat{\mathbb{B}}_{1kNT}-\mathbb{H}_{kNT}^{-1}\mathbb{B}_{1kNT}=(\widehat{\mathbb{H}}_{kNT}^{-1}-\mathbb{H}_k^{-1})\mathbb{B}_{1kNT}+(\widehat{\mathbb{H}}_{kNT}^{-1}-\mathbb{H}_{kNT}^{-1})(\hat{\mathbb{B}}_{1kNT}-\mathbb{B}_{1kNT})+\mathbb{H}_{kNT}^{-1}(\hat{\mathbb{B}}_{1kNT}-\mathbb{B}_{1kNT})$, $\mathbb{H}_{kNT}^{-1}=O(1)$, and $\mathbb{B}_{1kNT}=O_P(\sqrt{N_k/T})$, it suffices to show that (i) $\widehat{\mathbb{H}}_{kNT}-\mathbb{H}_{kNT}=o_P(\nu_{NT})$ and (ii) $\hat{\mathbb{B}}_{1kNT}-\mathbb{B}_{1kNT}=o_P(1)$, where $\nu_{NT}=\min(1,\sqrt{T/N_k})$.

We first prove (i). Let $\overline{\mathbb{H}}_{kNT}\equiv\frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^T\tilde{x}_{it}\tilde{x}_{it}'$. It suffices to show that (i1) $\widehat{\mathbb{H}}_{kNT}-\overline{\mathbb{H}}_{kNT}=o_P(\nu_{NT})$ and (i2) $\overline{\mathbb{H}}_{kNT}-\mathbb{H}_{kNT}=o_P(\nu_{NT})$. Note that

$$\begin{aligned} \widehat{\mathbb{H}}_{kNT}-\overline{\mathbb{H}}_{kNT}&=\frac{1}{\hat{N}_kT}\sum_{i\in\hat{G}_k}\sum_{t=1}^T\tilde{x}_{it}\tilde{x}_{it}'-\frac{1}{N_kT}\sum_{i\in G_k^0}\sum_{t=1}^T\tilde{x}_{it}\tilde{x}_{it}'\\ &=\frac{1}{\hat{N}_kT}\left(\sum_{i\in\hat{G}_k}-\sum_{i\in G_k^0}\right)\sum_{t=1}^T\tilde{x}_{it}\tilde{x}_{it}'+\frac{N_k-\hat{N}_k}{\hat{N}_kN_kT}\sum_{i\in G_k^0}\sum_{t=1}^T\tilde{x}_{it}\tilde{x}_{it}'\\ &\equiv H_{k,1}+H_{k,2}. \end{aligned}$$

By Corollary 2.3, we can readily show that $H_{k,2} = O_P(N_k^{-1}) = o_P(\nu_{NT})$. For any $\epsilon > 0$, we have by the proof of Theorem 2.2, $P(\|H_{k,1}\| \geq \nu_{NT}\epsilon) \leq P(\hat{F}_{kNT}) + P(\hat{E}_{kNT}) = o(1)$. It follows that $\widehat{\mathbb{H}}_{kNT} - \overline{\mathbb{H}}_{kNT} = o_P(\nu_{NT})$. Now,

$$\begin{aligned}
\overline{\mathbb{H}}_{kNT} - \mathbb{H}_{kNT} &= \frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{t=1}^{T} \left\{ \tilde{x}_{it}\tilde{x}_{it}' - \mathbb{E}\left\{ \left[x_{it} - \mathbb{E}(\overline{x}_{i\cdot}) \right] \left[x_{it} - \mathbb{E}(\overline{x}_{i\cdot}) \right]' \right\} \right\} \\
&= \frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{t=1}^{T} \Bigg\{ \left[x_{it} - \mathbb{E}(\overline{x}_{i\cdot}) \right] \left[x_{it} - \mathbb{E}(\overline{x}_{i\cdot}) \right] \\
&\quad - \mathbb{E}\left\{ \left[x_{it} - \mathbb{E}(\overline{x}_{i\cdot}) \right] \left[x_{it} - \mathbb{E}(\overline{x}_{i\cdot}) \right]' \right\} \\
&\quad + \frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{t=1}^{T} \left\{ \tilde{x}_{it}\tilde{x}_{it}' - \left[x_{it} - \mathbb{E}(\overline{x}_{i\cdot}) \right] \left[x_{it} - \mathbb{E}(\overline{x}_{i\cdot}) \right] \right\} \Bigg\} \\
&\equiv H_{k,3} + H_{k,4}.
\end{aligned}$$

Let ω_1 and ω_2 be arbitrary nonrandom $p \times 1$ vectors such that $\|\omega_1\| = \|\omega_2\| = 1$. By Assumption D1(i)–(ii) and the Davydov inequality, we can readily show that

$$\begin{aligned}
\mathbb{E}\left[\left(\omega_1' H_{k,3} \omega_2 \right)^2 \right] &= \operatorname{Var}\left(\omega_1' H_{k,3} \omega_2 \right) \\
&= \frac{1}{(N_k T)^2} \sum_{i \in G_k^0} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}(\zeta_{1,it}\zeta_{1,is}) = O\left((N_k T)^{-1} \right),
\end{aligned}$$

where $\zeta_{1,it} = \omega_1'[x_{it} - \mathbb{E}(\overline{x}_{i\cdot})][x_{it} - \mathbb{E}(\overline{x}_{i\cdot})] - \mathbb{E}\{[x_{it} - \mathbb{E}(\overline{x}_{i\cdot})][x_{it} - \mathbb{E}(\overline{x}_{i\cdot})]'\}\omega_2$. It follows that $H_{k,3} = O_P((N_k T)^{-1/2})$. Note that

$$\begin{aligned}
\omega_1' H_{k,4} \omega_2 &= \frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{t=1}^{T} \omega_1' x_{it} \left[\mathbb{E}(\overline{x}_{i\cdot}) - \overline{x}_{i\cdot} \right]' \omega_2 \\
&= \frac{-1}{N_k T^2} \sum_{i \in G_k^0} \sum_{t=1}^{T} \sum_{t=1}^{T} \zeta_{2,its},
\end{aligned}$$

where $\zeta_{2,its} = \omega_1'[x_{it} - \mathbb{E}(x_{it})][x_{is} - \mathbb{E}(x_{is})]'\omega_2$. By Assumption D1(i)–(ii) and Lemma A.2(ii) in Gao (2007), we have

$$\mathbb{E}\big[\big(\omega_1' H_{k,4}\omega_2\big)^2\big] = \mathrm{Var}\big(\omega_1' H_{k,4}\omega_2\big) = \frac{1}{N_k^2 T^4}\sum_{i\in G_k^0}\mathbb{E}\Bigg[\Bigg(\sum_{t=1}^{T}\sum_{s=1}^{T}\zeta_{2,its}\Bigg)^2\Bigg] = O\big(N_k^{-1}T^{-2}\big).$$

It follows that $H_{k,3} = O_P(N_k^{-1/2}T^{-1})$. Consequently, $\overline{\mathbb{H}}_{kNT} - \mathbb{H}_{kNT} = O_P((N_k T)^{-1/2}) = o_P(\nu_{NT})$. This completes the proof of (i).

We now prove (ii). Let

$$\overline{\mathbb{B}}_{1kNT} = \mathbb{E}(\mathbb{B}_{1kNT}) = \frac{-1}{N_k^{1/2}T^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}(\varepsilon_{it}x_{is}).$$

We prove (ii) by showing that (ii1) $\mathbb{B}_{1kNT} - \overline{\mathbb{B}}_{1kNT} = o_P(1)$, and (ii2) $\hat{\mathbb{B}}_{1kNT} - \overline{\mathbb{B}}_{1kNT} = o_P(1)$. For (ii1), we have

$$\begin{aligned}
&\omega_1'(\mathbb{B}_{1kNT} - \overline{\mathbb{B}}_{1kNT}) \\
&\quad = \frac{-1}{N_k^{1/2}T^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\omega_1'\big\{\varepsilon_{it}\big[x_{is} - \mathbb{E}(\overline{x}_{i\cdot})\big] - \mathbb{E}(\varepsilon_{it}x_{is})\big\} \\
&\quad = \frac{-1}{N_k^{1/2}T^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\zeta_{3,its},
\end{aligned}$$

where $\zeta_{3,its} = \omega_1'\{\varepsilon_{it}[x_{is} - \mathbb{E}(\overline{x}_{i\cdot})] - \mathbb{E}(\varepsilon_{it}x_{is})\}$. Then by Assumption D1(i)–(ii) and the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\mathbb{E}\big[\omega_1'(\mathbb{B}_{1kNT} - \overline{\mathbb{B}}_{1kNT})\big]^2 \\
&\quad = \mathrm{Var}\big(\omega_1'(\mathbb{B}_{1kNT} - \overline{\mathbb{B}}_{1kNT})\big) = \frac{1}{N_k T^3}\sum_{i\in G_k^0}\mathbb{E}\Bigg(\sum_{s=1}^{T}\sum_{t=1}^{T}\zeta_{3,its}\Bigg)^2 \\
&\quad \le \frac{2}{N_k T^3}\sum_{i\in G_k^0}\mathbb{E}\Bigg(\sum_{s=1}^{T}\sum_{t=1}^{T}\omega_1'\varepsilon_{it}\big[x_{is} - \mathbb{E}(\overline{x}_{i\cdot})\big]\Bigg)^2 \\
&\qquad + \frac{2}{N_k T^3}\sum_{i\in G_k^0}\Bigg(\sum_{s=1}^{T}\sum_{t=1}^{T}\omega_1'\mathbb{E}(\varepsilon_{it}x_{is})\Bigg)^2.
\end{aligned}$$

For the first term, we can apply Lemma A.2(ii) in Gao (2007) and show that it is $O(T^{-1})$. For the second term, we can apply the Davydov inequality directly to show that it is bounded from above by

$$\frac{2}{N_k T^3}\sum_{i\in G_k^0}\left(8T\|\varepsilon_{it}\|_{4q}\left\|\omega_1' x_{is}\right\|_{4q}\sum_{s=1}^{T}\alpha_i(s)^{(2q-1)/(2q)}\right)^2 = O\left(T^{-1}\right).$$

It follows that $\mathbb{B}_{1kNT} - \overline{\mathbb{B}}_{1kNT} = O(T^{-1/2}) = o_P(1)$.

We now show (ii2). We first make the following decomposition:

$$\begin{aligned}
\overline{\mathbb{B}}_{1kNT} - \hat{\mathbb{B}}_{1kNT} &= \frac{1}{\hat{N}_k^{1/2}T^{3/2}}\sum_{i\in\hat{G}_k}\sum_{s=1}^{T}\sum_{t=1}^{T}k_{M_T}(t,s)x_{is}\hat{\varepsilon}_{it} \\
&\quad - \frac{1}{N_k^{1/2}T^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}(\varepsilon_{it}x_{is}) \\
&= \frac{1}{\hat{N}_k^{1/2}T^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}k_{M_T}(t,s)x_{is}\hat{\varepsilon}_{it} \\
&\quad - \frac{1}{N_k^{1/2}T^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\mathbb{E}(\varepsilon_{it}x_{is}) + o_P(1) \\
&= \frac{1}{\hat{N}_k^{1/2}T^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}k_{M_T}(t,s)x_{is}(\hat{\varepsilon}_{it}-\varepsilon_{it}) \\
&\quad + \frac{1}{\hat{N}_k^{1/2}T^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}k_{M_T}(t,s)\left[x_{is}\varepsilon_{it} - \mathbb{E}(x_{is}\varepsilon_{it})\right] \\
&\quad + \frac{N_k^{-1/2}-\hat{N}_k^{-1/2}}{T^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}k_{M_T}(t,s)\mathbb{E}(x_{is}\varepsilon_{it}) \\
&\quad + \frac{1}{N_k^{1/2}T^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\left[1-k_{M_T}(t,s)\right]\mathbb{E}(x_{is}\varepsilon_{it}) \\
&\quad + o_P(1) \\
&\equiv \hat{B}_{kNT,1} + \hat{B}_{kNT,2} + \hat{B}_{kNT,3} + \hat{B}_{kNT,4} + o_P(1),
\end{aligned}$$

where the $o_P(1)$ term arises due to the replacement of $\hat{G}_k$ by G_k^0 and this can be easily justified by using the uniform classification consistency result and arguments as used in the proof of Theorem 2.5. We prove (ii) by demonstrating that $\hat{B}_{kNT,s} = o_P(1)$ for $s = 1, 2, 3,$ and 4.

We first study $\hat{B}_{kNT,1}$. Noting that $\hat{\varepsilon}_{it} = y_{it} - x_{it}'\hat{\alpha}_{\hat{G}_k} - \hat{\mu}_i = y_{it} - x_{it}'\hat{\alpha}_{\hat{G}_k} - \frac{1}{T}\sum_{t=1}^T (y_{it} - x_{it}'\hat{\alpha}_{\hat{G}_k})$ and $y_{it} = x_{it}'\alpha_k^0 + \mu_i + \varepsilon_{it}$ for $i \in G_k^0$, we have that, for $i \in G_k^0$,

$$\hat{\varepsilon}_{it} - \varepsilon_{it} = y_{it} - x_{it}'\hat{\alpha}_{\hat{G}_k} - \frac{1}{T}\sum_{t=1}^T \big(y_{it} - x_{it}'\hat{\alpha}_{\hat{G}_k}\big) - \varepsilon_{it} = \tilde{x}_{it}'\big(\alpha_k^0 - \hat{\alpha}_{\hat{G}_k}\big) - \overline{\varepsilon}_{i\cdot},$$

where $\overline{\varepsilon}_{i\cdot} = \frac{1}{T}\sum_{t=1}^T \varepsilon_{it}$. Then

$$\begin{aligned}
\hat{B}_{kNT,1} &= \frac{1}{\hat{N}_k^{1/2} T^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^T \sum_{t=1}^T k_{M_T}(t,s) x_{is} \tilde{x}_{it}' \big(\alpha_k^0 - \hat{\alpha}_{\hat{G}_k}\big) \\
&\quad - \frac{1}{\hat{N}_k^{1/2} T^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^T \sum_{t=1}^T k_{M_T}(t,s) x_{is} \overline{\varepsilon}_{i\cdot} \\
&\equiv B_{kNT,1}(1) - B_{kNT,1}(2), \quad \text{say.}
\end{aligned}$$

In view of the fact that $\hat{\alpha}_{\hat{G}_k} - \alpha_k^0 = O_P((N_k T)^{-1/2} + T^{-1})$ and $\hat{N}_k = N_k(1 + o_P(1))$, we have

$$\begin{aligned}
\big\| B_{kNT,1}(1) \big\| &= \frac{1}{\hat{N}_k^{1/2} T^{3/2}} \left\| \sum_{i\in G_k^0} \sum_{s=1}^T \sum_{t=1}^T k_{M_T}(t,s) x_{is} \tilde{x}_{it}' \big(\alpha_k^0 - \hat{\alpha}_{\hat{G}_k}\big) \right\| \\
&\le \frac{N_k T^{1/2}}{\hat{N}_k^{1/2}} \big\|\alpha_k^0 - \hat{\alpha}_{\hat{G}_k}\big\| \frac{1}{N_k T^2} \sum_{i\in G_k^0} \sum_{|s-t|\le M_T} \big\| x_{is}\tilde{x}_{it}' \big\| \\
&= N_k^{1/2} T^{1/2} O_P\big((N_k T)^{-1/2} + T^{-1}\big) O_P(M_T/T) \\
&= O_P\big(1 + N_k^{1/2} T^{-1/2}\big) O_P(M_T/T) = o_P(1),
\end{aligned}$$

where we use the fact that $\frac{1}{N_k T^2}\sum_{i\in G_k^0}\sum_{|s-t|\le M_T} \|x_{is}\tilde{x}_{it}'\| = O_P(M_T/T)$ by moment calculation and the Markov inequality. Let $\overline{B}_{kNT,1}(2) \equiv \frac{1}{N_k^{1/2} T^{3/2}} \times \sum_{i\in G_k^0}\sum_{s=1}^T \sum_{t=1}^T k_{M_T}(t,s)\omega' x_{is}\overline{\varepsilon}_{i\cdot}$, where ω is any $p\times 1$ nonrandom vector such that $\|\omega\| = 1$. Then by Assumption D1(i), (iii), and (iv),

$$\begin{aligned}
\left|\mathbb{E}\left[\overline{B}_{kNT,1}(2)\right]\right| &\leq \frac{1}{N_k^{1/2}T^{5/2}} \sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{r=1}^{T} k_{M_T}(t,s)\left|\mathbb{E}\left(\omega' x_{is}\varepsilon_{ir}\right)\right| \\
&\leq \frac{8}{N_k^{1/2}T^{5/2}} \sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{r=1}^{T} k_{M_T}(t,s) \\
&\quad \times \left\|\omega' x_{is}\right\|_{4q} \|\varepsilon_{ir}\|_{4q} \alpha_i\left(|r-s|\right)^{(2q-1)/(2q)} \\
&\leq \frac{CN_k^{1/2}}{T^{3/2}} \left\{\frac{1}{N_k}\sum_{i\in G_k^0} c_{\alpha,i}^{(2q-1)/(2q)}\right\} \\
&\quad \times \left\{\frac{1}{T}\sum_{t,s,r:\ |s-t|\leq M_T} \rho^{|r-s|(2q-1)/(2q)}\right\} \\
&= N_k^{1/2}T^{-3/2}O(1)O(M_T) = O\left(M_T N_k^{1/2} T^{-3/2}\right) = o(1).
\end{aligned}$$

Similarly, by Assumption D1(i)–(iv),

$$\begin{aligned}
\operatorname{Var}\left(\overline{B}_{kNT,1}(2)\right) &= \frac{1}{N_k T^5}\sum_{i\in G_k^0} \operatorname{Var}\left(\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{r=1}^{T} k_{M_T}(t,s)\omega' x_{is}\varepsilon_{ir}\right) \\
&\leq \frac{1}{N_k T^5}\sum_{i\in G_k^0} \mathbb{E}\left[\left(\sum_{s=1}^{T}\sum_{t=1}^{T}\sum_{r=1}^{T} k_{M_T}(t,s)\omega' x_{is}\varepsilon_{ir}\right)^2\right] \\
&= \frac{1}{N_k T^5}\sum_{i\in G_k^0} \sum_{1\leq t_1,t_2,\ldots,t_6\leq T} k_{M_T}(t_1,t_2) \\
&\quad \times k_{M_T}(t_4,t_5)\mathbb{E}\left(\omega' x_{it_2}\varepsilon_{it_3}\ \omega' x_{it_5}\varepsilon_{it_6}\right) \\
&\leq \frac{1}{N_k T^5}\sum_{i\in G_k^0} \sum_{\substack{1\leq t_1,t_2,\ldots,t_6\leq T \\ |t_1-t_2|\leq M_T, |t_4-t_5|\leq M_T}} \left|\mathbb{E}\left(\omega' x_{it_2}\varepsilon_{it_3}\ \omega' x_{it_5}\varepsilon_{it_6}\right)\right| \\
&= O\left(M_T^2/T\right) = o(1).
\end{aligned}$$

Consequently, $\overline{B}_{kNT,1}(2) = o_P(1)$. This, in conjunction with Corollary 2.3, implies that $B_{kNT,1}(2) = o_P(1)$ as ω is arbitrary. Thus we have shown that $\hat{B}_{kNT,1} = o_P(1)$.

For $\hat{B}_{kNT,2}$, note that $\hat{B}_{kNT,2} = \overline{B}_{kNT,2} N_k^{1/2}/\tilde{N}_k^{1/2} = \overline{B}_{kNT,2}(1 + o_P(1))$, where $\overline{B}_{kNT,2} = \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^T \sum_{t=1}^T k_{M_T}(t, s)[x_{is}\varepsilon_{it} - \mathbb{E}(x_{is}\varepsilon_{it})]$. By construction, $\mathbb{E}(\overline{B}_{kNT,2}) = 0$. By Assumption D1(ii)–(iii) and the Jensen inequality,

$$
\begin{aligned}
&\operatorname{Var}(\omega' \overline{B}_{kNT,2}) \\
&\quad = \frac{1}{N_k T^3} \sum_{i \in G_k^0} \operatorname{Var}\left[\sum_{s=1}^T \sum_{t=1}^T k_{M_T}(t, s)\omega'\left[x_{is}\varepsilon_{it} - \mathbb{E}(x_{is}\Delta\varepsilon_{it})\right]\right] \\
&\quad \le \frac{1}{N_k T^3} \sum_{i \in G_k^0} \sum_{s=1}^T \sum_{t=1}^T \sum_{r=1}^T \sum_{l=1}^T k_{M_T}(t, s) k_{M_T}(r, l) \mathbb{E}(\omega' x_{is}\varepsilon_{it}\varepsilon_{ir} x_{il}\omega) \\
&\quad \le \frac{1}{N_k T^3} \sum_{i \in G_k^0} \sum_{|s-t| \le M_T} \sum_{|r-l| \le M_T} \left|\mathbb{E}(\omega' x_{is}\varepsilon_{it}\varepsilon_{ir} x_{il}\omega)\right| \\
&\quad = O(M_T^2/T) = o(1),
\end{aligned}
$$

where the last equality follows from the fact that $\|\mathbb{E}(\omega' x_{is}\varepsilon_{it}\varepsilon_{ir}x_{il}\omega)\| \le \max_{i,s,t} \|x_{is}\varepsilon_{it}\|_2^2 \le \max_{i,t} \|x_{it}\|_4^2 \times \max_{i,t} \|\varepsilon_{it}\|_4^2 < C < \infty$ by Assumption D1(iii). Then $\overline{B}_{kNT,2} = o_P(1)$ by the Chebyshev inequality and thus $\hat{B}_{kNT,2} = o_P(1)$.

By Corollary 2.3 and the Davydov inequality,

$$
\begin{aligned}
\|\hat{B}_{kNT,3}\| &= \frac{\left|N_k^{-1} - \hat{N}_k^{-1}\right|}{T^{3/2}\left(N_k^{-1/2} + \tilde{N}_k^{-1/2}\right)} \left\| \sum_{i \in G_k^0} \sum_{s=1}^T \sum_{t=1}^T k_{M_T}(t, s)\mathbb{E}(x_{is}\varepsilon_{it}) \right\| \\
&\le \frac{|\hat{N}_k - N_k|}{T^{1/2}\hat{N}_k\left(N_k^{-1/2} + \hat{N}_k^{-1/2}\right)} \left\{ \frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{|s-t| \le M_T} \left\|\mathbb{E}(x_{is}\varepsilon_{it})\right\| \right\} \\
&= o_P\left(N_k^{-1/2} T^{-1/2}\right) O(1) = o_P(1).
\end{aligned}
$$

By Assumption D1(i)–(iv) and the Davydov inequality,

$$
\begin{aligned}
\|\hat{B}_{kNT,4}\| &= \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^T \sum_{t=1}^T \left[1 - k_{M_T}(t, s)\right]\mathbb{E}(x_{is}\varepsilon_{it}) \\
&= \left\| \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^T \sum_{t=1}^T \left[1 - k_{M_T}(t, s)\right]\mathbb{E}(x_{is}\varepsilon_{it}) \right\|
\end{aligned}
$$

$$\leq \frac{8}{N_k^{1/2}T^{3/2}} \sum_{i\in G_k^0} \sum_{|s-t|>M_T} \alpha_i\big(|s-t|\big)^{(2q-1)/(2q)} \|x_{is}\|_{4q} \|\varepsilon_{it}\|_{4q}$$

$$\leq CN_k^{-1/2}T^{1/2} \sum_{i\in G_k^0} \alpha_i(M_T)^{(2q-1)/(2q)} = o(1).$$

This completes the proof of the proposition. *Q.E.D.*

With the above result in hand, we can readily show that

$$\begin{aligned}
\sqrt{N_k T}\big(\hat{\alpha}_k^{(c)} - \alpha_k^0\big) &= \big[\sqrt{N_k T}\big(\hat{\alpha}_k - \alpha_k^0\big) - \mathbb{H}_{kNT}^{-1}\mathbb{B}_{1kNT}\big] \\
&\quad + (N_k/\hat{N}_k)^{1/2}\big[\mathbb{H}_{kNT}^{-1}\mathbb{B}_{1kNT} - \widehat{\mathbb{H}}_{kNT}^{-1}\hat{\mathbb{B}}_{1kNT}\big] \\
&\quad + \big[1 - (N_k/\hat{N}_k)^{1/2}\big]\mathbb{H}_{kNT}^{-1}\mathbb{B}_{1kNT} \\
&= \big[\sqrt{N_k T}\big(\hat{\alpha}_k - \alpha_k^0\big) - \mathbb{H}_{kNT}^{-1}\mathbb{B}_{1kNT}\big] \\
&\quad + o_P(1) + o_P\big(N_k^{-1}\big)O\big((N_K/T)^{1/2}\big) \\
&= \big[\sqrt{N_k T}\big(\hat{\alpha}_k - \alpha_k^0\big) - \mathbb{H}_{kNT}^{-1}\mathbb{B}_{1kNT}\big] + o_P(1).
\end{aligned}$$

That is, $\sqrt{N_k T}(\hat{\alpha}_k^{(c)} - \alpha_k^0)$ has the desired limiting distribution centered on the origin.

S2.2. *Bias Correction for the PGMM C-Lasso Estimator*

Bias correction for the PGMM C-Lasso estimator in dynamic panel data models can be done analogously. For simplicity, we focus on the case where $W_{iNT} = I_d$ for all i. Recall from Theorem 3.4 and the remark regarding Assumption B3(iii) (see (3.3) in particular) that

$$\begin{aligned}
&\sqrt{N_k T}\big(\tilde{\alpha}_k - \alpha_k^0\big) - \overline{A}_k^{-1} B_{kNT} \\
&\quad \xrightarrow{D} N\big(0, A_k^{-1}C_k A_k^{-1}\big) \quad \text{for} \quad k = 1, \ldots, K_0,
\end{aligned}$$

where $\overline{A}_k \equiv \frac{1}{N_k}\sum_{i\in G_k^0} \overline{Q}_{i,z\Delta x}' \overline{Q}_{i,z\Delta x}$ and

$$B_{kNT} = \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E}\big(\Delta x_{is} z_{is}' z_{it} \Delta\varepsilon_{it}\big).$$

Based on (3.3), in order to verify Assumption B3(iii) we also need to show

$$\text{(S21)}\qquad V_{kNT}=\frac{1}{N_k^{1/2}T^{1/2}}\sum_{i\in G_k^0}\sum_{t=1}^{T}\overline{Q}'_{i,z\Delta x}z_{it}\Delta\varepsilon_{it}\xrightarrow{D}N(0,C_k),\quad \text{and}$$

$$\begin{aligned}\text{(S22)}\qquad R_{kNT}&=\frac{1}{N_k^{1/2}T^{3/2}}\sum_{i\in G_k^0}\sum_{s=1}^{T}\sum_{t=1}^{T}\big\{\big[\Delta x_{is}z'_{is}-\mathbb{E}\big(\Delta x_{is}z'_{is}\big)\big]z_{it}\Delta\varepsilon_{it}\\&\quad-\mathbb{E}\big(\Delta x_{is}z'_{is}z_{it}\Delta\varepsilon_{it}\big)\big\}\\&=o_P(1).\end{aligned}$$

The first part is assured by a version of the CLT. Below, we first propose an estimate of the bias $\overline{A}_k^{-1}B_{kNT}$ and then demonstrate (S2.2).

To correct the bias, we propose to obtain consistent estimates of $\overline{A}_k$ and B_{kNT}, respectively, by

$$\tilde{A}_k=\frac{1}{\tilde{N}_k}\sum_{i\in\tilde{G}_k}\tilde{Q}'_{i,z\Delta x}\tilde{Q}_{i,z\Delta x}\quad\text{and}$$

$$\tilde{B}_{kNT}=\frac{1}{\tilde{N}_k^{1/2}T^{3/2}}\sum_{i\in\tilde{G}_k}\sum_{s=1}^{T}\sum_{t=1}^{T}k_{M_T}(t,s)\Delta x_{is}z'_{is}z_{it}\Delta\tilde{\varepsilon}_{it},$$

where $\Delta\tilde{\varepsilon}_{it}=\Delta y_{it}-\tilde{\alpha}'_{\tilde{G}_k}\Delta x_{it}$ for all $i\in\tilde{G}_k$,[3] $k_{M_T}(t,s)$ is as defined above: $k_{M_T}(t,s)=k^0_{M_T}(|t-s|)$ and $k^0_{M_T}(u)$ denotes the Bartlett kernel: $k^0_{M_T}(u)=(1-|u|/M_T)\mathbf{1}\{|u|\le M_T\}$. Note that we also allow dynamic misspecification here. If one is sure that the model is dynamically correctly specified in the sense that $\mathbb{E}(\Delta\varepsilon_{it}|\mathcal{F}_{i,t-1})=0$ where $\mathcal{F}_{i,t-1}=\sigma(\Delta\varepsilon_{i,t-1},\Delta x_{it-1},z_{it};\Delta\varepsilon_{i,t-2},\Delta x_{it-2},z_{i,t-1};\ldots)$, one can use the one-sided kernel: $k_{M_T}(t,s)=k^1_{M_T}(s-t)$, where $k^1_{M_T}(u)=(1-u/M_T)\mathbf{1}\{0\le u\le M_T\}$. The bias-corrected C-Lasso estimator of α_k^0 would be

$$\tilde{\alpha}_k^{(c)}=\tilde{\alpha}_k-\frac{1}{\sqrt{\tilde{N}_kT}}\tilde{A}_k^{-1}\tilde{B}_{kNT}.$$

Note that Theorem 3.4 indicates that there is no need to consider bias correction for the post-Lasso estimator $\tilde{\alpha}_{\tilde{G}_k}$.

Let $x_i\equiv(x_{i1},\ldots,x_{iT})'$ and $\varepsilon_i\equiv(\varepsilon_{i1},\ldots,\varepsilon_{iT})'$. We add the following assumption.

[3]Observe that $\tilde{\alpha}_k-\alpha_k^0=O_P((N_kT)^{-1/2}+T^{-1})$ and $\tilde{\alpha}_{\tilde{G}_k}-\alpha_k^0=O_P((N_kT)^{-1/2})$. We recommend using the post-Lasso estimator $\tilde{\alpha}_{\tilde{G}_k}$.

ASSUMPTION D2: (i) *For each* $i = 1, \ldots, N$, $\{(\Delta x_{it}, z_{it}, \Delta\varepsilon_{it}) : t = 1, 2, \ldots\}$ *is strong mixing with mixing coefficients* $\{\alpha_i(\cdot)\}$. *In addition*, $\alpha_i(\tau) \leq c_{\alpha,i}\rho^{\tau}$ *for some* $c_{\alpha,i} < \infty$ *and* $\rho \in (0, 1)$, *where* $\frac{1}{N_k}\sum_{i\in G_k^0} c_{\alpha,i}^{(2q-1)/(2q)} = O(1)$ *and* $\frac{1}{N_k}\sum_{i\in G_k^0} c_{\alpha,i}^{(q-1)/q} = O(1)$.

(ii) (x_i, ε_i) *are independent across* $i \in G_k^0$, *where* $k = 1, \ldots, K_0$.

(iii) $\max_{i,t} \mathbb{E}\|\Delta x_{it} z_{it}'\|^{4q} < C < \infty$ *and* $\max_{i,t} \mathbb{E}\|z_{it}\Delta\varepsilon_{it}\|^{4q} < C < \infty$ *for some* $q > 1$.

(iv) *As* $(N, T) \to \infty$, $M_T \to \infty$, $M_T^2/T \to 0$ *and* $N_k^{-1/2}T^{1/2} \times \sum_{i\in G_k^0} \alpha_i(M_T)^{(2q-1)/(2q)} \to 0$ *for each* $k = 1, \ldots, K_0$.

Assumption D2(i)–(iv) parallels Assumption D1(i)–(iv). The major difference is that we do not need $M_T^2 N_k/T^3 \to 0$ in Assumption D2(iv) but require $q > 1$ in Assumption D2(iii).

PROPOSITION S2.2: *Suppose that the conditions of Theorem* 3.4 *hold*. *Suppose Assumption* D2 *holds*. *Then* $\tilde{A}_k^{-1}\tilde{B}_{kNT} - \overline{A}_k^{-1}B_{kNT} = o_P(1)$.

PROOF: Noting that $\tilde{A}_k^{-1}\tilde{B}_{kNT} - \overline{A}_k^{-1}B_{kNT} = (\tilde{A}_k^{-1} - \overline{A}_k^{-1})B_{kNT} + (\tilde{A}_k^{-1} - \overline{A}_k^{-1})(\tilde{B}_{kNT} - B_{kNT}) + \overline{A}_k^{-1}(\tilde{B}_{kNT} - B_{kNT})$, $\overline{A}_k^{-1} = O(1)$, and $B_{kNT} = O(\sqrt{N_k/T})$, it suffices to show that (i) $\tilde{A}_k - \overline{A}_k = o_P(\nu_{NT})$ and (ii) $\tilde{B}_{kNT} - B_{kNT} = o_P(1)$, where $\nu_{NT} = \min(1, \sqrt{T/N_k})$.

We first prove (i). Note that

$$\begin{aligned}
\tilde{A}_k - \overline{A}_k &= \frac{1}{\tilde{N}_k}\sum_{i\in\tilde{G}_k} \tilde{Q}_{i,z\Delta x}'\tilde{Q}_{i,z\Delta x} - \frac{1}{N_k}\sum_{i\in G_k^0} \tilde{Q}_{i,z\Delta x}'\tilde{Q}_{i,z\Delta x} \\
&= \frac{1}{\tilde{N}_k}\left(\sum_{i\in\tilde{G}_k} - \sum_{i\in G_k^0}\right)\tilde{Q}_{i,z\Delta x}'\tilde{Q}_{i,z\Delta x} + \frac{N_k - \tilde{N}_k}{\tilde{N}_k N_k}\sum_{i\in G_k^0} \tilde{Q}_{i,z\Delta x}'\tilde{Q}_{i,z\Delta x} \\
&\equiv A_{k,1} + A_{k,2}, \quad \text{say.}
\end{aligned}$$

By Corollary 3.3, $A_{k,2} = O_P(N_k^{-1}) = o_P(\nu_{NT})$. For any $\epsilon > 0$, we have by the proof of Theorem 3.2, $P(\|A_{k,1}\| \geq \nu_{NT}\epsilon) \leq P(\tilde{F}_{kNT}) + P(\tilde{E}_{kNT}) = o(1)$. It follows that $\tilde{A}_k - \overline{A}_k = o_P(\nu_{NT})$.

Now we prove (ii). We make the following decomposition:

$$\begin{aligned}
&\tilde{B}_{kNT} - B_{kNT} \\
&\quad = \frac{1}{\tilde{N}_k^{1/2}T^{3/2}}\sum_{i\in\tilde{G}_k}\sum_{s=1}^{T}\sum_{t=1}^{T} k_{M_T}(t, s)\Delta x_{is}z_{is}'z_{it}\Delta\tilde{\varepsilon}_{it}
\end{aligned}$$

$$
\begin{aligned}
&- \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E}\big(\Delta x_{is} z_{is}' z_{it} \Delta \varepsilon_{it}\big) \\
&= \frac{1}{\tilde{N}_k^{1/2} T^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} k_{M_T}(t,s) \Delta x_{is} z_{is}' z_{it} \Delta \tilde{\varepsilon}_{it} \\
&\quad - \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \mathbb{E}\big(\Delta x_{is} z_{is}' z_{it} \Delta \varepsilon_{it}\big) + o_P(1) \\
&= \frac{1}{\tilde{N}_k^{1/2} T^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} k_{M_T}(t,s) \Delta x_{is} z_{is}' z_{it} (\Delta \tilde{\varepsilon}_{it} - \Delta \varepsilon_{it}) \\
&\quad + \frac{1}{\tilde{N}_k^{1/2} T^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} k_{M_T}(t,s) \\
&\quad \times \big[\Delta x_{is} z_{is}' z_{it} \Delta \varepsilon_{it} - \mathbb{E}\big(\Delta x_{is} z_{is}' z_{it} \Delta \varepsilon_{it}\big)\big] \\
&\quad + \frac{N_k^{-1/2} - \tilde{N}_k^{-1/2}}{T^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} k_{M_T}(t,s) \mathbb{E}\big(\Delta x_{is} z_{is}' z_{it} \Delta \varepsilon_{it}\big) \\
&\quad + \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \big[1 - k_{M_T}(t,s)\big] \mathbb{E}\big(\Delta x_{is} z_{is}' z_{it} \Delta \varepsilon_{it}\big) + o_P(1) \\
&\equiv B_{kNT,1} + B_{kNT,2} + B_{kNT,3} + B_{kNT,4} + o_P(1),
\end{aligned}
$$

where the $o_P(1)$ term arises due to the replacement of $\tilde{G}_k$ by G_k^0 and this can be easily justified by using the uniform classification consistency result and arguments as used in the proof of Theorem 2.5. We prove (ii) by demonstrating that $B_{kNT,s} = o_P(1)$ for $s = 1, 2, 3, 4$.

First, noting that $\Delta \tilde{\varepsilon}_{it} - \Delta \varepsilon_{it} = (\alpha_k^0 - \tilde{\alpha}_{\tilde{G}_k})' \Delta x_{it}$, $\tilde{\alpha}_{\tilde{G}_k} - \alpha_k^0 = O_P((N_k T)^{-1/2})$, and that $N_k / \tilde{N}_k = 1 + o_P(1)$ by Corollary 3.3, we have

$$
\begin{aligned}
\|B_{kNT,1}\| &= \frac{1}{\tilde{N}_k^{1/2} T^{3/2}} \left\| \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} k_{M_T}(t,s) \Delta x_{is} z_{is}' z_{it} (\Delta x_{it})' \big(\alpha_k^0 - \tilde{\alpha}_{\tilde{G}_k}\big) \right\| \\
&\leq (\tilde{N}_k T)^{1/2} \big\| \alpha_k^0 - \tilde{\alpha}_{\tilde{G}_k} \big\| \frac{N_k}{\tilde{N}_k} \frac{1}{N_k T^2}
\end{aligned}
$$

$$\times \sum_{i\in G_k^0} \sum_{|s-t|\le M_T} \left\|\Delta x_{is} z_{is}' z_{it} (\Delta x_{it})'\right\|$$
$$= O_P(1) b_{kNT,1},$$

where $b_{kNT,1} = \frac{1}{N_k T^2} \sum_{i\in G_k^0} \sum_{|s-t|\le M_T}^{T} \|\Delta x_{is} z_{is}' z_{it} (\Delta x_{it})'\|$. By the Markov inequality, $b_{kNT,1} = O_P(M_T/T)$. It follows that $\|B_{kNT,1}\| = O_P(M_T/T) = o_P(1)$ under Assumption D2(iv).

For $B_{kNT,2}$, note that $B_{kNT,2} = b_{kNT,2} N_k^{1/2}/\tilde{N}_k^{1/2} = b_{kNT,2}(1 + o_P(1))$, where

$$b_{kNT,2} = \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} k_{M_T}(t,s)$$
$$\times \left[\Delta x_{is} z_{is}' z_{it} \Delta\varepsilon_{it} - \mathbb{E}\left(\Delta x_{is} z_{is}' z_{it} \Delta\varepsilon_{it}\right)\right].$$

Let ω be any $p\times 1$ nonrandom vector such that $\|\omega\| = 1$. Then $\mathbb{E}(\omega' b_{kNT,2}) = 0$. By Assumption D2(ii)–(iv) and the Jensen inequality,

$$\operatorname{Var}\left(\omega' b_{kNT,2}\right)$$
$$= \frac{1}{N_k T^3} \sum_{i\in G_k^0} \operatorname{Var}\left[\sum_{s=1}^{T} \sum_{t=1}^{T} k_{M_T}(t,s) \times \omega'\left\{\Delta x_{is} z_{is}' z_{it} \Delta\varepsilon_{it} - \mathbb{E}\left(\Delta x_{is} z_{is}' z_{it} \Delta\varepsilon_{it}\right)\right\}\right]$$
$$\le \frac{1}{N_k T^3} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \sum_{r=1}^{T} \sum_{l=1}^{T} k_{M_T}(t,s) k_{M_T}(r,l)$$
$$\times \omega' \mathbb{E}\left[\Delta x_{is} z_{is}' z_{it} \Delta\varepsilon_{it} \Delta x_{il} z_{il}' z_{ir} \Delta\varepsilon_{ir}\right]\omega$$
$$\le \frac{1}{N_k T^3} \sum_{i\in G_k^0} \sum_{|s-t|\le M_T} \sum_{|r-l|\le M_T} \left\|\mathbb{E}\left[\omega' \Delta x_{is} z_{is}' z_{it} \Delta\varepsilon_{it} \Delta x_{il} z_{il}' z_{ir} \Delta\varepsilon_{ir} \omega\right]\right\|$$
$$= O\left(M_T^2/T\right) = o(1),$$

where the last equality follows from the fact that

$$\left\|\mathbb{E}\left[\omega' \Delta x_{is} z_{is}' z_{it} \Delta\varepsilon_{it} \Delta x_{il} z_{il}' z_{ir} \Delta\varepsilon_{ir} \omega\right]\right\|$$
$$\le \max_{i,s}\left\{\mathbb{E}\left\|\Delta x_{is} z_{is}'\right\|^4\right\}^{1/2} \times \max_{i,t}\left\{\mathbb{E}\|z_{it}\Delta\varepsilon_{it}\|^4\right\}^{1/2} < C < \infty$$

by Assumption D2(iii). It follows that $B_{kNT,2} = o_P(1)$.

By Corollary 3.3 and the Davydov inequality,

$$\begin{aligned}
&\|B_{kNT,3}\| \\
&= \frac{\left|N_k^{-1} - \tilde{N}_k^{-1}\right|}{T^{3/2}\left(N_k^{-1/2} + \tilde{N}_k^{-1/2}\right)} \left\| \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} k_{M_T}(t, s) \mathbb{E}\left(\Delta x_{is} z_{is}' z_{it} \Delta \varepsilon_{it}\right) \right\| \\
&\leq \frac{|\tilde{N}_k - N_k|}{T^{1/2} \tilde{N}_k \left(N_k^{-1/2} + \tilde{N}_k^{-1/2}\right)} \left\{ \frac{1}{N_k T} \sum_{i \in G_k^0} \sum_{|s-t| \leq M_T} \left\| \mathbb{E}\left(\Delta x_{is} z_{is}' z_{it} \Delta \varepsilon_{it}\right) \right\| \right\} \\
&= o_P\left(N_k^{-1/2} T^{-1/2}\right) O(1) = o_P(1).
\end{aligned}$$

By Assumption D2(i)–(iii) and the Davydov inequality,

$$\begin{aligned}
&\|B_{kNT,4}\| \\
&= \left\| \frac{1}{N_k^{1/2} T^{3/2}} \sum_{i \in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \left[1 - k_{M_T}(t, s)\right] \mathbb{E}\left(\Delta x_{is} z_{is}' z_{it} \Delta \varepsilon_{it}\right) \right\| \\
&\leq \frac{8}{N_k^{1/2} T^{3/2}} \sum_{i \in G_k^0} \sum_{|s-t| > M_T} \alpha_i\left(|s - t|\right)^{(2q-1)/(2q)} \left\| \Delta x_{is} z_{is}' \right\|_{4q} \| z_{it} \Delta \varepsilon_{it} \|_{4q} \\
&\leq C N_k^{-1/2} T^{1/2} \sum_{i \in G_k^0} \alpha_i (M_T)^{(2q-1)/(2q)} = o(1).
\end{aligned}$$

This completes the proof of the proposition. *Q.E.D.*

With the above result in hand, we can readily show that

$$\begin{aligned}
\sqrt{N_k T}\left(\tilde{\alpha}_k^{(c)} - \alpha_k^0\right) &= \left[\sqrt{N_k T}\left(\tilde{\alpha}_k - \alpha_k^0\right) - \overline{A}_k^{-1} B_{kNT}\right] \\
&\quad + (N_k/\tilde{N}_k)^{1/2} \left[\overline{A}_k^{-1} B_{kNT} - \tilde{A}_k^{-1} \tilde{B}_{kNT}\right] \\
&\quad + \left[1 - (N_k/\tilde{N}_k)^{1/2}\right] \overline{A}_k^{-1} B_{kNT} \\
&= \left[\sqrt{N_k T}\left(\tilde{\alpha}_k - \alpha_k^0\right) - \overline{A}_k^{-1} B_{kNT}\right] \\
&\quad + o_P(1) + o_P\left(N_k^{-1}\right) O\left((N_K/T)^{1/2}\right) \\
&= \left[\sqrt{N_k T}\left(\tilde{\alpha}_k - \alpha_k^0\right) - \overline{A}_k^{-1} B_{kNT}\right] + o_P(1).
\end{aligned}$$

That is, $\sqrt{N_k T}(\tilde{\alpha}_k^{(c)} - \alpha_k^0)$ has the desired limiting distribution centered on the origin.

Now, we demonstrate (S2.2). Let $\xi_{is} = \Delta x_{is} z_{is}' - \mathbb{E}(\Delta x_{is} z_{is}')$ and $\eta_{it} = z_{it}\Delta\varepsilon_{it}$. Noting that $\mathbb{E}(\xi_{is}) = 0$ and $\mathbb{E}(\eta_{it}) = 0$, we have

$$\begin{aligned} R_{kNT} &= \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i\in G_k^0} \sum_{s=1}^{T} \sum_{t=1}^{T} \left[\xi_{is}\eta_{it} - \mathbb{E}(\xi_{is}\eta_{it})\right] \\ &= \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i\in G_k^0} \sum_{t=1}^{T} \left[\xi_{it}\eta_{it} - \mathbb{E}(\xi_{it}\eta_{it})\right] \\ &\quad + \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i\in G_k^0} \sum_{1\le s<t\le T} \left[\xi_{is}\eta_{it} - \mathbb{E}(\xi_{is}\eta_{it})\right] \\ &\quad + \frac{1}{N_k^{1/2}T^{3/2}} \sum_{i\in G_k^0} \sum_{1\le t<s\le T} \left[\xi_{is}\eta_{it} - \mathbb{E}(\xi_{is}\eta_{it})\right] \\ &\equiv R_{kNT,1} + R_{kNT,2} + R_{kNT,3}, \quad \text{say.} \end{aligned}$$

It is trivial to show that $R_{kNT,1} = O_P(T^{-1})$ by the Chebyshev and Davydov inequalities. For $R_{kNT,2}$, we have $\mathbb{E}(R_{kNT,2}) = 0$ by construction, and by Assumption D2(ii) and the Jensen inequality,

$$\begin{aligned} &\mathbb{E}\big(R_{kNT,2}^2\big) \\ &\quad = \frac{1}{N_kT^3} \sum_{i\in G_k^0} \operatorname{Var}\bigg(\sum_{1\le t_1<t_2\le T} \left[\xi_{it_1}\eta_{it_2} - \mathbb{E}(\xi_{it_1}\eta_{it_2})\right]\bigg) \\ &\quad \le \frac{1}{N_kT^3} \sum_{i\in G_k^0} \sum_{1\le t_1<t_2\le T} \sum_{1\le t_3<t_4\le T} \mathbb{E}(\xi_{it_1}\eta_{it_2}\xi_{it_3}\eta_{it_4}) \equiv S_{kNT}, \quad \text{say.} \end{aligned}$$

To bound S_{kNT}, we can consider three subcases: (a) $\#\{t_1, t_2, t_3, t_4\} = 4$, (b) $\#\{t_1, t_2, t_3, t_4\} = 3$, and (c) $\#\{t_1, t_2, t_3, t_4\} = 2$, and use $S_{kNT,a}$, $S_{kNT,b}$, and $S_{kNT,c}$ to denote the last summation when the time indices are restricted to these three cases in order. Apparently, $S_{kNT,c} = O(1/T)$ under Assumption D2(iii). In case (a), without loss of generality (wlog) assume that $1 \le t_1 < t_2 < t_3 < t_4 \le T$ and denote $S_{kNT,a}^{(1)}$ as $S_{kNT,a}$ when the time indices are restricted to this subcase. (Note that the other subcases can be analyzed analogously.)

Let d_c be the cth largest difference among $t_{j+1}-t_j$ for $j=1,2,3$. Then

$$S^{(1)}_{kNT,a}=\frac{1}{N_kT^3}\sum_{i\in G_k^0}\Bigg\{\sum_{1\le t_1<t_2<t_3<t_4\le T,t_2-t_1=d_1}+\sum_{1\le t_1<t_2<t_3<t_4\le T,t_3-t_2=d_1}$$
$$+\sum_{1\le t_1<t_2<t_3<t_4\le T,t_4-t_3=d_1}\Bigg\}\mathbb{E}(\xi_{it_1}\eta_{it_2}\xi_{it_3}\eta_{it_4})$$
$$\equiv S^{(1)}_{kNT,a1}+S^{(1)}_{kNT,a2}+S^{(1)}_{kNT,a3},\quad \text{say.}$$

By the Davydov inequality and Assumption D2(i) and (iii),

$$S^{(1)}_{NT,a1}\le\frac{1}{N_kT^3}\sum_{i\in G_k^0}\sum_{t_1=1}^{T-3}\sum_{t_2=t_1+\max_{j\ge3}\{t_j-t_{j-1}\}}^{T-2}\sum_{t_3=t_2+1}^{T-1}\sum_{t_4=t_3+1}^{T}\|\xi_{it_1}\|_{4q}$$
$$\times\|\eta_{it_2}\xi_{it_3}\eta_{it_4}\|_{4q/3}\alpha_i(t_2-t_1)^{(q-1)/q}$$
$$\le\frac{C}{N_kT^3}\sum_{i=1}^{N}\sum_{t_1=1}^{T-3}\sum_{t_2=t_1+1}^{T-2}(t_2-t_1)^2\alpha_i(t_2-t_1)^{(q-1)/q}$$
$$\le\frac{1}{N_kT}\sum_{i=1}^{N}\sum_{\tau=1}^{\infty}\tau\alpha_i(\tau)^{(q-1)/q}=O\big(T^{-1}\big).$$

Similarly, we can show that $S^{(1)}_{kNT,as}=O(1/T)$ for $s=2,3$. It follows that $S^{(1)}_{kNT,a}=O(1/T)$ and $S^{(1)}_{kNT,a}=O(1/T)=o(1)$. In case (b), wlog assume that $t_4=t_2$ and $1\le t_1<t_2<t_3\le T$ and we use $S^{(1)}_{kNT,b}$ to $S_{kNT,b}$ when the time indices are restricted to this subcase. Then by the Davydov inequality and Assumption D2(i) and (iii),

$$\big|S^{(1)}_{kNT,b}\big|=\frac{1}{N_kT^3}\sum_{i\in G_k^0}\sum_{1\le t_1<t_2<t_3\le T}\big|\mathbb{E}\big(\xi_{it_1}\eta^2_{it_2}\xi_{it_3}\big)\big|$$
$$\le\frac{8}{N_kT^3}\sum_{i=1}^{N}\sum_{1\le t_1<t_2<t_3\le T}\big\|\xi_{it_1}\eta^2_{it_2}\big\|_{4q/3}\|\xi_{it_3}\|_{4q}\alpha_i(t_3-t_2)^{(q-1)/q}$$
$$\le\frac{8C}{N_kT}\sum_{i=1}^{N}\sum_{\tau=1}^{\infty}\alpha_i(\tau)^{(q-1)/q}=O\big(T^{-1}\big).$$

So $S_{kNT,b}=O(T^{-1})$. Consequently, $S_{kNT}=O(T^{-1})$ and $R_{kNT,2}=O_P(T^{-1/2})$ by the Chebyshev inequality. By the same token, $R_{kNT,3}=O_P(T^{-1/2})$. Thus we have shown that $R_{kNT}=O_P(T^{-1/2})=o_P(1)$.

S3. NUMERICAL ALGORITHM AND ADDITIONAL SIMULATION RESULTS

S3.1. *Numerical Algorithm*

In this section, we propose an iterative algorithm to obtain the PPL estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ in Section 2. A similar algorithm applies for PGMM estimation. Documented computer code is available online.

Step 1. Start with an initial value $\hat{\boldsymbol{\alpha}}^{(0)} = (\hat{\alpha}_1^{(0)}, \dots, \hat{\alpha}_K^{(0)})$ and $\hat{\boldsymbol{\beta}}^{(0)} = (\hat{\beta}_1^{(0)}, \dots, \hat{\beta}_N^{(0)})$ such that $\sum_{i=1}^N \|\hat{\beta}_i^{(0)} - \hat{\alpha}_k^{(0)}\| \neq 0$ for each $k = 2, \dots, K$. Set the iteration index $r = 1$.

Step 2. Given $\hat{\boldsymbol{\alpha}}^{(r-1)} = (\hat{\alpha}_1^{(r-1)}, \dots, \hat{\alpha}_K^{(r-1)})$ and $\hat{\boldsymbol{\beta}}^{(r-1)} = (\hat{\beta}_1^{(r-1)}, \dots, \hat{\beta}_N^{(r-1)})$, we first choose $(\boldsymbol{\beta}, \alpha_1)$ to minimize

$$Q_{1NT,\lambda_1}^{(r,1,K)}(\boldsymbol{\beta}, \alpha_1) = Q_{1,NT}(\boldsymbol{\beta}) + \frac{\lambda_1}{N} \sum_{i=1}^N \|\beta_i - \alpha_1\| \prod_{k\neq 1}^K \big\|\hat{\beta}_i^{(r-1)} - \hat{\alpha}_k^{(r-1)}\big\|,$$

and obtain the updated estimate $(\hat{\boldsymbol{\beta}}^{(r,1)}, \hat{\alpha}_1^{(r)})$ of $(\boldsymbol{\beta}, \alpha_1)$. Next we choose $(\boldsymbol{\beta}, \alpha_2)$ to minimize

$$\begin{aligned} &Q_{1NT,\lambda_1}^{(r,2,K)}(\boldsymbol{\beta}, \alpha_2) \\ &\quad = Q_{1,NT}(\boldsymbol{\beta}) + \frac{\lambda_1}{N} \sum_{i=1}^N \|\beta_i - \alpha_2\| \big\|\hat{\beta}_i^{(r,1)} - \hat{\alpha}_1^{(r)}\big\| \prod_{k\neq 1,2}^K \big\|\hat{\beta}_i^{(r-1)} - \hat{\alpha}_k^{(r-1)}\big\|, \end{aligned}$$

and obtain the updated estimate $(\hat{\boldsymbol{\beta}}^{(r,2)}, \hat{\alpha}_2^{(r)})$ of $(\boldsymbol{\beta}, \alpha_2)$. Repeat this procedure until we choose $(\boldsymbol{\beta}, \alpha_K)$ to minimize

$$Q_{1NT,\lambda_1}^{(r,K,K)}(\boldsymbol{\beta}, \alpha_K) = Q_{1,NT}(\boldsymbol{\beta}) + \frac{\lambda_1}{N} \sum_{i=1}^N \|\beta_i - \alpha_K\| \prod_{k=1}^{K-1} \big\|\hat{\beta}_i^{(r,k)} - \hat{\alpha}_k^{(r)}\big\|,$$

and obtain the updated estimate $(\hat{\boldsymbol{\beta}}^{(r,K)}, \hat{\alpha}_K^{(r)})$ of $(\boldsymbol{\beta}, \alpha_K)$. Let $\hat{\boldsymbol{\alpha}}^{(r)} = (\hat{\alpha}_1^{(r)}, \dots, \hat{\alpha}_K^{(r)})$ and $\hat{Q}_{1NT}^{(r,K)} = \sum_{k=1}^K Q_{1NT,\lambda_1}^{(r,k,K)}(\hat{\boldsymbol{\beta}}^{(r,k)}, \hat{\boldsymbol{\alpha}}_k^{(r)})$. Update the iteration index from r to $r+1$.

Step 3. Repeat Step 2 until a convergence criterion is achieved, for example, when

$$\hat{Q}_{1NT}^{(r-1,K)} - \hat{Q}_{1NT}^{(r,K)} < \boldsymbol{\epsilon}_{\text{tol}} \quad \text{and} \quad \frac{\sum_{k=1}^K \big\|\hat{\alpha}_k^{(r)} - \hat{\alpha}_k^{(r-1)}\big\|^2}{\sum_{k=1}^K \big\|\hat{\alpha}_k^{(r-1)}\big\|^2 + 0.0001} < \boldsymbol{\epsilon}_{\text{tol}},$$

where ϵ_{tol} is some prescribed tolerance level (e.g., 0.0001). Define the final iterative estimate of $\boldsymbol{\alpha}$ as $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1^{(R)}, \ldots, \hat{\alpha}_K^{(R)})$ for a sufficiently large R such that the convergence criterion is met. Intuitively, individual i is classified to group $\hat{G}_k$ if $\hat{\beta}_i^{(R,k)} = \hat{\alpha}_k$; otherwise, $\hat{\beta}_i$ is assigned to be the $\alpha_k^{(R)}$ that is closest to some $\hat{\beta}_i^{(R,l)}$, $l = 1, \ldots, K$. In either case, we can write the individual estimate as $\hat{\beta}_i = \hat{\alpha}_{k^*}^{(R)}$, where $k^* = \arg\min_{k \in \{1,\ldots,K\}} \|\hat{\beta}_i^{(R,l^*(k))} - \hat{\alpha}_k^{(R)}\|$ and $l^*(k) = \arg\min_{l \in \{1,\ldots,K\}} \|\hat{\beta}_i^{(R,l)} - \hat{\alpha}_k^{(R)}\|$.

S3.2. *Convexity, Choice of Initial Value, and Convergence of the Algorithm*

The optimization of $Q_{1NT,\lambda_1}^{(r,k,K)}(\boldsymbol{\beta}, \alpha_k)$ is conducted on the $(Np + p)$-dimensional parameter space for $(\boldsymbol{\beta}, \alpha_k)$. When N is nontrivial, this is a high-dimensional optimization problem. Obviously, in the penalty term, $\beta_1, \ldots, \beta_N$ and α_k are jointly convex, given $\prod_{l=k+1}^{K} \|\hat{\beta}_i^{(r-1)} - \hat{\alpha}_l^{(r-1)}\|$ and $\prod_{l=1}^{k-1} \|\hat{\beta}_i^{(r)} - \hat{\alpha}_l^{(r)}\|$ for each $i = 1, \ldots, N$. If $Q_{1,NT}(\boldsymbol{\beta})$ is convex in $\boldsymbol{\beta}$, then $Q_{1NT,\lambda_1}^{(r,k,K)}(\boldsymbol{\beta}, \alpha_k)$, as the summation of $Q_{1,NT}(\boldsymbol{\beta})$ and the penalty, is also convex in $(\boldsymbol{\beta}, \alpha_k)$. Convexity can substantially reduce the computational burden of high-dimensional optimization.

A convex $Q_{1,NT}(\boldsymbol{\beta})$ is common in panel data models. Convexity apparently holds in the linear models in Examples 1 and 2. It also holds in the nonlinear models in Example 3 with $F(\cdot)$ as the standard logistic or normal CDF, and in Example 4 after reparameterizing the original parameter $(\beta_i, \mu_i, \sigma_\varepsilon^2)$ into $(\theta_{1i} = \beta_i/\sigma_\varepsilon^2, \theta_{2i} = \mu_i/\sigma_\varepsilon^2, \theta_3 = 1/\sigma_\varepsilon^2)$. We utilize the convexity throughout our numerical works.

Given the convexity in each substep (r, k), the proposed algorithm consists of a sequence of convex problems implemented in an iterative manner. In particular, the only difference between the standard Lasso and a single substep of PPL is that Lasso shrinks the coefficients to a known center (zero), while the center of PPL is determined in the convex programming. Thus, a PPL iteration has the same computational complexity as Lasso, which is $O(N^3T)$ in our context of panel linear regression (Efron, Hastie, and Johnstone (2004, p. 443)). The computational cost of a single iteration is minimal.

Since the additive-multiplicative penalty is not jointly convex in all the parameters $(\boldsymbol{\beta}, \boldsymbol{\alpha})$, we can take advantage of convexity in each substep for $(\boldsymbol{\beta}, \alpha_k)$ but not simultaneously for $(\boldsymbol{\beta}, \boldsymbol{\alpha})$. As a consequence of the non-convexity, the sequence $\hat{\mathcal{Q}}_{1NT}^{(r,K)}$, $r = 1, \ldots, R$, depends on the initial value $(\hat{\boldsymbol{\alpha}}^{(0)}, \hat{\boldsymbol{\beta}}^{(0)})$, and $\hat{\mathcal{Q}}_{1NT}^{(R,K)}$ might terminate at a local minimum but not a global minimum.

A natural initial value is to set $(\alpha_k^{(0)} = 0)_{k=1}^K$ and each $\beta_i^{(0)}$ as the QMLE from the ith individual time series $(w_{i1}, \ldots, w_{iT})$. Denote this particular choice $(\boldsymbol{\beta}^{\text{init}}, \boldsymbol{\alpha}^{\text{init}})$, and we use it in all the simulations as well as applications, if not explicitly stated otherwise. As we compare it with other possible choices, for

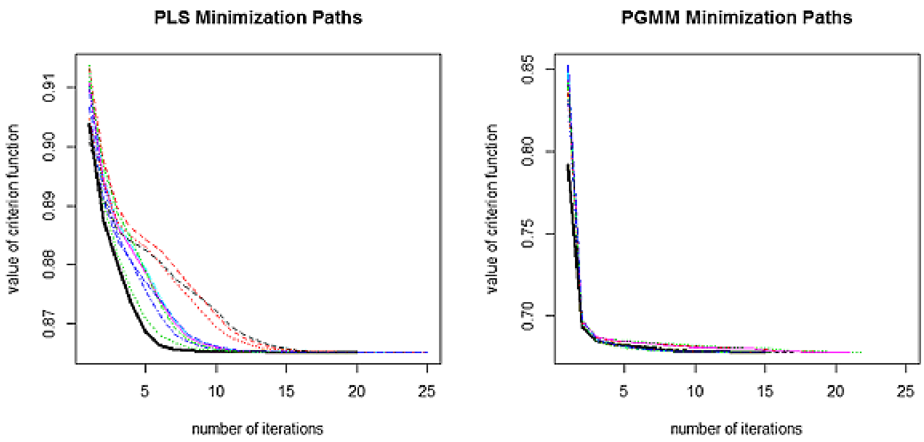

FIGURE S1.—PLS and PGMM paths starting at different initial values.

example, $(\alpha_k^{(0)}=0)_{k=1}^K$ and $(\beta_i^{(0)}=1)_{i=1}^N$, starting at $(\boldsymbol{\beta}^{\text{init}}, \boldsymbol{\alpha}^{\text{init}})$ often makes the algorithm converge in fewer iterations.

Although a formal investigation of the algorithm's computational complexity to attain the global optimum is beyond the scope of the paper, we explore its numerical convergence and sensitivity to initial values through a numerical example. We use the real data of savings rate in Section 5.1, and apply PLS and PGMM given the number of groups and the tuning parameters selected according to the IC. The left subgraph in Figure S1 shows the path of $\hat{\mathcal{Q}}_{1NT}^{(r,K)}$, $r=1,\dots,R$, and the right subgraph displays its PGMM counterpart. Each of the ten paths is associated with a different starting point. First, the bold black curve is the path that starts at $(\boldsymbol{\beta}^{\text{init}}, \boldsymbol{\alpha}^{\text{init}})$. Next, we perturb the initial value to be $((\beta_i^{\text{init}}+e_i)_{i=1}^N, \boldsymbol{\alpha}^{\text{init}})$, where e_i is a vector of p elements, each of which is randomly drawn from Uniform$(-1,1)$ and is independent across i. This is a substantial perturbation, in view of the magnitude of the estimates in Table III. We use the perturbed initial values to generate the other nine curves.

Figure S1 illustrates the robustness of C-Lasso to initial values. We observe in each subgraph that the criterion functions descend fast in the first few iterations, and then the paths turn almost flat until the tolerance level is reached. All paths converge to the same value of the criterion function in this experiment.

S3.3. *Additional Simulation Results*

In this section, we carry out two more simulation exercises, one using PGMM to estimate a static panel model with endogenous regressors as in DGP 4 below, and the other using PLS to estimate the linear panel AR(1) in DGP 2.

DGP 4 (Linear Static Panel With Endogeneity). We maintain the linear panel structure model with two explanatory variables as in DGP 1, but the first regressor is endogenous, as it is generated from the following underlying reduced-form equation: $x_{it1} = 0.2\mu_i^0 + 0.5z_{it1} + 0.5z_{it2} + 0.5e_{it}$, where z_{it1} and z_{it2} are two excluded instrumental variables, and the reduced-form error e_{it} and the structural-equation idiosyncratic shock ε_{it} follow a bivariate normal distribution:

$$\begin{pmatrix} \varepsilon_{it} \\ e_{it} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}\right).$$

The second regressor x_{it2} is exogenous, and $(x_{it2}, z_{it1}, z_{it2}) \sim$ i.i.d. $N(0, I_3)$ is independent of $(e_{it}, \varepsilon_{it})$. All variables are independent across i and t. The econometrician observes $(y_{it}, x_{it1}, x_{it2}, z_{it1}, z_{it2})$. The true coefficients of the three groups are $(0.2, 1.8)$, $(1, 1)$ and $(1.8, 0.2)$, respectively.

We report the statistics in Tables SI and SII, which correspond to Tables I and II, respectively, in the main text. The choice of tuning parameters is exactly the same as described in Section 4. When we compare PLS estimation with PGMM in DGP 2, we find that the PLS works better in determining the correct number of groups and in classifying the individual units. The 95% coverage probabilities are comparable to those of PGMM when $T = 50$, but are lower than PGMM when T is small. Similarly to PPL in DGP 3, the lower coverage probabilities are caused by the bias. The analytical bias correction removes the bias asymptotically, but the effect is limited when T is small, as is shown in the oracle. The post-Lasso has larger coverage probability than the oracle, as the estimated standard deviation is inflated by a few misclassified units.

Table SIII reports the RMSE and bias of $\boldsymbol{\alpha}_1$ from post-Lasso and C-Lasso under the true K_0 and the IC-determined $\hat{K}$ (or $\tilde{K}$ for PGMM). These estimates are bias-corrected whenever necessary in the DGPs. For example,

TABLE SI

FREQUENCY OF SELECTING $K = 1, \dots, 5$ GROUPS WHEN $K_0 = 3$

		DGP 4					DGP 2 (PLS)				
N	T	1	2	**3**	4	5	1	2	**3**	4	5
100	15	0	0.022	**0.902**	0.076	0	0	0.106	**0.894**	0	0
100	25	0	0	**0.966**	0.028	0.006	0	0	**1**	0	0
100	50	0	0	**0.996**	0.004	0	0	0	**1**	0	0
200	15	0	0	**0.940**	0.058	0.002	0	0	**1**	0	0
200	25	0	0	**0.950**	0.046	0.004	0	0	**1**	0	0
200	50	0	0	**0.994**	0.006	0	0	0	**1**	0	0

TABLE SII

CLASSIFICATION AND POINT ESTIMATION OF $\boldsymbol{\alpha}_1$ IN ADDITIONAL SIMULATIONS

	N	T	% of Correct Classification	Post-Lasso RMSE	Post-Lasso Bias	Post-Lasso Coverage	Oracle RMSE	Oracle Bias	Oracle Coverage
DGP 4	100	15	0.8287	0.1583	0.0462	0.7850	0.0806	0.0018	0.9344
	100	25	0.9281	0.0883	0.0195	0.8880	0.0617	0.0009	0.9380
	100	50	0.9885	0.0517	0.0075	0.9406	0.0437	−0.0012	0.9422
	200	15	0.8378	0.1155	0.0484	0.7860	0.0577	−0.0016	0.9454
	200	25	0.9320	0.0643	0.0199	0.8742	0.0436	0.0001	0.9506
	200	50	0.9881	0.0364	0.0074	0.9356	0.0311	−0.0005	0.9450
DGP 2	100	15	0.8907	0.0413	0.0061	0.9148	0.0352	0.0041	0.8524
(PLS)	100	25	0.9511	0.0261	0.0041	0.9710	0.0241	0.0028	0.9076
	100	50	0.9908	0.0160	0.0015	0.9908	0.0156	0.0013	0.9334
	200	15	0.8949	0.0294	0.0064	0.9154	0.0253	0.0052	0.8576
	200	25	0.9520	0.0188	0.0037	0.9714	0.0178	0.0036	0.8808
	200	50	0.9912	0.0113	0.0017	0.9934	0.0111	0.0015	0.9282

the RMSE of PPL under K_0 is calculated as $(\frac{1}{S}\sum_{s=1}^{S}\sum_{k=1}^{K_0}\frac{\hat{N}_k^{(s)}}{N}(\hat{\alpha}_{k,1}^{(s)}(K_0,\lambda_1)-\alpha_{k,1}^0)^2)^{1/2}$, where s and S are the index and the total number of simulation replications, respectively, and $\hat{N}_k^{(s)}=\sum_{i=1}^{N}\mathbf{1}\{i\in\hat{G}_k^{(s)}(K,\lambda_1)\}$ is the estimated number of units in the kth group. This quantity differs from its counterpart in Table II as each group-specific estimate is weighted by $\hat{N}_k^{(s)}/N$, instead of N_k/N, to take into account the uncertainty in classification. The bias is computed similarly. The post-Lasso's RMSE and bias under the known K_0 are close to the oracle. The performance of C-Lasso is in general comparable to that of post-Lasso, although C-Lasso appears to have larger RMSE in the PGMM estimation of DGP 2, where it does not enjoy the oracle property.

When $K\neq K_0$, we generalize the definition of the set of true group-specific parameters. For $K<K_0$, we shrink $\boldsymbol{\alpha}_1^0=(\alpha_{1,1}^0,\dots,\alpha_{K_0,1}^0)$ into a K-element subset $\boldsymbol{\alpha}_1^0(K)$. For $K>K_0$, we augment $\boldsymbol{\alpha}_1^0$ by adding $K-K_0$ elements choosing from $\alpha_{k,1}^0,\dots,\alpha_{K_0,1}^0$ so that the resulting $\boldsymbol{\alpha}_1^0(K)$ contains $\boldsymbol{\alpha}_1^0$. Elements are eliminated or concatenated in each replication to fit $\hat{\boldsymbol{\alpha}}(\hat{K}^{(s)},\lambda_1)$. In this scenario, the RMSE is calculated as $(\frac{1}{S}\sum_{s=1}^{S}\sum_{k=1}^{\hat{K}^{(s)}}\frac{\hat{N}_k^{(s)}}{N}(\hat{\alpha}_{k,1}(\hat{K}^{(s)},\lambda_1)-\alpha_{k,1}^0(\hat{K}^{(s)}))^2)^{1/2}$. According to the simulation, the effect of not knowing K_0 is noticeable when $T=15$ in the linear models and $T=25$ in the nonlinear model, but it does not necessarily enlarge the RMSE, for the estimator under K_0 is also noisy when T is small. The discrepancy of the RMSE and bias between K_0 and $\hat{K}$ (or $\tilde{K}$) quickly vanishes when T grows.

TABLE SIII

ESTIMATION OF $\boldsymbol{\alpha}_1$ BY POST-LASSO AND C-LASSO UNDER K_0 AND $\hat{K}$ OR $\tilde{K}$

			Post-Lasso				C-Lasso					
			$K = K_0$		$K = \hat{K}$		$K = K_0$		$K = \hat{K}$		Oracle	
	N	T	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
DGP 1	100	15	0.0596	0.0108	0.0829	0.0092	0.0619	0.0133	0.0839	0.0120	0.0463	0.0012
	100	25	0.0385	0.0019	0.0385	0.0019	0.0396	0.0040	0.0396	0.0040	0.0353	0.0001
	100	50	0.0249	0.0000	0.0249	0.0000	0.0255	0.0011	0.0255	0.0011	0.0245	−0.0002
	200	15	0.0434	0.0079	0.1373	0.0081	0.0457	0.0107	0.1353	0.0114	0.0324	−0.0013
	200	25	0.0273	0.0015	0.0273	0.0015	0.0280	0.0040	0.0280	0.0040	0.0250	−0.0006
	200	50	0.0174	−0.0001	0.0174	−0.0001	0.0181	0.0011	0.0181	0.0011	0.0171	−0.0002
DGP 2	100	15	0.0848	−0.0090	0.0787	−0.0016	0.1311	−0.0372	0.1188	−0.0250	0.0502	−0.0037
(PGMM)	100	25	0.0556	−0.0055	0.0561	−0.0051	0.1042	−0.0267	0.1045	−0.0255	0.0351	0.0011
	100	50	0.0278	−0.0012	0.0278	−0.0012	0.0418	−0.0130	0.0418	−0.0130	0.0242	−0.0010
	200	15	0.0712	−0.0141	0.0743	−0.0145	0.1491	−0.0399	0.1483	−0.0383	0.0352	−0.0017
	200	25	0.0333	−0.0051	0.0333	−0.0051	0.0932	−0.0284	0.0932	−0.0284	0.0252	−0.0006
	200	50	0.0193	−0.0014	0.0193	−0.0014	0.0277	−0.0134	0.0277	−0.0134	0.0164	0.0000
DGP 3	100	25	0.1722	0.0587	0.1516	0.0727	0.2154	0.0615	0.1641	0.0688	0.1077	0.0114
	100	50	0.0853	0.0379	0.0878	0.0383	0.1178	0.0487	0.1191	0.0489	0.0752	0.0090
	200	25	0.1342	0.0483	0.1401	0.0649	0.1826	0.0487	0.1441	0.0573	0.0821	0.0116
	200	50	0.0632	0.0264	0.0632	0.0264	0.0948	0.0372	0.0948	0.0372	0.0573	0.0121

(*Continues*)

TABLE SIII—*Continued*

			Post-Lasso				C-Lasso					
			$K=K_0$		$K=\hat{K}$		$K=K_0$		$K=\hat{K}$		Oracle	
	N	T	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
DGP 4	100	15	0.1691	0.0487	0.1803	0.0376	0.2148	0.1087	0.2102	0.0941	0.0806	0.0018
	100	25	0.0724	0.0189	0.1217	0.0207	0.0882	0.0523	0.1323	0.0539	0.0617	0.0009
	100	50	0.0450	0.0031	0.0645	0.0042	0.0532	0.0204	0.0707	0.0215	0.0437	−0.0012
	200	15	0.1271	0.0512	0.1348	0.0466	0.1777	0.1128	0.1793	0.1074	0.0577	−0.0016
	200	25	0.0513	0.0153	0.1392	0.0235	0.0720	0.0498	0.1485	0.0577	0.0436	0.0001
	200	50	0.0314	0.0036	0.0549	0.0049	0.0399	0.0221	0.0602	0.0234	0.0311	−0.0005
DGP 2	100	15	0.0482	0.0081	0.0487	0.0065	0.0747	0.0297	0.0715	0.0254	0.0352	0.0041
(PLS)	100	25	0.0263	0.0043	0.0263	0.0043	0.0418	0.0189	0.0418	0.0189	0.0241	0.0028
	100	50	0.0160	0.0016	0.0160	0.0016	0.0218	0.0085	0.0218	0.0085	0.0156	0.0013
	200	15	0.0295	0.0064	0.0295	0.0064	0.0567	0.0293	0.0567	0.0293	0.0253	0.0052
	200	25	0.0188	0.0037	0.0188	0.0037	0.0307	0.0174	0.0307	0.0174	0.0178	0.0036
	200	50	0.0113	0.0017	0.0113	0.0017	0.0171	0.0084	0.0171	0.0084	0.0111	0.0015

TABLE SIV
SUMMARY STATISTICS FOR THE SAVINGS DATA SET

	Mean	Median	S.E.	Min	Max
Savings rate	22.099	20.790	8.833	−3.207	53.434
Inflation rate	7.724	4.853	15.342	−3.846	293.679
Real interest rate	7.422	5.927	10.062	−63.761	93.915
Per capita GDP growth rate	2.855	2.971	3.865	−17.545	14.060

S4. ADDITIONAL APPLICATION RESULTS

S4.1. *More on Savings Rate Modeling and Classification*

All data are downloaded from the World Bank.[4] We extract all countries with all the variables in (5.1) available. Using the time span 1995–2010, we were able to construct a balanced panel of 57 countries. We remove one outlier, Bulgaria, whose 1997 economic collapse produced hyperinflation in the CPI that significantly distorted the overall mean and the standard deviation. In total, we collect 56 countries. The summary statistics are shown in Table SIV.

In the implementation, we scale-normalize all the variables for each individual unit to guarantee that the coefficients are comparable. Moreover, in PGMM we use Δy_{t-2} and a constant as two excluded IVs. Although the constant is uncorrelated with the endogenous variable, adding it here stabilizes the post-Lasso estimation in finite samples.

Table SV displays the group membership. The country names in bold are the 47 coincidences of PLS and PGMM classification out of the total 56 countries.

S4.2. *More on the Civil War Application*

The replication data of Fearon and Laitin (2003) can be downloaded from Fearon's personal web page.[5] The data span from 1945 to 1998,[6] but the panel is highly unbalanced. Following Collier and Hoeffler (2004), Djankov and Reynal-Querol (2010), and Blattman and Miguel (2010), we choose 1960 as the starting year to generate a balanced panel of $N = 38$, as many countries' civil war incidence is always 0 or 1 between 1960 and 1998.

In the regression, the dependent variable is the civil war incidence, and the explanatory variables are the lagged civil war incidence, the one-period difference of log GDP per capita, and the one-period difference of log population. Moreover, in view of the natural scaling of the binary variable, we keep the

[4]http://data.worldbank.org/data-catalog/world-development-indicators.

[5]https://www.stanford.edu/group/fearon-research/cgi-bin/wordpress/wp-content/uploads/2013/10/apsr03repdata.zip.

[6]The original data end at 1999, but no population information is provided for any country in the last year.

TABLE SV

ESTIMATED GROUP MEMBERSHIP

PLS	PGMM
Group 1: (31 countries) *Armenia*, *Australia*, *Bahamas*, *Belarus*, *Bolivia*, *Botswana*, *Cape Verde*, *China*, *Czech*, Guatemala, *Honduras*, *Hungary*, *Indonesia*, *Israel*, *Italy*, *Japan*, *Jordan*, *Latvia*, *Malawi*, *Malaysia*, *Mauritius*, *Mexico*, Mongolia, *Panama*, *Paraguay*, *Philippines*, *Romania*, *South Africa*, *Sri Lanka*, *Thailand*, *Ukraine*	Group 1: (36 countries) *Armenia*, *Australia*, *Bahamas*, *Belarus*, *Bolivia*, *Botswana*, *Cape Verde*, *China*, *Czech*, Egypt, *Honduras*, *Hungary*, India, *Indonesia*, *Israel*, *Italy*, *Japan*, *Jordan*, Kenya, *Latvia*, *Malawi*, *Malaysia*, Malta, *Mauritius*, *Mexico*, *Panama*, *Paraguay*, *Philippines*, *Romania*, *South Africa*, *Sri Lanka*, Swaziland, Switzerland, *Thailand*, *Ukraine*, United Kingdom
Group 2: (25 countries) *Bangladesh*, *Canada*, *Costa Rica*, *Dominican*, Egypt, *Guyana*, *Iceland*, India, Kenya, *Korea (Rep.)*, *Lithuania*, Malta, *Netherlands*, *Papua New Guinea*, *Peru*, *Russian*, *Singapore*, Swaziland, Switzerland, *Syrian*, *Tanzania*, *Uganda*, United Kingdom, *United States*, *Uruguay*	Group 2: (20 countries) *Bangladesh*, *Canada*, *Costa Rica*, *Dominican Republic*, Guatemala, *Guyana*, *Iceland*, *Korea (Rep.)*, *Lithuania*, Mongolia, *Netherlands*, *Papua New Guinea*, *Peru*, *Russian*, *Singapore*, *Syrian*, *Tanzania*, *Uganda*, *United States*, *Uruguay*

original dependent variable and the lagged dependent variable. For the other two continuously distributed variables, we follow the practice as in the savings rate application to scale-normalize each time series by the individual sample standard deviation. To ensure that the estimated coefficients are comparable, we further multiply these two scale-normalized variables by the overall standard deviation of the lagged dependent variable so that all the explanatory regressors are of the same scale. Furthermore, the Probit regressions for the individual time series are unstable in those countries with only one or two incidences. Therefore, the C-Lasso initial values are set as the pooled FE Probit coefficient estimates.

The summary statistics are displayed in Table SVI. Membership is reported under "high-occurrence" and "low-occurrence" groups with results as follows.

High-occurrence group (23 countries): Guatemala, Peru, Argentina, Mali, Senegal, Chad, Congo (Dem.), Congo (Rep.), Somalia, Morocco, Sudan,

TABLE SVI

SUMMARY STATISTICS FOR THE CIVIL WAR DATA SET

	Mean	Median	S.E.	Min	Max
Civil war incidence	0.352	0	0.478	0	1
GDP per capita growth	0.020	0.024	0.040	−0.811	0.306
Population growth	0.012	0.015	0.076	−0.507	0.661

TABLE SVII
SUMMARY STATISTICS FOR THE DEMOCRACY DATA SET

	Mean	Median	S.E.	Min	Max
Democracy index	0.5760	0.6667	0.3712	0	1
GDP per capita (in logarithm)	8.2981	8.3039	1.0685	6.0937	10.4450

Turkey, Iraq, Lebanon, Afghanistan, China, Pakistan, Sri Lanka, Nepal, Cambodia, Laos, Philippines, Indonesia.

Low-occurrence group (15 countries): Haiti, Dominican Republic, El Salvador, Nicaragua, United Kingdom, Yugoslavia, Cyprus, Russia, Liberia, Nigeria, Central African Republic, Ethiopia, South Africa, Iran, Jordan.

S4.3. *Linear Dynamic Modeling of Democracy*

In this section, we use the data provided by Bonhomme and Manresa (2015) to revisit the link between income growth and democracy across countries. Following BM's Equation (22), we specify a linear dynamic model, where the dependent variable is a country's democracy index (measured by Freedom House indicator between 0 (the lowest) and 1 (the highest)), and the explanatory variables are the first-order lagged democracy index and the income (measured by the logarithm of GDP per capita).

The data set contains a balanced panel of 84 countries and 8 periods at a five-year interval over 1965–2000. The summary statistics are reported in Table SVII. We use PLS to estimate the model in this short panel. Many developed countries, such as the United States or United Kingdom, kept their democracy index at the highest level throughout the time. Due to the lack of within-group variation in these countries, we scale-normalize each variable by its pooled standard deviation. This standardization makes sure that the parameter y_{it-1} can still be interpreted as the autoregressive coefficient, and the magnitude is comparable with the income coefficient.

Following practice in the simulation, the IC with $\rho_{1NT} = \frac{2}{3}(NT)^{-1/2}$ picks out $K = 3$ and $c_{\lambda_1} = 1.20$ in all combinations of $K = 1, \dots, 5$ and c_{λ_1} in a geometrically increasing sequence of 10 points in $(0.2, \dots, 2)$. Under $K = 3$ and $c_{\lambda_1} = 1.20$, C-Lasso categorizes the 84 countries into the following groups:

Group 1 (30 countries): Belgium, Bolivia, Brazil, Canada, Dominican Republic, Ecuador, El Salvador, Finland, Guatemala, Guinea, Iceland, Indonesia, Italy, Japan, Jordan, Luxembourg, Mali, Morocco, Nepal, Panama, Peru, Philippines, Portugal, Romania, South Africa, Thailand, Turkey, United Kingdom, Uruguay, Venezuela.

Group 2 (36 countries): Algeria, Argentina, Australia, Austria, Barbados, Burkina Faso, Burundi, Cameroon, Chile, China, Colombia, Costa Rica, Cote

TABLE SVIII

PLS ESTIMATION RESULTS[a]

	Pooled FE		PLS					
			Group 1		Group 2		Group 3	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Lagged democracy	0.4993***	0.0491	0.5141***	0.0643	0.0954	0.0733	−0.0543	0.0521
Income	0.2552***	0.0489	0.6545***	0.0930	0.1550***	0.0448	−0.5542***	0.0860

[a]Note: *** 1% significant, ** 5% significant, * 10% significant.

d'Ivoire, Denmark, Egypt, France, Gabon, Ghana, Greece, India, Iran, Israel, Jamaica, Kenya, Malawi, Malaysia, Mexico, Nigeria, Norway, Paraguay, Rwanda, Spain, Sweden, Togo, Trinidad and Tobago, United States.

Group 3 (18 countries): Benin, Chad, Congo (Rep.), Honduras, Ireland, Korea (Rep.), Madagascar, Netherlands, New Zealand, Nicaragua, Niger, Sri Lanka, Switzerland, Syrian, Tanzania, Tunisia, Uganda, Zambia.

The post-Lasso and pooled FE estimates are shown in Table SVIII. We focus on the coefficient for income. The common FE coefficient is positive and significant. The positive effect is echoed by Groups 1 and 2, but contrasts with Group 3, which consists mainly of low-income and low-democracy nations combined with a few selected OECD countries. OECD countries such as Ireland, Netherlands, New Zealand, and Switzerland maintained their democracy index at 1 throughout the sample period. The lack of variation in the dependent variable makes them uninformative about the income coefficient.

REFERENCES

BLATTMAN, C., AND E. MIGUEL (2010): "Civil Wars," *Journal of Economic Literature*, 48, 3–57. [61]

BONHOMME, S., AND E. MANRESA (2015): "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica*, 83, 1147–1184. [63]

COLLIER, P., AND A. HOEFFLER (2004): "Greed and Grievance in Civil Wars," *Oxford Economic Papers*, 56, 563–595. [61]

DJANKOV, S., AND M. REYNAL-QUEROL (2010): "Poverty and Civil War: Revisiting the Evidence," *Review of Economics and Statistics*, 92, 1035–1041. [61]

EFRON, B., T. HASTIE, I. JOHNSTONE, AND R. TIBSHIRANI (2004): "Least Angle Regression," *The Annals of Statistics*, 32, 407–499. [55]

FEARON, J. D., AND D. D. LAITIN (2003): "Ethnicity, Insurgency, and Civil War," *The American Political Science Review*, 97, 75–90. [61]

GAO, J. (2007): *Nonlinear Time Series: Semiparametric and Nonparametric Methods,*, New York: Chapman & Hall/CRC. [41,42]

HAHN, J., AND G. KUERSTEINER (2011): "Bias Reduction for Dynamic Nonlinear Panel Models With Fixed Effects," *Econometric Theory*, 27, 1152–1191. [29]

HAHN, J., AND W. NEWEY (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72, 1295–1319. [14]

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Applications*. New York: Academic Press. [8,12]

MERLEVÈDE, F., M. PEILGRAD, AND E. RIO (2009): "Bernstein Inequality and Moderate Deviations Under Strong Mixing Conditions," in *IMS Collections: High Dimensional Probability*, Vol. V. Beachwood, OH: Inst. Math. Statist., 273–292. [1]

WHITE, H. (2001): *Asymptotic Theory for Econometricians*, London: Emerald. [29]

School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903, Singapore; ljsu@smu.edu.sg,

Dept. of Economics, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China; zhentao.shi@cuhk.edu.hk,

and

Cowles Foundation for Research in Economics, Yale University, P.O. Box 208281, New Haven, CT 06520, U.S.A., University of Auckland, University of Southampton, and Singapore Management University; peter.phillips@yale.edu.

Co-editor Elie Tamer handled this manuscript.

Manuscript received June, 2014; final revision received December, 2015.