Supplement to "Confidence set for group membership"

(*Quantitative Economics*, Vol. 15, No. 2, May 2024, 245–277)

Andreas Dzemski
Department of Economics, University of Gothenburg

Ryo Okui
Graduate School of Economics and Faculty of Economics, University of Tokyo

Appendix A: Proofs of main results

A.1 *Notation*

We introduce additional notation. We consider statistics that replace some (not all) estimated components in the statistics defined in the main text with population quantities. Let

$$D_i(h) = \frac{\sum_{t=1}^{T} d_{it}(g_i^0, h)/\sqrt{T}}{\sqrt{\Xi_i(h, h)}}$$

and

$$\widetilde{H}_{ij}(h, h') = \frac{1}{T} \sum_{t=|j|+1}^{T} \big(d_{i,t+\min(0,j)}(g_i^0, h) - \bar{d}_i(g_i^0, h)\big)\big(d_{i,t-\max(0,j)}(g_i^0, h') - \bar{d}_i(g_i^0, h')\big),$$

where $\bar{d}_i(g_i^0, h) = \sum_{t=1}^{T} d_{it}(g^0, h)/T$. Let

$$\widetilde{\Xi}_i(h, h') = \sum_{j=-T+1}^{T-1} K\left(\frac{j}{\kappa_N}\right) \widetilde{H}_{ij}(h, h').$$

Let

$$\widetilde{D}_i(h) = \frac{\sum_{t=1}^{T} d_{it}(g_i^0, h)/\sqrt{T}}{\sqrt{\widetilde{\Xi}_i(h, h)}},$$

Andreas Dzemski: andreas.dzemski@economics.gu.se
Ryo Okui: okuiryo@e.u-tokyo.ac.jp

Available at http://qeconomics.org. https://doi.org/10.3982/QE1975

and let $\widetilde{\Omega}_i(g_i^0)$ denote the $(G-1)\times(G-1)$ matrix with entries

$$\big(\widetilde{\Omega}_i(g_i^0)\big)_{j,j'} = \frac{\widetilde{\Xi}_i(h,h')}{\sqrt{\widetilde{\Xi}_i(h,h)\widetilde{\Xi}_i(h',h')}}.$$

We write $d_{it}(h) = d_{it}(g_i^0, h)$.

A.2 *Proofs*

Proof of Theorem 1. The result follows directly from Lemma A.1. □

Proof of Theorem 2. Write $k_{\alpha,N}(\Omega)$ for the $1-\alpha/N$-quantile of a $N(0,\Omega)$ random variable. Abbreviate $\Omega_i = \Omega_i(g_i^0)$ and $\widehat{\Omega}_i = \widehat{\Omega}_i(g_i^0)$.

Let ζ_N denote a diverging sequence, $\zeta_N \to \infty$. For $\bar{\alpha}_N = \alpha(1+2C_{\text{A.3}}\sqrt{\epsilon_N \log(N/\alpha)})$ and $C_{\text{A.3}}$ the constant from Lemma A.3, we first establish the following chain of inequalities:

$$k_{\bar{\alpha}_N,N}(\Omega_i) \le k_{\alpha,N}\big(\rho(\widehat{\Omega}_i,\epsilon_N)\big) \le c_{\alpha,N}\big(\rho(\widehat{\Omega}_i,\epsilon_N)\big). \tag{A.1}$$

To prove the second inequality, let $t_{T-1}(\cdot)$ denote the cumulative distribution function of a t-distributed random variable with $(T-1)$-degrees of freedom and let X denote a $(G-1)$ random vector distributed according to centered multivariate t-distribution with $T-1$ degrees of freedom and scale matrix $\rho(\widehat{\Omega}_i,\epsilon_N)$. The marginal distribution of the first component of X, denoted by X_1, is $X_1 \sim t_{T-1}$. Let $d_N = t_{T-1}^{-1}(1-\alpha/N)$ and note that $d_N \to \infty$. Moreover,

$$\alpha/N = P(X_1 > d_N) \le P\Big(\max_{h\in 1,\dots,G-1} X_h > d_N\Big).$$

Therefore, $c_{\alpha,N}(\rho(\widehat{\Omega}_i,\epsilon_N)) \ge \sqrt{T/(T-1)}d_N$ and for N_0 and T_0 independent of $\rho(\widehat{\Omega}_i(g_i^0),\epsilon_N)$ and t^*, the constant defined in Lemma A.2 we can take

$$c_{\alpha,N}\big(\rho(\widehat{\Omega}_i,\epsilon_N)\big) > t^*,$$

for all $N \ge N_0$, $T \ge T_0$. Therefore, the assumptions of Lemma A.2 are satisfied for $t = c_{\alpha,N}(\rho(\widehat{\Omega}_i,\epsilon_N))$ and N, T large enough and Lemma A.2 implies

$$\begin{aligned}
&\Phi_{\max,\rho(\widehat{\Omega}_i,\epsilon_N)}\big(k_{\alpha,N}\big(\rho(\widehat{\Omega}_i,\epsilon_N)\big)\big)\\
&\quad = 1-\alpha/N\\
&\quad = t_{\max,\rho(\widehat{\Omega}_i,\epsilon_N),T-1}\big(\sqrt{(T-1)/T}\,c_{\alpha,N}\big(\rho(\widehat{\Omega}_i,\epsilon_N)\big)\big)\\
&\quad \le \Phi_{\max,\rho(\widehat{\Omega}_i,\epsilon_N)}\big(c_{\alpha,N}\big(\rho(\widehat{\Omega}_i,\epsilon_N)\big)\big)
\end{aligned}$$

and, therefore, $k_{\alpha,N}(\rho(\widehat{\Omega}_i,\epsilon_N)) \le c_{\alpha,N}(\rho(\widehat{\Omega}_i,\epsilon_N))$. We now establish the first inequality in (A.1). Note that

$$\alpha_N \equiv \alpha\big(1 + C_{\text{A.3}}c_{\alpha,N}\big(\rho(\widehat{\Omega}_i,\epsilon_N)\big)\big) \le \alpha\big(1+2C_{\text{A.3}}\sqrt{\log(N/\alpha)}\big)$$

by Lemma A.10 and that $k_{\alpha,N}$ is decreasing in α. Therefore, it suffices to establish

$$k_{\alpha_N,N}(\Omega_i) \le k_{\alpha,N}\big(\rho(\widehat{\Omega}_i, \epsilon_N)\big). \tag{A.2}$$

It can be established by applying Lemma A.3. Lemma A.4 yields

$$\|\widehat{\Omega}_i - \Omega_i\|_{\max} \le \zeta_N \left(\frac{r_{\theta,N}}{\iota_N \min\limits_{1\le i\le N} \sigma_i} + T^{-c_1}(\log N)^{c_2} + T^{-\rho} \right) \equiv \delta_N,$$

on an event whose probability approaches one. On this set, Lemma A.10 states that we can take $k_{\alpha,N}(\rho(\widehat{\Omega}_i, \epsilon_N)) > \sqrt{\log(N/\alpha)} > \sqrt{2}$ for N large enough. We now show that $\delta_N/\epsilon_N \to 0$. Since ζ_N can be taken to diverge at an arbitrarily slow rate, it suffices to show that, for $\pi > 0$,

$$\begin{aligned} r_{\theta,N}\Big/\Big(\iota_N \min_{1\le i\le N} \sigma_i\Big) &\le \epsilon_N T^{-\pi}, \\ T^{-c_1}(\log N)^{c_2} &\le \epsilon_N T^{-\pi}, \\ T^{-\rho} &\le \epsilon_N T^{-\pi}. \end{aligned}$$

Since $r_{\theta,N}\sqrt{T\log N} = o(\iota_N \min_{1\le i\le N} \sigma_i)$ and $\epsilon_N \ge (\log N)^{-k_2}$, the first condition is met provided

$$T^{-1+2\pi}(\log N)^{-1+2k_2} = O(1).$$

Under the assumed rate condition $N \le o(1)T_2^{\delta}$, this holds for any $0 < \pi < \frac{1}{2}$. The second and third conditions can be checked similarly. Now that we have established $\delta_N/\epsilon_N \to 0$ we can take $4\delta_N \le \epsilon_N$, satisfying one of the conditions of Lemma A.3. The condition $\epsilon_N \le 4c_\omega/3$ is assumed. This argument verifies all conditions of Lemma A.3 and yields inequality (A.2) and, therefore, (A.1).

Our assumptions,

$$r_{\theta,N}\sqrt{T\log N}\Big/\Big(\iota_N \min_{1\le i\le N} \sigma_i\Big) \to 0$$

and $N \le o(1)T^{\delta_2}$ guarantee that $(\sqrt{T} \vee \log N) r_{\theta,N}/(\iota_N \min_{1\le i\le N} \sigma_i)$ and $T^{-c_1}(\log N)^{c_2}$ vanish. Therefore,

$$b_N^{\mathrm{LV}*} = \frac{(\sqrt{T} \vee \log N) r_{\theta,N}}{\iota_N \min\limits_{1\le i\le N} \sigma_i} + T^{-c_1}(\log N)^{c_2} + T^{-\rho}$$

vanishes and Lemma A.14 can be applied and yields

$$P\Big(\max_{1\le i\le N} \max_{h\in\mathbb{G}\setminus\{g_i^0\}} \big|\widehat{D}_i(h) - D_i(h)\big| > \zeta_N b_N^{\mathrm{LV}*}\Big) = o(1). \tag{A.3}$$

We now prepare to apply the high-dimensional CLT in Lemma A.13. Let $\delta_i(h) = \theta_{g_i^0} - \theta_h$ and

$$X_{it}(h) = \frac{d_{it}(h)}{\sqrt{T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}_p\big[d_{it}(h)d_{is}(h)\big]}} = -\frac{x_{it}'\big(\delta_i(h)/\|\delta_i(h)\|\big)v_{it}}{\xi_i(h)/\big(\sigma_i\|\delta_i(h)\|\big)},$$

where

$$\xi_i(h) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}\big[d_{it}(h)d_{is}(h)\big]} = \sqrt{\delta_i(h)'\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}_P\big[x_{it}x_{is}'u_{it}u_{is}\big]\right)\delta_i(h)}.$$

Define the vector

$$X_t = \big(\big(X_{1t}(h)\big)_{h\in\mathbb{G}\setminus\{g_1^0\}}, \dots, \big(X_{Nt}(h)\big)_{h\in\mathbb{G}\setminus\{g_N^0\}}\big)'.$$

This vector has length $J = N(G-1)$. Let Ξ denote the long-run variance of the time series X_t defined as the $J \times J$ matrix with entry (j,k) given by

$$\operatorname{cov}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_{t,j}, \frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_{t,k}\right).$$

Let G denote a centered normal vector with variance matrix Ξ. Clearly, X_i is a normal random vector with covariance matrix Ω_i. Taking complements, we have

$$\begin{aligned}
&\sup_{(r_1,\dots,r_N)\in\mathbb{R}^N_{++}}\Big|P\Big(\max_{1\le i\le N}\Big(\max_{h\in\mathbb{G}\setminus\{g_i^0\}} D_i(h) - r_i\Big) > 0\Big) \\
&\quad - P\Big(\max_{1\le i\le N}\Big(\max_{1\le h\le G-1} X_{i,h} - r_i\Big) > 0\Big)\Big| \\
&= \sup_{(r_1,\dots,r_N)\in\mathbb{R}^N_{++}}\big|P\big(D_i(h)\le r_i \text{ for all } h\in\mathbb{G}\setminus\{g_i^0\} \text{ and } i=1,\dots,N\big) \\
&\quad - P(X_{i,h}\le r_i \text{ for all } h=1,\dots,G-1 \text{ and } i=1,\dots,N)\big| \\
&\le \sup_{a\in\mathbb{R}^{(G-1)N}}\left|P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Z_t\le a\right) - P(X\le a)\right| \\
&\le C\left(\frac{(\log N)^{(1+2d_1)/(3d_1)}}{T^{1/9}} + \frac{(\log N)^{7/6}}{T^{1/9}}\right) = o(1).
\end{aligned} \tag{A.4}$$

Here, the last inequality holds by Lemma A.13 and the asymptotic order follows from $N \le o(1)T^{\delta_2}$. Now, we have

$$\begin{aligned}
&P\big(\exists i\in 1,\dots,N \text{ such that } \hat{T}_i(g_i^0) > c_{\alpha,N}(\widehat{\Omega}_i)\big) \\
&\quad\le P\big(\exists i\in 1,\dots,N \text{ such that } \hat{T}_i\big(g_i^0\big) > k_{\bar{\alpha}_N,N}(\Omega_i)\big)
\end{aligned}$$

$$
\begin{aligned}
&\le P\Big(\max_{1\le i\le N}\max_{h\in\mathbb{G}\setminus\{g_i^0\}}\big(D_i(h)-k_{\bar\alpha_N,N}(\Omega_i)+\zeta_N b_N^{\mathrm{LV}*}\big)>0\Big)+o(1).\\
&\le P\Big(\max_{1\le i\le N}\Big(\max_{1\le h\le G-1}X_{i,h}-k_{\bar\alpha_N,N}(\Omega_i)+\zeta_N b_N^{\mathrm{LV}*}\Big)>0\Big)\\
&\quad+\sup_{(r_1,\dots,r_N)\in\mathbb{R}^N_{++}}\Big|P\Big(\max_{1\le i\le N}\Big(\max_{h\in\mathbb{G}\setminus\{g_i^0\}}D_i(h)-r_i\Big)>0\Big)\\
&\quad-P\Big(\max_{1\le i\le N}\Big(\max_{1\le h\le G-1}X_{i,h}-r_i\Big)>0\Big)\Big|+o(1)\\
&\le P\Big(\max_{1\le i\le N}\Big(\max_{1\le h\le G-1}X_{i,h}-k_{\bar\alpha_N,N}(\Omega_i)+\zeta_N b_N^{\mathrm{LV}*}\Big)>0\Big)+o(1),
\end{aligned}
\tag{A.5}
$$

where the first inequality follows by (A.1), the second inequality follows by (A.3), the third inequality holds because the sup bounds deviations for all choices of r_i and, therefore, in particular $r_i=k_{\bar\alpha_N,N}(\Omega_i)-\zeta_N b_N^{\mathrm{LV}*}$, and the fourth inequality follows by (A.4).

Next, applying an anticoncentration argument eliminates the $\zeta_N b_N^{\mathrm{LV}*}$ term on the right-hand side of the previous display. To this end, let $a>0$ and write $k_{N,i}=k_{\bar\alpha_N,N}(\Omega_i)$. Then

$$
\begin{aligned}
&P\Big(\max_{1\le i\le N}\Big(\max_{1\le h\le G-1}X_{i,h}-k_{N,i}\Big)+a>0\Big)-P\Big(\max_{1\le i\le N}\Big(\max_{1\le h\le G-1}X_{i,h}-k_{N,i}\Big)>0\Big)\\
&=P\Big(\max_{1\le i\le N}\Big(\max_{1\le h\le G-1}X_{i,h}-k_{N,i}\Big)\le 0\Big)-P\Big(\max_{1\le i\le N}\Big(\max_{1\le h\le G-1}X_{i,h}-k_{N,i}\Big)\le -a\Big)\\
&=P(X\le x+a)-P(X\le x),
\end{aligned}
$$

where

$$
x=(\underbrace{k_{N,1}-a,\dots,k_{N,1}-a}_{G-1\text{ times}},\underbrace{k_{N,2}-a,\dots,k_{N,2}-a}_{G-1\text{ times}},\dots,\underbrace{k_{N,N}-a,\dots,k_{N,N}-a}_{G-1\text{ times}})
$$

The Nasarov-type inequality from Lemma A.1 in Chernozhukov, Chetverikov, and Kato (2017) applies with $b=1$ and $p=(G-1)N$ and yields

$$
P(X\le x+a)-P(X\le x)\le C_{\text{Nasarov}}a\sqrt{\log\big(N(G-1)\big)}\le O(1)a\log N.
$$

Therefore,

$$
\begin{aligned}
&P\Big(\max_{1\le i\le N}\Big(\max_{1\le h\le G-1}X_{i,h}-k_{N,i}\Big)+a>0\Big)\\
&\quad\le P\Big(\max_{1\le i\le N}\Big(\max_{1\le h\le G-1}X_{i,h}-k_{N,i}\Big)>0\Big)+O(1)a\sqrt{\log N}
\end{aligned}
$$

Now, combining this inequality with (A.5) by putting $a=\zeta_N b_N^{\mathrm{LV}*}$ yields

$$
\begin{aligned}
&P\Big(\max_{1\le i\le N}\max_{h\in\mathbb{G}\setminus\{g_i^0\}}D_i(h)-k_{\bar\alpha_N,N}(\Omega_i)+\zeta_N b_N^{\mathrm{LV}*}>0\Big)\\
&\quad\le P\Big(\max_{1\le i\le N}\Big(\max_{1\le h\le G-1}X_{i,h}-k_{\bar\alpha_N,N}(\Omega_i)\Big)+\zeta_N b_N^{\mathrm{LV}*}>0\Big)+o(1)
\end{aligned}
$$

$$\leq P\Big(\max_{1\leq i\leq N}\Big(\max_{1\leq h\leq G-1} X_{i,h} - k_{\bar{\alpha}_N,N}(\Omega_i)\Big) > 0\Big) + O(1)\zeta_N b_N^{\mathrm{LV}*}\sqrt{\log N} + o(1)$$

$$\leq \sum_{1\leq i\leq N} P\Big(\max_{1\leq h\leq G-1} X_{i,h} - k_{\bar{\alpha}_N,N}(\Omega_i) > 0\Big) + o(1)$$

$$= \bar{\alpha}_N + o(1)$$

$$= \alpha + 2C_{\mathrm{A.3}}\sqrt{\epsilon_N \log(N/\alpha)} + o(1) \to \alpha. \qquad \square$$

Proof of Theorem 3. Since the marginal distributions of a multivariate t-distribution with ν degrees of freedom are Student-t with ν degrees of freedom, it holds that

$$(t_{\max,\rho(\widehat{\Omega}_i(g_i^0),\epsilon_N),T-1})^{-1}\left(1-\frac{\alpha}{N}\right) \leq t_{T-1}^{-1}\left(1-\frac{\alpha}{(G-1)N}\right)$$

by the union bound. Thus, replacing our critical value with the SNS critical values yields a more conservative test. Now, inspection of the proof of Theorem 2 shows that Assumption 1.9 is only used to argue that $\widehat{\Omega}_i(g_i^0)$ in the definition of the critical value can be replaced by the population quantity $\Omega_i(g_i^0)$. Since we are replacing the critical value that depends on $\widehat{\Omega}_i(g_i^0)$ by a critical value that is independent of $\widehat{\Omega}_i(g_i^0)$, this step is not needed. Similarly, the SNS critical value is independent of the regularization sequence, and the assumptions on ϵ_N are therefore unnecessary. $\square$

Proof of Theorem 4. The proof is similar to the proof of Theorem 2, replacing the application of Lemma A.4 and Lemma A.14 by and application of Lemma A.16. $\square$

Proof of Theorem 5. Let

$$d_{it}^U(g,h) = \big(y_{it} - w_{it}'\theta^w - x_{it}'\theta_g\big)^2 - \big(y_{it} - w_{it}'\theta^w - x_{it}'\theta_h\big)^2$$

and $d_i^U(h) = d_{it}^U(g_i^0,h)$, $\hat{d}_i^U(h) = \hat{d}_i^U(g_i^0,h)$, $\bar{d}_i^U(h) = N^{-1}\sum_{t=1}^T d_{it}^U(h)$, and $\bar{\hat{d}}_i^U(h) = N^{-1}\sum_{t=1}^T \hat{d}_{it}^U(h)$. We note that the hypothesis selection part of the procedure does not affect the theoretical analysis. This is because, here, we focus on size, and thus need to consider only the behavior of the test statistics under $\{g_i^0\}_{i=1}^N$.

In the following, $o(1)$ is understood such that $a = o(1)$ if $\limsup_{N,T\to\infty}|a| = 0$.

Let $J = \{(i,h) \mid i\in\{1,\ldots,N\}, h\in\mathbb{G}\backslash\{g_i^0\}\}$ and

$$J_1 = \left\{(i,h) \mid i\in\{1,\ldots,N\}, h\in\mathbb{G}\backslash\{g_i^0\}, \frac{\sqrt{T}\mathbb{E}_P\big(\bar{d}_i^U(h)\big)}{s_{i,T}^U(h)} > -c_{\beta,N}^{\mathrm{SNS}}\right\},$$

where $(s_i^U(h))^2 = \sum_{t=1}^T \mathrm{var}(d_{it}^U(h))/T = \mathrm{var}(d_{it}^U(h))$ (the equality follows by stationarity). Roughly speaking, J_1 is the set of pairs of units and groups that are difficult to distinguish from the true group membership.

Step 1 We first prove that $\inf_{P\in\mathbb{P}_N} P(\max_{(i,h)\in J_1^c} \hat{\bar{d}}_i^U(h) \le 0) > 1-\beta-N^{-1}-CT^{-c}-a_{\theta,N}$. Note that $\hat{\bar{d}}_i^U(h) > 0$ for some $(i,h)\in J_1^c$ implies that

$$\max_{(i,h)\in J} \frac{\sqrt{T}\big(\hat{\bar{d}}_i^U(h) - \mathbb{E}_P\big(\bar{d}_i^U(h)\big)\big)}{s_{i,T}^U(h)} > c_{\beta,N}^{\text{SNS}}.$$

We observe that

$$\begin{aligned}
&\sup_{P\in\mathbb{P}_N} P\bigg(\max_{(i,h)\in J} \frac{\sqrt{T}\big(\hat{\bar{d}}_i^U(h) - \mathbb{E}_P\big(\bar{d}_i^U(h)\big)\big)}{s_i^U(h)} > c_{\beta,N}^{\text{SNS}}\bigg)\\
&\quad\le \sup_{P\in\mathbb{P}_N} P\bigg(\max_{(i,h)\in J} \frac{\sqrt{T}\big(\bar{d}_i^U(h) - \mathbb{E}_P\big(\bar{d}_i^U(h)\big)\big)}{s_i^U(h)} > c_{\beta,N}^{\text{SNS}} - e_{N,1}^U\bigg)\\
&\qquad+ \sup_{P\in\mathbb{P}_N} P\bigg(\max_{(i,h)\in J}\bigg|\frac{\sqrt{T}\big(\hat{\bar{d}}_i^U(h) - \bar{d}_i^U(h)\big)}{s_i^U(h)}\bigg| > e_{N,1}^U\bigg),
\end{aligned}$$

where

$$e_{N,1}^U = C\frac{r_{\theta,N}}{\iota_N + \min\limits_{1\le i\le N} \sigma_i}.$$

The second term on the right-hand side converges to zero by (A.19) in Lemma A.19. Let β_N solve $c_{\beta_N,N}^{\text{SNS}} = c_{\beta,N}^{\text{SNS}} - e_{N,1}^U$. To see that β_N is well-defined, note that since $c_{\beta,N}^{\text{SNS}} \to \infty$ and $e_{N,1}^U \to 0$, the right-hand side of the equation is diverging and, therefore, positive for large N. Moreover, $c_{p,N}^{\text{SNS}} \downarrow 0$ as $p \uparrow N/2$. We thus establish the existence of β_N. Uniqueness follows from the strict monotonicity of the distribution function of the t-distribution. Thus, we have

$$\begin{aligned}
&\sup_{P\in\mathbb{P}_N} P\bigg(\max_{(i,h)\in J} \frac{\sqrt{T}\big(\hat{\bar{d}}_i^U(h)\big) - \mathbb{E}_P\big(\bar{d}_i^U(h)\big)\big)}{s_i^U(h)} > c_{\beta,N}^{\text{SNS}}\bigg)\\
&\quad\le \sup_{P\in\mathbb{P}_N} P\bigg(\max_{(i,h)\in J} \frac{\sqrt{T}\big(\bar{d}_i^U(h)\big) - \mathbb{E}_P\big(\bar{d}_i^U(h)\big)\big)}{s_i^U(h)} > c_{\beta_N,N}^{\text{SNS}}\bigg) + o(1)\\
&\quad\le 1-(G-1)N\Phi\big(c_{\beta_N,N}^{\text{SNS}}\big) + o(1)\\
&\quad\le \beta_N + o(1)\\
&\quad= \beta + o(1),
\end{aligned}$$

where the second inequality follows by Lemma A.13 and the Bonferroni inequality, the third inequality holds because $c_{\beta_N,N}^{\text{SNS}}$ becomes sufficiently large as $N\to\infty$ and the tail of the t-distribution is heavier than that of the standard normal distribution (Lemma A.2 under the unidimensional case), and the last inequality follows by the fact that $|\beta_N - \beta| \le Ce_{N,1}^U\sqrt{\log((G-1)N/\beta)} \to 0$. We now show that $|\beta_N - \beta| \le$

$Ce^U_{N,1}\sqrt{\log((G-1)N/\beta)}$. Let F_T denote the distribution function of a t-distributed random variable with $T-1$ degrees of freedom, and let f_T denote its density function. Let $c(\beta)=t^{-1}_{T-1}(1-\beta/((G-1)N))$ and $e^{U*}_{N,1}=\sqrt{(T-1)/T}e^U_{N,1}$. By the mean-value theorem,

$$\begin{aligned}\frac{\beta_N}{(G-1)N}-\frac{\beta}{(G-1)N}&=F_T\big(c(\beta)\big)-F_T\big(c(\beta_N)\big)\\&=F_T\big(c(\beta)\big)-F_T\big(c(\beta)-e^U_{N,1}\big)=f_T\big(c^*\big)e^{U*}_{N,1},\end{aligned}$$

where c^* is a value between $c(\beta_N)$ and $c(\beta)$. Noting that $c(\beta_N)<c(\beta)$ and that f_T is decreasing on the positive axis, rearranging this equality yields

$$\begin{aligned}|\beta_N-\beta|&\le f_T\big(c(\beta_N)\big)(G-1)Ne^{U*}_{N,1}\\&\le 2c(\beta_N)(1-F_T\big(c(\beta_N)\big)(G-1)Ne^{U*}_{N,1}\\&\le 4e^{U*}_{N,1}\beta_N\sqrt{\log\big((G-1)N/\beta_N\big)}\\&\le 4e^U_{N,1}\beta\sqrt{\log\big((G-1)N/\beta\big)}+4e^U_{N,1}|\beta_N-\beta|\sqrt{\log\big((G-1)N/\beta\big)}\\&\le 4e^U_{N,1}\sqrt{\log\big((G-1)N/\beta\big)}+o\big(|\beta_N-\beta|\big),\end{aligned}$$

where the second inequality follows from Lemma A.18, the third inequality follows from Lemma A.17 (with $\epsilon_N=1$), the fourth inequality follows from $e^U_{N,1}=\sqrt{T/(T-1)}e^{U*}_{N,1}\ge e^{U*}_{N,1}$ and $\beta_N\ge\beta$, the fifth inequality follows from $e^U_{N,1}\sqrt{\log N}\to 0$. This recursion implies

$$|\beta_N-\beta|\le 5e^U_{N,1}\sqrt{\log\big((G-1)N/\beta\big)}$$

for N large enough.

An implication of Step 1 is as follows. Let

$$\mathbb{N}=\left\{i\in\{1,\dots,N\}\;\middle|\;\max_{h\in\mathbb{G}\setminus\{g^0_i\}}\frac{\sqrt{T}\mathbb{E}_P\big(\bar d^U_i(h)\big)}{s^U_i(h)}>-c^{\mathrm{SNS}}_{\beta,N}\right\}.$$

Then

$$\inf_{P\in\mathbb{P}_N}P\Big(\max_{i\in\mathbb{N}^c}\max_{h\in\mathbb{G}\setminus\{g^0_i\}}\bar{\hat d}^U_i(h)\le 0\Big)\ge 1-\beta+o(1).$$

Step 2 Next, we prove that $\inf_{P\in\mathbb{P}_N}P(\prod_{i=1}^N\hat M_i(g^0_i)\supseteq J_1)\ge 1-\beta+o(1)$, where $\prod_{i=1}^N$ denotes the N-fold Cartesian product over the sets indexed by i. Here, we drop the g argument for simplicity of notation when arguments are g^0_i and h.

We note that

$$\begin{aligned}&\sup_{P\in\mathbb{P}_N}P\left(\prod_{i=1}^N\widehat M_i\big(g^0_i\big)\not\supseteq J_1\right)\\&=\sup_{P\in\mathbb{P}_N}P\left(\exists(i,h);\,\widehat D^U_i(h)\le -2c^{\mathrm{SNS}}_{\beta_N,N}\text{ and }\frac{\sqrt{T}\mathbb{E}_P\big(\bar d^U_i(h)\big)}{s^U_i(h)}>-c^{\mathrm{SNS}}_{\beta,N}\right)\end{aligned}$$

$$
\begin{aligned}
&\le \sup_{P\in\mathbb{P}_N} P\Big(\exists(i,h);\, D_i^U(h) \le -2c_{\beta,N}^{\text{SNS}} + e_{N,2}^U \text{ and } \frac{\sqrt{T}\mathbb{E}_P\big(\bar{d}_i^U(h)\big)}{s_i^U(h)} > -c_{\beta,N}^{\text{SNS}}\Big) \\
&\quad + \sup_{P\in\mathbb{P}_N} P\Big(\max_{1\le i\le N}\max_{h\in\mathbb{G}\setminus\{g_i^0\}} \big|\widehat{D}_i^U(h) - D_i^U(h)\big| > e_{N,2}^U\Big),
\end{aligned}
$$

where

$$
e_{N,2}^U = C\frac{r_{\theta,N}(\sqrt{T}+\sqrt{\log N})}{\iota_N \wedge \min\limits_{1\le i\le N}\sigma_i}.
$$

By (A.20) in Lemma A.19, it holds that

$$
\sup_{P\in\mathbb{P}_N} P\Big(\max_{1\le i\le N}\max_{h\in\mathbb{G}\setminus\{g_i^0\}} \big|\widehat{D}_i^U(h) - \tilde{D}_i^U(h)\big| > e_{N,2}^U\Big) = o(1).
$$

We observe

$$
\begin{aligned}
&\sup_{P\in\mathbb{P}_N} P\Big(\exists(i,h);\, D_i^U(h) \le -2c_{\beta,N}^{\text{SNS}} + e_{N,2}^U \text{ and } \frac{\sqrt{T}\mathbb{E}_P\big(\bar{d}_i^U(h)\big)}{s_i^U(h)} > -c_{\beta,N}^{\text{SNS}}\Big) \\
&\quad \le \sup_{P\in\mathbb{P}_N} P\Big(\max_{1\le i\le N}\max_{h\in\mathbb{G}\setminus\{g_i^0\}} \frac{\sqrt{T}(\mathbb{E}_P\big(\bar{d}_i^U(h) - \bar{d}_i^U(h)\big)}{s_{i,T}^U(h)} \\
&\qquad > \frac{2\tilde{s}_i^U(h) - s_i^U(h)}{s_i^U(h)}c_{\beta,N}^{\text{SNS}} - \frac{\tilde{s}_i^U(h)}{s_i^U(h)}e_{N,2}^U\Big).
\end{aligned}
$$

Note that $\tilde{s}_i^U(h)s_i^U(h) > 1 - r/2$ is equivalent to

$$
\frac{2\tilde{s}_i^U(h) - s_i^U(h)}{s_i^U(h)} > 1 - r.
$$

Thus, we have

$$
\begin{aligned}
&\sup_{P\in\mathbb{P}_N} P\Big(\max_{1\le i\le N}\max_{h\in\mathbb{G}\setminus\{g_i^0\}} \frac{\sqrt{T}(\mathbb{E}_P\big(\bar{d}_i^U(h) - \bar{d}_i^U(h)\big)}{s_{i,T}^U(h)} > \frac{2\tilde{s}_i^U(h) - s_i^U(h)}{s_i^U(h)}c_{\beta,N}^{\text{SNS}} - \frac{\tilde{s}_i^U(h)}{s_i^U(h)}e_{N,2}^U\Big) \\
&\le \sup_{P\in\mathbb{P}_N} P(\max_{1\le i\le N}\max_{h\in\mathbb{G}\setminus\{g_i^0\}} \frac{\sqrt{T}\big(\mathbb{E}_P\big(\bar{d}_i^U(h) - \bar{d}_i^U(h)\big)}{s_{i,T}^U(h)} > (1-r)c_{\beta,N}^{\text{SNS}} - Ae_{N,2}^U) \\
&\quad + \sup_{P\in\mathbb{P}_N} P\Big(\Big|\frac{\tilde{s}_i^U(h)}{s_i^U(h)} - 1\Big| > r/2\Big) \\
&\quad + \sup_{P\in\mathbb{P}_N} P\Big(\max_{1\le i\le N}\max_{h\in\mathbb{G}\setminus\{g_i^0\}} \Big|\frac{\tilde{s}_i^U(h)}{s_i^U(h)}\Big| > A\Big),
\end{aligned}
$$

where $A > 1$ is a fixed number. We now note that

$$\begin{aligned}
&\tilde{s}_i^U(h)^2 - s_i^U(h)^2 \\
&\quad = \frac{1}{T}\sum_{t=1}^{T}\big\{\big(d_{it}^U(h)\big)^2 - \mathbb{E}_P\big[\big(d_{it}^U(h)\big)^2\big)\big] - \big(\bar{d}_i^U(h) - \mathbb{E}_P\big[d_{it}^U(h)\big]\big)\big(\bar{d}_i^U(h) + \mathbb{E}_P\big[d_{it}^U(h)\big]\big\}.
\end{aligned}$$

By Lemma A.15 and (A.24), it holds that

$$\sup_{P\in\mathbb{P}}\left(\max_{1\le i\le N}\left|\frac{\tilde{s}_i^U(h)^2 - s_i^U(h)^2}{\sigma_i^2\left\|\delta_i(h)\right\|^2}\right| \ge CT^{-1/2}\log N\right) = o(1). \tag{A.6}$$

Because $s_i^U(h) > s_i(h)$ and $s_i(h)/(\sigma_i\|\delta_i(h)\|)$ is bounded from above and from below by Assumption 1.4, it holds that

$$\sup_{P\in\mathbb{P}}\left(\max_{1\le i\le N}\left|\frac{\tilde{s}_i^U(h)}{s_i^U(h)} - 1\right| \ge CT^{-1/4}\sqrt{\log N}\right) = o(1). \tag{A.7}$$

We take $r = T^{-1/4}\sqrt{\log N}$. We then have

$$\sup_{P\in\mathbb{P}_N} P\left(\left|\frac{\tilde{s}_i^U(h)}{s_i^U(h)} - 1\right| > r/2\right) + \sup_{P\in\mathbb{P}_N} P\left(\max_{1\le i\le N}\max_{h\in\mathbb{G}\setminus\{g_i^0\}}\left|\frac{\tilde{s}_i^U(h)}{s_i^U(h)}\right| > A\right) = o(1).$$

Let β_N' be such that $c_{\beta_N',N}^{\text{SNS}} = (1-r)c_{\beta,N}^{\text{SNS}} - Ae_{N,2}^U$. We then examine

$$\begin{aligned}
&\sup_{P\in\mathbb{P}_N} P\big(\max_{1\le i\le N}\max_{h\in\mathbb{G}\setminus\{g_i^0\}}\frac{\sqrt{T}\big(\mathbb{E}_P\big(\bar{d}_i^U(h) - \bar{d}_i^U(h)\big)}{s_{i,T}^U(h)} > (1-r)c_{\beta,N}^{\text{SNS}} - Ae_{N,2}^U\big) \\
&\quad = \sup_{P\in\mathbb{P}_N} P\big(\max_{1\le i\le N}\max_{h\in\mathbb{G}\setminus\{g_i^0\}}\frac{\sqrt{T}\big(\mathbb{E}_P\big(\bar{d}_i^U(h) - \bar{d}_i^U(h)\big)}{s_{i,T}^U(h)} > c_{\beta_N',N}^{\text{SNS}}\big) \\
&\quad \le 1 - (G-1)N\Phi\big(c_{\beta_N',N}^{\text{SNS}}\big) + o(1) \\
&\quad \le \beta_N' + o(1) \\
&\quad = \beta + o(1),
\end{aligned}$$

where the first inequality follows by Lemma A.13 and the Bonferroni inequality, the second inequality holds because $c_{\beta_N',N}^{\text{SNS}}$ becomes sufficiently large as $N \to \infty$ and the tail of the t-distribution is heavier than that of the standard normal distribution (Lemma A.2 under the unidimensional case), and the last inequality follows by the fact that $|\beta_N' - \beta| \le Ce_{N,2}^U\sqrt{\log((G-1)N/\beta)} \to 0$ shown in Step 1.

Summing up, we have

$$\sup_{P\in\mathbb{P}_N} P\left(\prod_{i=1}^{N}\widehat{M}_i(g_i^0) \not\supseteq J_1\right) \le \beta + o(1).$$

An implication of Step 2 is as follows. Let

$$\hat{\mathbb{N}} = \{ i \in \{1, \dots, N\} \mid M_i(g_i^0) \neq \emptyset \}.$$

Then

$$\inf_{P \in \mathbb{P}_N} P(\hat{\mathbb{N}} \supseteq \mathbb{N}) \geq 1 - \beta + o(1).$$

Step 3 First, consider the case in which $J_1 = \emptyset$. In this case, the argument in Step 1 yields that

$$\inf_{P \in \mathbb{P}_N} P\big(\hat{g}_i = g_i^0, \forall i\big) = \inf_{P \in \mathbb{P}_N} P\Big(\max_{1 \leq i \leq N} \max_{h \in \mathbb{G} \setminus \{g_i^0\}} \widehat{D}_i^U(h) \leq 0 \Big) \geq 1 - \beta + o(1).$$

The equality in the above display follows because $\hat{g}_i$ minimizes the squared loss in the two-step procedure (see (11) in the main text). Because $\{\hat{g}_i\}_{i=1}^N$ is always included in the confidence set, the limiting probability of the confidence set not including $\{g_i^0\}_{i=1}^N$ is less than $\beta < \alpha$.

Next, consider the case in which $|J_1| \geq 1$. Observe that

$$\begin{aligned}
&\sup_{P \in \mathbb{P}_N} P\big(\{g_i^0\}_{i=1}^N \notin \widehat{C}_{\mathrm{sel}, \alpha, \beta} \big) \\
&= \sup_{P \in \mathbb{P}_N} P\Bigg(\bigcup_{i=1}^N \Big(\{ \widehat{T}_i(g_i^0) > \hat{c}_{\alpha - 2\beta, \hat{N}, i}(g_i^0) \} \cap \Big\{ \max_{h \in \mathbb{G} \setminus \{g_i^0\}} \widehat{D}_i^U(h) > 0 \Big\} \Big) \Bigg) \\
&\leq \sup_{P \in \mathbb{P}_N} P\Bigg(\bigcup_{i \in \mathbb{N}} \{ \widehat{T}_i(g_i^0) > \hat{c}_{\alpha - 2\beta, \hat{N}, i}(g_i^0) \} \cup \bigcup_{i \in \mathbb{N}^c} \Big\{ \max_{h \in \mathbb{G} \setminus \{g_i^0\}} \widehat{D}_i^U(h) > 0 \Big\} \Bigg) \\
&\leq \sup_{P \in \mathbb{P}_N} P\Bigg(\bigcup_{i \in \mathbb{N}} \{ \widehat{T}_i(g_i^0) > \hat{c}_{\alpha - 2\beta, \hat{N}, i}(g_i^0) \} \Bigg) + \sup_{P \in \mathbb{P}_N} P\Bigg(\bigcup_{i \in \mathbb{N}^c} \Big\{ \max_{h \in \mathbb{G} \setminus \{g_i^0\}} \widehat{D}_i^U(h) > 0 \Big\} \Bigg).
\end{aligned}$$

By Step 1, we have

$$\sup_{P \in \mathbb{P}_N} P\Bigg(\bigcup_{i \in \mathbb{N}^c} \Big\{ \max_{h \in \mathbb{G} \setminus \{g_i^0\}} \widehat{D}_i^U(h) > 0 \Big\} \Bigg) \leq \beta + o(1).$$

By Step 2, we have

$$\begin{aligned}
&\sup_{P \in \mathbb{P}_N} P\Bigg(\bigcup_{i \in \mathbb{N}} \{ \widehat{T}_i(g_i^0) > \hat{c}_{\alpha - 2\beta, \hat{N}, i}(g_i^0) \} \Bigg) \\
&\leq \sup_{P \in \mathbb{P}_N} P\Bigg(\{\hat{\mathbb{N}} \supseteq \mathbb{N}\} \cap \bigcup_{i \in \mathbb{N}} \{ \widehat{T}_i(g_i^0) > \hat{c}_{\alpha - 2\beta, \hat{N}, i}(g_i^0) \} \Bigg) + \sup_{P \in \mathbb{P}_N} P\big(\{\hat{\mathbb{N}} \not\supseteq \mathbb{N}\} \big) \\
&\leq \sup_{P \in \mathbb{P}_N} P\Bigg(\bigcup_{i \in \mathbb{N}} \{ \widehat{T}_i(g_i^0) > \hat{c}_{\alpha - 2\beta, |\mathbb{N}|}(g_i^0) \} \Bigg) + \beta + o(1).
\end{aligned}$$

Thus, we have

$$\sup_{P\in\mathbb{P}_N} P\big(\{g_i^0\}_{i=1}^N \notin \widehat{C}_{\mathrm{sel},\alpha,\beta}\big) \le \sup_{P\in\mathbb{P}_N} P\Big(\bigcup_{i\in\mathbb{N}}\{\widehat{T}_i(g_i^0) > \hat{c}_{\alpha-2\beta,|\mathbb{N}|,i}(g_i^0)\}\Big) + 2\beta + o(1).$$

Theorem 2 implies

$$\limsup_{N,T\to\infty}\sup_{P\in\mathbb{P}_N} P\big(\{g_i^0\}_{i=1}^N \notin \widehat{C}_{\mathrm{sel},\alpha,\beta}\big) \le \alpha. \qquad \square$$

A.3 *Supporting lemmas*

LEMMA A.1. *Let* $(\phi_i)_{i=1}^n$ *denote a collection of independent, nonrandomized tests, and suppose that*

$$\alpha_i = n\mathbb{P}(\phi_i > 0)$$

with $\alpha_{\max} := \max_{i=1,\dots,n}\alpha_i < 1$. *Then*

$$\alpha_{\min} - \frac{\alpha_{\min}^2}{2} \le \mathbb{P}\Big(\max_{i=1,\dots,n}\phi_i > 0\Big) \le \alpha_{\max} - \frac{\alpha_{\max}^2}{2}\bigg(1 - \frac{\alpha_{\max}}{3} + \frac{1}{n}\Big(1 - \frac{\alpha_{\max}}{n}\Big)^{-2}\bigg),$$

where $\alpha_{\min} := \min_{i=1,\dots,n}\alpha_i$.

PROOF. For fixed $0 < x < 1$, let $\bar{x}$ denote a generic intermediate value between zero and x. By a Taylor expansion around $x = 0$,

$$\exp(-x) = 1 - x + \frac{1}{2}x^2 - \frac{1}{6}\exp(-\bar{x})x^3 \ge 1 - x + x^2\Big(\frac{1}{2} - \frac{x}{6}\Big). \tag{A.8}$$

Moreover,

$$\log(1-x) = 0 - x - \frac{x^2}{2(1-\bar{x})^2} \ge -x - \frac{x^2}{2(1-x)^2}. \tag{A.9}$$

Now, for $0 < \alpha < 1$,

$$\begin{aligned}
\Big(1 - \frac{\alpha}{n}\Big)^n &= \exp\Big(n\log\Big(1 - \frac{\alpha}{n}\Big)\Big) \\
&\ge \exp(-\alpha)\exp\bigg(-\frac{\alpha^2}{2n}\Big(1 - \frac{\alpha}{n}\Big)^{-2}\bigg) \\
&\ge \bigg(1 - \alpha + \alpha^2\Big(\frac{1}{2} - \frac{\alpha}{6}\Big)\bigg)\bigg(1 - \frac{\alpha^2}{2n}\Big(1 - \frac{\alpha}{n}\Big)^{-2}\bigg) \\
&\ge 1 - \alpha + \frac{\alpha^2}{2}\Big(1 - \frac{\alpha}{3}\Big) - \frac{\alpha^2}{2n}\Big(1 - \frac{\alpha}{n}\Big)^{-2},
\end{aligned}$$

where the first inequality uses (A.9), the second inequality uses (A.8) and the last inequality uses

$$1-\alpha+\alpha^2\left(\frac{1}{2}-\frac{\alpha}{6}\right)\le 1.$$

We conclude that

$$\mathbb{P}\Big(\max_{i=1,\dots,n}\phi_i>0\Big)=1-\mathbb{P}\Big(\max_{i=1,\dots,n}\phi_i=0\Big)$$
$$\le 1-\left(1-\frac{\alpha_{\max}}{n}\right)^n\le \alpha_{\max}-\frac{\alpha_{\max}^2}{2}\left(1-\frac{\alpha_{\max}}{3}\right)+\frac{\alpha_{\max}^2}{2n}\left(1-\frac{\alpha_{\max}}{n}\right)^{-2}.$$

Next, note that

$$\left(1-\frac{\alpha}{n}\right)^n\le\exp(-\alpha)\le 1-\alpha+\frac{\alpha^2}{2}$$

and, therefore,

$$\mathbb{P}\Big(\max_{i=1,\dots,n}\phi_i>0\Big)=1-\mathbb{P}\Big(\max_{i=1,\dots,n}\phi_i=0\Big)\ge 1-\left(1-\frac{\alpha_{\min}}{n}\right)^n\ge\alpha_{\min}-\frac{\alpha_{\min}^2}{2}. \qquad \square$$

LEMMA A.2. *Let V denote a correlation matrix, which is possibly singular, and let $\Phi_{\max,V}$ denote the distribution function of the maximum element of a multivariate normal random vector with covariance matrix V. There is $t^*\in\mathbb{R}$ independent of T and V such that for all $t>t^*$,*

$$t_{\max,V,T-1}\left(\sqrt{\frac{T-1}{T}}t\right)\le\Phi_{\max,V}(t).$$

PROOF. Let $\mathbf{x}$ be a vector of random variables such that $\mathbf{x}\sim N(0,V)$. By the definitions of $\Phi_{\max,V}$ and $t_{\max,V,T-1}$, we have

$$\Phi_{\max,V}(t)=P(\mathbf{x}\le t)$$

and

$$t_{\max,V,T-1}\left(\sqrt{\frac{T-1}{T}}t\right)=P\left(\frac{1}{\sqrt{z/(T-1)}}\mathbf{x}\le\sqrt{\frac{T-1}{T}}t\right),$$

where an inequality such as $\mathbf{x}\le t$ is understood in an elementwise way, and z is a χ^2 random variable with $(T-1)$ degrees of freedom that is independent of $\mathbf{x}$.

Let r be the rank of V. We have the following eigendecomposition of V:

$$V=U\Sigma U',$$

where Σ is a diagonal matrix with nonnegative elements and U is a unitary matrix. We arrange the elements of U and Σ such that the first r diagonal elements of Σ are nonzero and its other diagonal elements are zero. Let Σ_r be the $r \times r$ upper-left block of Σ. Let

$$\mathbf{x}^* = U'\mathbf{x}.$$

By construction, $\mathbf{x}^* \sim N(0, \Sigma)$. Because Σ is diagonal and only the first r diagonal elements are nonzero, the first r elements of $\mathbf{x}^*$ can be nonzero, and its other elements are zero. Let $\mathbf{x}_r$ be the vector of the first r elements of $\mathbf{x}^*$. Note that by the definition of Σ_r, $\mathbf{x}^* \sim N(0, \Sigma_r)$. This observation implies that

$$\mathbf{x} = U\mathbf{x}^* = U_r\mathbf{x}_r,$$

where U_r is the matrix that consists of the first r columns of U.

We can then write

$$\Phi_{\max,V}(t) = \int_{\mathbf{x}\le t} \phi_{\Sigma_r}(\mathbf{x}_r)\, d\mathbf{x}_r$$

and

$$t_{\max,V,T-1}\left(\sqrt{\frac{T-1}{T}}t\right) = \int_{\mathbf{x}\le\sqrt{\frac{T-1}{T}}t} f^t_{\Sigma_r,T-1}(\mathbf{x}_r)\, d\mathbf{x}_r = \int_{\mathbf{x}\le t} f^{t,*}_{\Sigma_r,T-1}(\mathbf{x}_r)\, d\mathbf{x}_r,$$

where

$$\phi_{\Sigma_r}(\mathbf{x}_r) = (2\pi)^{-r/2}\big(\det(\Sigma_r)\big)^{-1/2} \exp\left(-\frac{1}{2}\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r\right),$$

and

$$f^t_{\Sigma_r,T-1}(\mathbf{x}_r) = \big(\pi(T-1)\big)^{-r/2}\big(\det(\Sigma_r)\big)^{-1/2}\Gamma\left(\frac{T+r-1}{2}\right)\left(\Gamma\left(\frac{T-1}{2}\right)\right)^{-1} \times \left(1+\frac{1}{T-1}\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r\right)^{-(T+r-1)/2}$$

is the density of the multivariate t distribution with scale matrix V and $(T-1)$ degrees of freedom, and

$$f^{t,*}_{\Sigma_r,T-1}(\mathbf{x}_r) = (\pi T)^{-r/2}\big(\det(\Sigma_r)\big)^{-1/2}\Gamma\left(\frac{T+r-1}{2}\right)\left(\Gamma\left(\frac{T-1}{2}\right)\right)^{-1} \times \left(1+\frac{1}{T}\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r\right)^{-(T+r-1)/2}.$$

We now identify a region in which $f^{t,*}_{\Sigma_r,T-1}(\mathbf{x}_r) > \phi_{\Sigma_r}(\mathbf{x}_r)$. We have

$$\log f_{\Sigma_r,T-1}(\mathbf{x}_r) - \log\phi_{\Sigma_r}(\mathbf{x}_r) = A_T - \frac{T+r-1}{2}\log\left(1+\frac{1}{T}\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r\right) + \frac{1}{2}\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r,$$

where

$$A_T = -\frac{r}{2}\log(T) + \log\Gamma\left(\frac{T+r-1}{2}\right) - \log\Gamma\left(\frac{T-1}{2}\right) + \frac{r}{2}\log(2).$$

By the property of the logarithm function and the linear function, there is a unique value, denoted by x_T^*, such that $f^{t,*}_{\Sigma_r,T-1}(\mathbf{x}_r) \le \phi_{\Sigma_r}(\mathbf{x}_r)$ implies $\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r \le x_T^*$. To see this, we consider the two functions $\log(1+y)$ and $ay+b$, where $a = T/(T+r-1)$ and $b = 2A_T/(T+r-1)$. We want to find a value of y, say y', such that if $y \ge y'$ then $\log(1+y) \le ay+b$. Because $\log(1+y)$ is increasing and concave and $a > 0$ there are two possibilities: (1) $ay+b \ge \log(1+y)$ for any y and $ay+b > \log(1+y)$ almost always; (2) the curves $\log(1+y)$ and $ay+b$ intersect with each other at two points, say y_1 and y_2 such that $\log(1+y) < ay+b$ for $y < y_1$, $\log(1+y) \ge ay+b$ for $y_1 \le y \le y_2$, and $\log(1+y) < ay+b$ for $y > y_2$. The first case does not apply to our situation because if this was the case, then $f^{t,*}_{\Sigma_r,T-1}(\mathbf{x}_r) > \phi_{\Sigma_r}(\mathbf{x}_r)$ almost always, contradicting the fact that both curves integrate to one. Thus, the second case applies. The values of y_1 and y_2 can be obtained by solving $\log(1+y) = ay+b$. It holds $y_2 > 0$ because the slope of $\log(1+y)$ at y_2 must be smaller than a and $0 < a < 1$.

Choose t large enough such that $\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r \le x_T^*$ implies $\mathbf{x} \le t$. This choice of t depends on T only through x_T^*. In particular, if $x_T^* = O(1)$ then t can be chosen independently of T. To prove this set, $t = \sqrt{x_T^* \dim(\mathbf{x})}$. Since V is a correlation matrix, its largest eigenvalue is bounded by r and $\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r \ge \|\mathbf{x}_r\|^2/r$. Because $\mathbf{x}^*$ is a vector whose first r elements are those of $\mathbf{x}_r$ and whose other elements are zero, $\|\mathbf{x}_r\|^2 = \|\mathbf{x}^*\|^2$. By the definition of $\mathbf{x}^*$, it holds that $\|\mathbf{x}^*\|^2 = \|U'\mathbf{x}\|^2 = \|\mathbf{x}\|^2$, where the last equality uses the fact that U is a unitary matrix. Observe that if $\mathbf{x} \not\le t$ so that an element of $\mathbf{x}$ exceeds t, then $\|\mathbf{x}\|^2 > t^2 \ge x_T^* r$. This implies that $\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r \ge \|\mathbf{x}\|^2/r > x_T^* r/r = x_T^*$.

We have

$$
\begin{aligned}
\Phi_{\max,\Sigma_r}(t) - t_{\max,\Sigma_r,T-1}\left(\sqrt{\frac{T-1}{T}}t\right) &= \int_{\mathbf{x}\le t} \left(\phi_{\Sigma_r}(\mathbf{x}_r) - f^{t,*}_{\Sigma_r,T-1}(\mathbf{x}_r)\right) d\mathbf{x}_r \\
&= \int_{\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r \le x_T^*} \left(\phi_{\Sigma_r}(\mathbf{x}_r) - f^{t,*}_{\Sigma_r,T-1}(\mathbf{x}_r)\right) d\mathbf{x}_r \\
&\quad + \int_{\mathbf{x}\le t, \mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r > x_T^*} \left(\phi_{\Sigma_r}(\mathbf{x}_r) - f^{t,*}_{V,T-1}(\mathbf{x}_r)\right) d\mathbf{x}_r,
\end{aligned}
$$

where the first integral on the right-hand side of the equation is taken over $\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r \le a$ because $\{\mathbf{x}_r : \mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r \le x_T^*, \mathbf{x} \le t\} = \{\mathbf{x}_r : \mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r \le x_T^*\}$ by our choice of t. Because both $\phi_{\Sigma_r}(\mathbf{x}_r)$ and $f^{t,*}_{\Sigma_r,T-1}(\mathbf{x}_r)$ are densities and integrate to one, we have

$$
\int_{\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r \le x_T^*} \left(\phi_{\Sigma_r}(\mathbf{x}_r) - f^{t,*}_{\Sigma_r,T-1}(\mathbf{x}_r)\right) d\mathbf{x}_r = -\int_{\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r > x_T^*} \left(\phi_{\Sigma_r}(\mathbf{x}_r) - f^{t,*}_{\Sigma_r,T-1}(\mathbf{x}_r)\right) d\mathbf{x}_r,
$$

Thus, for t large enough such that $\mathbf{x}'V^{-1}\mathbf{x} \le x_T^*$ implies $\mathbf{x} \le t$, we have

$$
\begin{aligned}
&\Phi_{\max,\Sigma_r}(t) - t_{\max,\Sigma_r,T-1}\left(\sqrt{\frac{T-1}{T}}t\right) \\
&= -\int_{\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r > x_T^*} \left(\phi_{\Sigma_r}(\mathbf{x}_r) - f^{t,*}_{\Sigma_r,T-1}(\mathbf{x}_r)\right) d\mathbf{x}_r
\end{aligned}
$$

$$
\begin{aligned}
&+\int_{\mathbf{x}\le t,\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r>x_T^*}\left(\phi_{\Sigma_r}(\mathbf{x}_r)-f_{V,T-1}^{t,*}(\mathbf{x}_r)\right)d\mathbf{x}_r\\
&=\int_{\mathbf{x}\not\le t,\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r>x_T^*}\left(\phi_{\Sigma_r}(\mathbf{x}_r)-f_{V,T-1}^{t,*}(\mathbf{x}_r)\right)d\mathbf{x}_r\ge 0,
\end{aligned}
$$

where the last inequality follows because $\mathbf{x}_r'\Sigma_r^{-1}\mathbf{x}_r > x_T^*$ implies $\phi_{\Sigma_r}(\mathbf{x}_r) > f_{V,T-1}^{t,*}(\mathbf{x}_r)$.

Next, we evaluate the order of x_T^*. Note that x_T^* solves

$$
\frac{1}{2}x_T^*+A_T=\frac{T+r-1}{2}\log\left(1+\frac{1}{T}x_T^*\right).
$$

We first show that $A_T=O(1)$ where the order is taken with respect to T. We separately examine the cases of odd and even G. Suppose that r is even (we may assume $r\ge 2$). Then the recurrent relation of the Gamma function implies that

$$
\begin{aligned}
A_T&=-\frac{r}{2}\log(T)+\sum_{j=0}^{r/2-1}\log\left(\frac{T-1}{2}+j\right)+\frac{r}{2}\log(2)\\
&=-\frac{r}{2}\log(T)+\sum_{j=0}^{r/2-1}\log(T-1+2j)-\frac{r}{2}\log(2)+\frac{r}{2}\log(2)\\
&=\sum_{j=0}^{r/2-1}\log\left(\frac{T-1+2j}{T}\right)=O(1)
\end{aligned}
$$

as $T\to\infty$. Next, we consider cases in which r is odd. For $r=1$, $A_T=O(1)$ follows from

$$
\sqrt{\frac{T}{2}}\frac{\Gamma\left(\frac{T-1}{2}\right)}{\Gamma\left(\frac{T}{2}\right)}\to 1. \tag{A.10}
$$

For $r\ge 3$, by the recurrent relation of the Gamma function, we have

$$
\begin{aligned}
A_T&=-\frac{r}{2}\log(T)+\sum_{j=0}^{r/2-1}\log\left(\frac{T}{2}+j\right)+\log\Gamma\left(\frac{T}{2}\right)-\log\Gamma\left(\frac{T-1}{2}\right)+\frac{r}{2}\log(2)\\
&=\sum_{j=0}^{(r-1)/2-1}\log\left(\frac{T+2j}{T}\right)+\frac{1}{2}\log\left(\frac{2}{T}\right)+\log\Gamma\left(\frac{T}{2}\right)-\log\Gamma\left(\frac{T-1}{2}\right).
\end{aligned}
$$

By (A.10),

$$
\log\Gamma\left(\frac{T}{2}\right)-\log\left(\Gamma\left(\frac{T-1}{2}\right)\left(\frac{T}{2}\right)^{1/2}\right)=O(1).
$$

We have established that $A_T=O(1)$ for all $r\ge 1$. To prove the lemma, it now suffices to prove $x_T^*=O(1)$. Suppose the opposite is true. Then there is a subsequence

$T_1, \dots, T_k, \dots$ such that $x^*_{T_k}$ monotonically diverges to infinity. By the definition of x^*_T, we have

$$x^*_T + A_T = (T + r - 1)\log\left(1 + \frac{1}{T}x^*_T\right).$$

For sufficiently large y, $y/2 > \log(1 + y)$. Therefore, for sufficiently large k, we have

$$x^*_{T_k} + A_T < \frac{T + r - 1}{2T} x^*_{T_k}$$

Rearranging terms yields

$$\frac{T - r + 1}{2T} x^*_{T_k} + A_T < 0,$$

contradicting that $A_T = O(1)$ and $x^*_{T_k}$ diverging to infinity can both be true. This proves $x^*_T = O(1)$. □

LEMMA A.3 (Comparison bound for critical values with regularization). *Let Ω and $\widehat{\Omega}$ denote $p \times p$ correlation matrices and let ϵ, δ, and c_ω denote positive constants such that $4\delta \le \epsilon \le 4c_\omega/3$. Suppose that $\Omega_{ij} > -1 + c_\omega$ for all $i, j = 1, \dots, p$, and*

$$\|\widehat{\Omega} - \Omega\|_{\max} \le \delta.$$

Let $X \sim N(0, \Omega)$ and $\hat{X}^\epsilon \sim N(0, \rho(\widehat{\Omega}, \epsilon))$. Then there is a universal constant C such that for all $a > \sqrt{2}$,

$$\frac{P\left(\max_{j=1,\dots,p} X_j > a\right)}{P\left(\max_{j=1,\dots,p} \hat{X}^\epsilon_j > a\right)} - 1 \le Ca\sqrt{\epsilon}.$$

In particular, suppose that $\hat{c}_{\alpha,N}$ is the $1 - \alpha/N$ quantile of $\max_{j=1,\dots,p} \hat{X}^\epsilon_j$ and let $c_{\alpha_N,N}$ denote the $1 - \alpha_N/N$ quantile of $\max_{j=1,\dots,p} X_j$, where

$$\alpha_N = \alpha(1 + \hat{c}_{\alpha,N} C\sqrt{\epsilon}).$$

If $\hat{c}_{\alpha,N} > \sqrt{2}$, then $\hat{c}_{\alpha,N} \ge c_{\alpha_N,N}$.

PROOF. Write $\hat{\epsilon} = \epsilon^*(\widehat{\Omega}, \epsilon)$. By the Cholesky decomposition, there is a lower-triangular matrix L (possibly with some diagonal elements equal to zero) such that $\widehat{\Omega} = LL'$. $\widehat{\Omega}$ can be interpreted as the covariance matrix of the random vector LW, where W is a random vector in $\mathbb{R}^p$ with expectation zero and covariance matrix I_p. $V = \widehat{\Omega} + \hat{\epsilon} I_p$ can be interpreted as the covariance matrix of the random vector that is generated by adding independent, component-specific noise E_i to the ith components of LW, where E_i has mean zero and variance $\hat{\epsilon}$. Then $\rho(\widehat{\Omega}, \hat{\epsilon})$ transforms V into a correlation matrix. Since $\rho(\widehat{\Omega}, \epsilon)$ and $\widehat{\Omega}$ are both correlation matrices, $\rho(\widehat{\Omega}, \epsilon)_{ii} = \widehat{\Omega}_{ii} = 1$, for all $i = 1, \dots, p$. Let ℓ_i

denote the ith row of L. For $i \neq j$, $\widehat{\Omega}_{ij}$ is equal to the covariance between $\ell_i'W$ and $\ell_j'W$, that is,

$$\widehat{\Omega}_{ij} = \mathrm{cov}\big(\ell_i'W, \ell_j'W\big) = \ell_i' \,\mathrm{cov}(W)\ell_j = \ell_i'\ell_j.$$

V_{ij} is equal to the covariance between $\ell_i'W + E_i$ and $\ell_j'W + E_j$, that is,

$$V_{ij} = \mathrm{cov}\big(\ell_i'W + E_i, \ell_j'W + E_j\big) = \ell_i' \,\mathrm{cov}(W)\ell_j = \ell_i'\ell_j = \widehat{\Omega}_{ij}.$$

For $i = 1, \dots, p$,

$$V_{ii} = \mathrm{cov}\big(\ell_i'W + E_i, \ell_i'W + E_i\big) = \ell_i' \,\mathrm{cov}(W)\ell_i + \hat{\epsilon} = \widehat{\Omega}_{ii} + \hat{\epsilon} = 1 + \hat{\epsilon}.$$

Therefore, for $i \neq j$,

$$\rho(\widehat{\Omega}, \epsilon)_{ij} = \frac{V_{ij}}{\sqrt{V_{ii}V_{jj}}} = \frac{\widehat{\Omega}_{ij}}{1 + \hat{\epsilon}}.$$

We now derive a bound on

$$\Delta_{ij} = \big(\arcsin\big(\rho(\widehat{\Omega}, \epsilon)_{ij}\big) - \arcsin(\Omega_{ij})\big)^+.$$

Since $\arcsin(\cdot)$ is strictly increasing on $(0, 1)$, a necessary condition for $\Delta_{ij} \neq 0$ is $\rho(\widehat{\Omega}, \epsilon)_{ij} > \Omega_{ij}$. This condition bounds $\rho(\widehat{\Omega}, \epsilon)_{ij}$ and Ω_{ij} away from -1 and 1. In particular,

$$\widehat{\Omega}/(1 + \hat{\epsilon}) = \rho(\widehat{\Omega}, \epsilon)_{ij} > \Omega_{ij}$$

implies

$$\widehat{\Omega}_{ij} > (1 + \hat{\epsilon})\Omega_{ij}.$$

Since we also have $\widehat{\Omega}_{ij} \leq \Omega_{ij} + \delta$, an $\widehat{\Omega}_{ij}$ fulfilling both conditions exists only if

$$(1 + \hat{\epsilon})\Omega_{ij} < \Omega_{ij} + \delta$$

or equivalently if $\hat{\epsilon}\Omega_{ij} < \delta$. Suppose that $\Omega_{ij} > 1 - \epsilon/2$. Then we have

$$\begin{aligned}\hat{\epsilon} &\geq \epsilon - (1 - \widehat{\Omega}_{ij}) \\ &\geq \epsilon + \Omega_{ij} - \delta - 1 > \epsilon + (1 - \epsilon/2) - \delta - 1 = \epsilon/2 - \delta.\end{aligned}$$

Therefore, $\hat{\epsilon}\Omega_{ij} < \delta$ is only possible if

$$(\epsilon/2 - \delta)(1 - \epsilon/2) < \delta.$$

This inequality contradicts $\epsilon \geq 4\delta$, and hence we can take $\Omega_{ij} \leq 1 - \epsilon/2$. Moreover, since $\epsilon \leq 2c_\omega$, $\Omega_{ij} \geq -1 - c_\omega \geq -1 - \epsilon/2$.

For an upper bound on $\rho(\widehat{\Omega}, \epsilon)_{ij}$, we have

$$\rho(\widehat{\Omega}, \epsilon)_{ij} = \frac{\widehat{\Omega}_{ij}}{1+\hat{\epsilon}} \leq \frac{\widehat{\Omega}_{ij}}{1+\epsilon-(1-\widehat{\Omega}_{ij})} \leq \frac{\widehat{\Omega}_{ij}}{\widehat{\Omega}_{ij}+\epsilon} \leq \frac{1}{1+\epsilon} \leq 1-\epsilon/2.$$

For a lower bound on $\rho(\widehat{\Omega}, \epsilon)_{ij}$, suppose that $\widehat{\Omega}_{ij} < 0$, in which case,

$$\rho(\widehat{\Omega}, \epsilon)_{ij} = \frac{\widehat{\Omega}_{ij}}{1+\hat{\epsilon}} \geq \widehat{\Omega}_{ij} = \Omega_{ij} + \widehat{\Omega}_{ij} - \Omega_{ij} \geq -1 + c_\omega - \delta \geq -1+\epsilon/2,$$

provided that $c_\omega - \epsilon/2 - \delta \geq 0$. This condition is satisfied if $\epsilon \leq 4c_\omega/3$. Therefore, we have the bounds

$$-1+\epsilon/2 \leq \Omega_{ij} \leq 1-\epsilon/2$$

and

$$-1+\epsilon/2 \leq \rho(\widehat{\Omega}_{ij}, \epsilon) \leq 1-\epsilon/2.$$

We also have

$$\rho(\widehat{\Omega}, \epsilon)_{ij} - \Omega_{ij} = \widehat{\Omega}_{ij}/(1+\epsilon) - \Omega_{ij} \leq \delta + \epsilon \leq 5\epsilon.$$

By the intermediate value theorem, there is an intermediate value ρ^* between $-1+\epsilon/2$ and $1-\epsilon/2$ such that, on $\rho(\widehat{\Omega}, \epsilon)_{ij} > \Omega_{ij}$,

$$\Delta_{ij} = \frac{\rho(\widehat{\Omega}, \epsilon)_{ij} - \Omega_{ij}}{\sqrt{1-\left(\rho^*\right)^2}} \leq \frac{5\epsilon}{\sqrt{1-(1-\epsilon/2)^2}} \leq \frac{5\epsilon}{\sqrt{\epsilon(1-\epsilon/4)}} \leq \frac{10\sqrt{\epsilon}}{\sqrt{3}}.$$

Let Φ denote the cumulative distribution function of a standard normal random variable, and let ϕ denote its probability density function. Gordon's lower bound (see, e.g., Duembgen (2010)) states that

$$1-\Phi(a) > \frac{\phi(a)}{a\left(1-1/a^2\right)}$$

for $a > 0$, and thus $1-\Phi(a) > \frac{1}{2}\phi(a)/a$ for $a > \sqrt{2}$. Therefore,

$$P\Big(\max_{j=1,\ldots,p} \hat{X}_j > a\Big) \geq P\big(\hat{X}_1^\epsilon > a\big) = 1-\Phi(a) > \frac{\phi(a)}{2a} = \frac{\exp\Big(-\frac{a^2}{2}\Big)}{a\sqrt{8\pi}}$$

By Theorem 2.1 in Li and Shao (2002),

$$P\Big(\max_{j=1,\ldots,p} X_j > a\Big) - P\Big(\max_{j=1,\ldots,p} \hat{X}_j^\epsilon > a\Big)$$
$$= P\Big(\max_{j=1,\ldots,p} \hat{X}_j^\epsilon \leq a\Big) - P\Big(\max_{j=1,\ldots,p} X_j \leq a\Big)$$

$$\leq \frac{1}{2\pi}\exp\left(-\frac{a^2}{2}\right)\sum_{1\leq i<j\leq p}\Delta_{ij} \leq \frac{5}{\pi\sqrt{3}}\exp\left(-\frac{a^2}{2}\right)\sqrt{\epsilon}$$

We may assume $P(\max_{j=1,\dots,p} X_j > a) > P(\max_{j=1,\dots,p} \hat{X}_j^{\epsilon} > a)$ since the statement of the theorem holds trivially otherwise. Then, combining the bounds derived above, yields

$$\frac{P\left(\max_{j=1,\dots,p} X_j > a\right) - P\left(\max_{j=1,\dots,p} \hat{X}_j^{\epsilon} > a\right)}{P\left(\max_{j=1,\dots,p} \hat{X}_j^{\epsilon} > a\right)} < \frac{10a\sqrt{\epsilon}}{\sqrt{3}}.$$

This proves the first claim of the lemma. To prove the second claim of the lemma, note that the first claim of the lemma implies

$$P\left(\max_{j=1,\dots,p} X_j > \hat{c}_{\alpha,N}\right) = \frac{P\left(\max_{j=1,\dots,p} X_j > \hat{c}_{\alpha,N}\right)}{P\left(\max_{j=1,\dots,p} \hat{X}_j > \hat{c}_{\alpha,N}\right)}\alpha/N$$

$$\leq \alpha/N(1+\hat{c}_{\alpha,N}C\sqrt{\epsilon}) \leq \alpha_N/N. \qquad \square$$

LEMMA A.4 (Consistency of $\widehat{\Omega}$). *Let $\mathbb{P}_N$ be the set of probability measures, which satisfy Assumptions* 1 *with identical choices of a, b, d_1, and d_2. Assume $T^{\delta_1} \leq N$ for some universal constant $\delta_1 > 0$. Let Assumption* 2 *hold and $\kappa_N \asymp T^{\rho}$ where $0 < \rho < (\vartheta - 1)/(3\vartheta - 2)$. Let*

$$b_N^{\mathrm{LV}} = \frac{r_{\theta,N}}{\iota_N \min_{1\leq i\leq N}\sigma_i} + T^{-c_1}(\log N)^{c_2} + T^{-\rho},$$

where $c_1 > 0$ and $c_2 > 0$ are two constants defined in Lemma A.5 *with c_1 depending only on (ρ, ϑ) and c_2 depending only on (d_2, ϑ). Assume that $b_N^{\mathrm{LV}} \to 0$. For any sequence ζ_N such that $\zeta_N \to \infty$ as $N, T \to \infty$,*

$$\sup_{P\in\mathbb{P}_N} P\left(\max_{1\leq i\leq N}\max_{(h,h^*)\in\mathbb{G}^2}\left|\left(\widehat{\Omega}_i(g_i^0)\right)_{h,h^*} - \left(\Omega_i(g_i^0)\right)_{h,h^*}\right| > \zeta_N b_N^{\mathrm{LV}}\right) = o(1).$$

PROOF. Throughout the proof, let C, C', and C'' denote generic constants that do not depend on $P \in \mathbb{P}$. Let $\xi_i(h) = \sqrt{\Xi_i(h,h)}$ and $\widehat{\xi}_i(h) = \sqrt{\widehat{\Xi}_i(h,h)}$.

We observe the following decomposition:

$$\begin{aligned}
&(\widehat{\Omega}_i)_{h,h^*} - (\Omega_i)_{h,h^*} \\
&= \frac{\widehat{\Xi}_i(h,h^*)}{\hat{\xi}_i(h)\hat{\xi}_i(h^*)} - \frac{\Xi_i(h,h^*)}{\xi_i(h)\xi_i(h^*)} \\
&= \left[\left(\frac{\xi_i(h)}{\hat{\xi}_i(h)}\right)\left(\frac{\xi_i(h^*)}{\hat{\xi}_i(h^*)}\right) - 1\right]\left(\frac{\widehat{\Xi}_i(h,h^*)}{\xi_i(h)\xi_i(h^*)} - \frac{\Xi_i(h,h^*)}{\xi_i(h)\xi_i(h^*)}\right) \\
&\quad + \left[\left(\frac{\xi_i(h)}{\hat{\xi}_i(h)}\right)\left(\frac{\xi_i(h^*)}{\hat{\xi}_i(h^*)}\right) - 1\right]\frac{\Xi_i(h,h^*)}{\xi_i(h)\xi_i(h^*)} + \left(\frac{\widehat{\Xi}_i(h,h^*)}{\xi_i(h)\xi_i(h^*)} - \frac{\Xi_i(h,h^*)}{\xi_i(h)\xi_i(h^*)}\right).
\end{aligned}$$

Noting that

$$\left|\Xi_i(h,h)\right| = \left|\sigma_i^2\delta_i(h)'\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}_P\left[v_{it}v_{is}x_{it}x_{is}'\right]\right)\delta_i(h)\right|,$$

Assumption 1.4 implies that $\xi_i(h)/(\sigma_i)\|\delta(h)\|$ (and $\xi_i(h^*)/(\sigma_i)\|\delta(h^*)\|$) is bounded away from zero. The inequality $|\sqrt{a}-1| \leq |a-1|$ for $a>0$ implies that

$$\left|\left(\frac{\widehat{\xi}_i(h)}{\xi_i(h)}\right)-1\right| \leq \left|\left(\frac{\widehat{\Xi}_i(h,h)}{\Xi_i(h,h)}\right)-1\right|.$$

Thus, we have

$$\begin{aligned}
&\left|\left(\frac{\xi_i(h)}{\hat{\xi}_i(h)}\right)\left(\frac{\xi_i(h^*)}{\hat{\xi}_i(h^*)}\right)-1\right| \\
&= \left|\left(\frac{\xi_i(h)}{\hat{\xi}_i(h)}-1+1\right)\left(\frac{\xi_i(h^*)}{\hat{\xi}_i(h^*)}\right)-1\right| \\
&= \left|\left(\frac{\xi_i(h)}{\hat{\xi}_i(h)}-1\right)\left(\frac{\xi_i(h^*)}{\hat{\xi}_i(h^*)}\right)+\left(\frac{\xi_i(h^*)}{\hat{\xi}_i(h^*)}-1\right)\right| \\
&= \left|\left(\frac{\xi_i(h)}{\hat{\xi}_i(h)}-1\right)\left(\frac{\xi_i(h^*)}{\hat{\xi}_i(h^*)}\right)+\left(\frac{\xi_i(h^*)}{\hat{\xi}_i(h^*)}-1\right)\right| \\
&\leq \left|\frac{\Xi_i(h,h)}{\widehat{\Xi}_i(h,h)}-1\right|\left(\frac{\Xi_i(h^*,h^*)}{\widehat{\Xi}_i(h^*,h^*)}\right)^{1/2}+\left|\frac{\Xi_i(h^*,h^*)}{\widehat{\Xi}_i(h^*,h^*)}-1\right|.
\end{aligned}$$

Now, it holds that

$$\left|\frac{\Xi_i(h,h)}{\widehat{\Xi}_i(h,h)}-1\right| \leq \left|\frac{\widehat{\Xi}_i(h,h)-\Xi_i(h,h)}{\Xi_i(h,h)}+1\right|^{-1}\left|\frac{\widehat{\Xi}_i(h,h)-\Xi_i(h,h)}{\Xi_i(h,h)}\right|.$$

Provided that $b_N^{\mathrm{LV}} \to 0$, we have

$$\left|\left(\frac{\xi_i(h)}{\hat{\xi}_i(h)}\right)\left(\frac{\xi_i(h^*)}{\hat{\xi}_i(h^*)}\right)-1\right| < C\zeta_N b_N^{\mathrm{LV}}$$

with probability approaching one by Lemma A.5. Therefore, by Lemma A.5, the desired result holds. □

LEMMA A.5 (Consistency of long-run variance estimator). *Let $\mathbb{P}_N$ be the set of probability measures, which satisfy Assumption* 1 *with identical choices of a, b, d_1, and d_2. Assume $T^{\delta_1} \leq N$ for some universal constant $\delta_1 > 0$. Let Assumption* 2 *hold and $\kappa_N \asymp T^{\rho}$ where* $0<\rho<(\vartheta-1)/(3\vartheta-2)$. *Let*

$$b_N^{\mathrm{LV}} = \frac{r_{\theta,N}}{\iota_N \min_{1\leq i\leq N}\sigma_i} + T^{-c_1}(\log N)^{c_2} + T^{-\rho},$$

and assume that $b_N^{\mathrm{LV}} \to 0$. Then there exist two constants $c_1 > 0$ depending only on (ρ, ϑ) and $c_2 > 0$ depending only on (d_2, ϑ), such that, for any sequence ζ_N such that $\zeta_N \to \infty$ as $N, T \to \infty$,

$$\sup_{P \in \mathbb{P}_N} P\left(\max_{1 \le i \le N} \max_{(h,h') \in \mathbb{G}^2} \left| \frac{\widehat{\Xi}_i(h, h') - \Xi_i(h, h')}{\sigma_i^2 \| \delta_i(h) \| \| \delta_i(h') \|} \right| > \zeta_N b_N^{\mathrm{LV}} \right) = o(1).$$

Proof. To conserve notation, we introduce the shorthand

$$K_N^{(j)} = K\left(\frac{j}{\kappa_N} \right)$$

and

$$\begin{aligned} u_{it}^{(+j)} &= u_{i,t+\max(0,j)}, & v_{it}^{(+j)} &= v_{i,t+\max(0,j)}, \\ u_{it}^{(-j)} &= u_{i,t-\max(0,j)}, & v_{it}^{(-j)} &= v_{i,t-\max(0,j)}, \\ w_{it}^{(+j)} &= w_{i,t+\max(0,j)}, & x_{it}^{(+j)} &= x_{i,t+\max(0,j)}, \\ w_{it}^{(-j)} &= w_{i,t-\max(0,j)}, & x_{it}^{(-j)} &= x_{i,t-\max(0,j)} \end{aligned}$$

and

$$\begin{aligned} \overline{x_i x_i} &= \frac{1}{T} \sum_{u=1}^{T} x_{iu} x_{iu}', & \overline{x_i w_i} &= \frac{1}{T} \sum_{u=1}^{T} x_{iu} w_{iu}', \\ \overline{u_i x_i} &= \frac{1}{T} \sum_{u=1}^{T} u_{iu} x_{iu}, & \overline{v_i x_i} &= \frac{1}{T} \sum_{u=1}^{T} v_{iu} x_{iu}. \end{aligned}$$

By the triangular inequality, we have

$$\left| \frac{\widehat{\Xi}_i(h, h') - \Xi_i(h, h')}{\sigma_i^2 \| \delta_i(h) \| \| \delta_i(h') \|} \right| \le \left| \frac{\widehat{\Xi}_i(h, h') - \widetilde{\Xi}_i(h, h')}{\sigma_i^2 \| \delta_i(h) \| \| \delta_i(h') \|} \right| + \left| \frac{\widetilde{\Xi}_i(h, h') - \Xi_i(h, h')}{\sigma_i^2 \| \delta_i(h) \| \| \delta_i(h') \|} \right|.$$

We examine each of the two terms on the right-hand side. We first examine the second term and then the first term. We note that

$$\widetilde{\Xi}_i(h, h') = \sigma_i^2 \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \left(v_{it}^{(+j)} x_{it}^{(+j)} - \overline{v_i x_i} \right) \left(v_{it}^{(-j)} x_{it}^{(-j)} - \overline{v_i x_i} \right)' \delta_i(h')$$

and

$$\Xi_i(h, h') = \sigma_i^2 \delta_i(h)' \left(\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}_P\left[v_{it} v_{is} x_{it} x_{is}' \right] \right) \delta_i(h').$$

Lemma A.6 gives the bound of

$$\sup_{i,h,h'}\Bigg|\sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T}\sum_{t=|j|+1}^{T}\big(v_{it}^{(+j)}x_{it}^{(+j)}-\overline{v_i x_i}\big)\big(v_{it}^{(-j)}x_{it}^{(-j)}-\overline{v_i x_i}\big)' \\ -\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}_P\big[v_{it}v_{is}x_{it}x_{is}'\big]\right)\Bigg|_\infty .$$

We thus have

$$\sup_{P\in\mathbb{P}_N} P\left(\max_{1\le i\le N}\max_{(h,h')\in\mathbb{G}^2}\left|\frac{\widetilde{\Xi}_i(h,h')-\Xi_i(h,h')}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|}\right| > \zeta_{1,N}\big(T^{-c_1}(\log N)^{c_2}+T^{-\rho}\big)\right)=o(1),$$

where $\zeta_{1,N}\to\infty$. Next, we derive the bound of

$$\sup_{i,h,h'}\frac{1}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|}\big|\widehat{\Xi}_i(h,h')-\widetilde{\Xi}_i(h,h')\big|.$$

First, note that

$$d_{it}(h)=-\sigma_i v_{it}x_{it}'(\theta_{g_i^0}-\theta_h)=-\sigma_i v_{it}x_{it}'\delta_i(h),$$

and

$$\bar{d}_i(h)=-\sigma_i\overline{v_i x_i}'(\theta_{g_i^0}-\theta_h)=-\sigma_i\overline{v_i x_i}'\delta_i(h).$$

Let

$$\hat{u}_{it}=y_{it}-x_{it}'\hat{\theta}_{g_i^0}-w_{it}'\hat{\theta}^w,$$

so that

$$\hat{u}_{it}-u_{it}=-x_{it}'(\hat{\theta}_{g_i^0}-\theta_{g_i^0})-w_{it}'\big(\hat{\theta}^w-\theta^w\big).$$

With this notation,

$$\hat{d}_{it}\big(g_i^0,h\big)=\hat{d}_{it}(h)=-\,\hat{u}_{it}x_{it}'\hat{\delta}_i(h).$$

Consider the decomposition

$$\begin{aligned}
&\big(\hat{d}_{it}(g^0,h)-\bar{\hat{d}}_i(g^0,h)\big)\big(\hat{d}_{is}(g^0,h')-\bar{\hat{d}}_i(g^0,h')\big)\\
&\quad-\big(d_{it}(g^0,h)-\bar{d}_i(g^0,h)\big)\big(d_{is}(g^0,h')-\bar{d}_i(g^0,h')\big)\\
&=\big(\hat{\delta}_i(h)-\delta_i(h)\big)'\left(\hat{u}_{it}x_{it}-\frac{1}{T}\sum_{u=1}^{T}\hat{u}_{iu}x_{iu}\right)\left(\hat{u}_{is}x_{is}-\frac{1}{T}\sum_{u=1}^{T}\hat{u}_{iu}x_{iu}\right)'\big(\hat{\delta}_i(h')-\delta_i(h')\big)\\
&\quad-\big(\hat{\delta}_i(h)-\delta_i(h)\big)'\left(u_{it}x_{it}-\frac{1}{T}\sum_{u=1}^{T}u_{iu}x_{iu}\right)\left(u_{is}x_{is}-\frac{1}{T}\sum_{u=1}^{T}u_{iu}x_{iu}\right)'\big(\hat{\delta}_i(h')-\delta_i(h')\big)
\end{aligned}$$

$$
\begin{aligned}
&+ \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \left(\hat{u}_{it} x_{it} - \frac{1}{T}\sum_{u=1}^{T} \hat{u}_{iu} x_{iu} \right) \left(\hat{u}_{is} x_{is} - \frac{1}{T}\sum_{u=1}^{T} \hat{u}_{iu} x_{iu} \right)' \delta_i\big(h'\big) \\
&- \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \left(u_{it} x_{it} - \frac{1}{T}\sum_{u=1}^{T} u_{iu} x_{iu} \right) \left(u_{is} x_{is} - \frac{1}{T}\sum_{u=1}^{T} u_{iu} x_{iu} \right)' \delta_i\big(h'\big) \\
&+ \delta_i(h)' \left(\hat{u}_{it} x_{it} - \frac{1}{T}\sum_{u=1}^{T} \hat{u}_{iu} x_{iu} \right) \left(\hat{u}_{is} x_{is} - \frac{1}{T}\sum_{u=1}^{T} \hat{u}_{iu} x_{iu} \right)' \big(\hat{\delta}_i\big(h'\big) - \delta_i\big(h'\big)\big) \\
&- \delta_i(h)' \left(u_{it} x_{it} - \frac{1}{T}\sum_{u=1}^{T} u_{iu} x_{iu} \right) \left(u_{is} x_{is} - \frac{1}{T}\sum_{u=1}^{T} u_{iu} x_{iu} \right)' \big(\hat{\delta}_i\big(h'\big) - \delta_i\big(h'\big)\big) \\
&+ \delta_i(h)' \left(\hat{u}_{it} x_{it} - \frac{1}{T}\sum_{u=1}^{T} \hat{u}_{iu} x_{iu} \right) \left(\hat{u}_{is} x_{is} - \frac{1}{T}\sum_{u=1}^{T} \hat{u}_{iu} x_{iu} \right)' \delta_i\big(h'\big) \\
&- \delta_i(h)' \left(u_{it} x_{it} - \frac{1}{T}\sum_{u=1}^{T} u_{iu} x_{iu} \right) \left(u_{is} x_{is} - \frac{1}{T}\sum_{u=1}^{T} u_{iu} x_{iu} \right)' \delta_i\big(h'\big) \\
&+ \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \left(u_{it} x_{it} - \frac{1}{T}\sum_{u=1}^{T} u_{iu} x_{iu} \right) \left(u_{is} x_{is} - \frac{1}{T}\sum_{u=1}^{T} u_{iu} x_{iu} \right)' \big(\hat{\delta}_i\big(h'\big) - \delta_i\big(h'\big)\big) \\
&+ \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \left(u_{it} x_{it} - \frac{1}{T}\sum_{u=1}^{T} u_{iu} x_{iu} \right) \left(u_{is} x_{is} - \frac{1}{T}\sum_{u=1}^{T} u_{iu} x_{iu} \right)' \delta_i\big(h'\big) \\
&+ \delta_i(h)' \left(u_{it} x_{it} - \frac{1}{T}\sum_{u=1}^{T} u_{iu} x_{iu} \right) \left(u_{is} x_{is} - \frac{1}{T}\sum_{u=1}^{T} u_{iu} x_{iu} \right)' \big(\hat{\delta}_i\big(h'\big) - \delta_i\big(h'\big)\big).
\end{aligned}
$$

Next, we consider the following decomposition:

$$
\begin{aligned}
\hat{u}_{it} x_{it} - \frac{1}{T}\sum_{u=1}^{T} \hat{u}_{iu} x_{iu} = u_{it} x_{it} - \frac{1}{T}\sum_{u=1}^{T} u_{iu} x_{iu} - \big(x_{it} x_{it}' - \overline{x_i x_i}\big)\big(\hat{\theta}_{g_i^0} - \theta_{g_i^0}\big) \\
- \big(x_{it} w_{it}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big).
\end{aligned}
$$

We thus have

$$
\begin{aligned}
&\left(\hat{u}_{it} x_{it} - \frac{1}{T}\sum_{u=1}^{T} \hat{u}_{iu} x_{iu} \right) \left(\hat{u}_{is} x_{is} - \frac{1}{T}\sum_{u=1}^{T} \hat{u}_{iu} x_{iu} \right)' \\
&\quad = \big(u_{it} x_{it} - \overline{u_i x_i}\big)\big(u_{is} x_{is} - \overline{u_i x_i}\big)' \\
&\qquad - \big(u_{it} x_{it} - \overline{u_i x_i}\big)\Big(\big(x_{is} x_{is}' - \overline{x_i x_i}\big)\big(\hat{\theta}_{g_i^0} - \theta_{g_i^0}\big) + \big(x_{is} w_{is}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\Big)' \\
&\qquad - \Big(\big(x_{it} x_{it}' - \overline{x_i x_i}\big)\big(\hat{\theta}_{g_i^0} - \theta_{g_i^0}\big) + \big(x_{it} w_{it}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\Big)\big(u_{is} x_{is} - \overline{u_i x_i}\big)' \\
&\qquad + \Big(\big(x_{it} x_{it}' - \overline{x_i x_i}\big)\big(\hat{\theta}_{g_i^0} - \theta_{g_i^0}\big) + \big(x_{it} w_{it}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\Big)
\end{aligned}
$$

$$\begin{aligned}
&\times \left(\left(x_{is}x_{is}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0}) + \left(x_{is}w_{is}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)' \\
&= (u_{it}x_{it} - \overline{u_i x_i})(u_{is}x_{is} - \overline{u_i x_i})' \\
&\quad - (u_{it}x_{it} - \overline{u_i x_i})\left(\left(x_{is}x_{is}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \\
&\quad - (u_{it}x_{it} - \overline{u_i x_i})\left(\left(x_{is}w_{is}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)' \\
&\quad - \left(\left(x_{it}x_{it}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)(u_{is}x_{is} - \overline{u_i x_i})' \\
&\quad - \left(\left(x_{it}w_{it}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)(u_{is}x_{is} - \overline{u_i x_i})' \\
&\quad + \left(\left(x_{it}x_{it}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)\left(\left(x_{is}x_{is}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \\
&\quad + \left(\left(x_{it}x_{it}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)\left(\left(x_{is}w_{is}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)' \\
&\quad + \left(\left(x_{it}w_{it}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)\left(\left(x_{is}x_{is}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \\
&\quad + \left(\left(x_{it}w_{it}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)\left(\left(x_{is}w_{is}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)'.
\end{aligned}$$

Combining these two decomposition results, we have

$$\begin{aligned}
&\left(\hat{d}_{it}(g^0, h) - \bar{\hat{d}}_i(g^0, h)\right)\left(\hat{d}_{is}(g^0, h') - \bar{\hat{d}}_i(g^0, h')\right) \\
&\quad - \left(d_{it}(g^0, h) - \bar{d}_i(g^0, h)\right)\left(d_{is}(g^0, h') - \bar{d}_i(g^0, h')\right) \\
&= \left(\hat{\delta}_i(h) - \delta_i(h)\right)'\Big(-(u_{it}x_{it} - \overline{u_i x_i})\left(\left(x_{is}x_{is}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \\
&\quad - (u_{it}x_{it} - \overline{u_i x_i})\left(\left(x_{is}w_{is}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)' \\
&\quad - \left(\left(x_{it}x_{it}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)(u_{is}x_{is} - \overline{u_i x_i})' \\
&\quad - \left(\left(x_{it}w_{it}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)(u_{is}x_{is} - \overline{u_i x_i})' \\
&\quad + \left(\left(x_{it}x_{it}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)\left(\left(x_{is}x_{is}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \\
&\quad + \left(\left(x_{it}x_{it}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)\left(\left(x_{is}w_{is}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)' \\
&\quad + \left(\left(x_{it}w_{it}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)\left(\left(x_{is}x_{is}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \\
&\quad + \left(\left(x_{it}w_{it}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)\left(\left(x_{is}w_{is}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)'\Big)\left(\hat{\delta}_i(h') - \delta_i(h')\right) \\
&\quad + \left(\hat{\delta}_i(h) - \delta_i(h)\right)'\Big(-(u_{it}x_{it} - \overline{u_i x_i})\left(\left(x_{is}x_{is}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \\
&\quad - (u_{it}x_{it} - \overline{u_i x_i})\left(\left(x_{is}w_{is}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)' \\
&\quad - \left(\left(x_{it}x_{it}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)(u_{is}x_{is} - \overline{u_i x_i})' \\
&\quad - \left(\left(x_{it}w_{it}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)(u_{is}x_{is} - \overline{u_i x_i})' \\
&\quad + \left(\left(x_{it}x_{it}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)\left(\left(x_{is}x_{is}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \\
&\quad + \left(\left(x_{it}x_{it}' - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)\left(\left(x_{is}w_{is}' - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)'
\end{aligned}$$

$$
\begin{aligned}
&+ \big(\big(x_{it}w_{it}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)\big(\big(x_{is}x_{is}' - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \\
&+ \big(\big(x_{it}w_{it}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)\big(\big(x_{is}w_{is}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)'\big)\delta_i(h') \\
&+ \delta_i(h)'\big(-(u_{it}x_{it} - \overline{u_i x_i})\big(\big(x_{is}x_{is}' - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \\
&- (u_{it}x_{it} - \overline{u_i x_i})\big(\big(x_{is}w_{is}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)' \\
&- \big(\big(x_{it}x_{it}' - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)(u_{is}x_{is} - \overline{u_i x_i})' \\
&- \big(\big(x_{it}w_{it}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)(u_{is}x_{is} - \overline{u_i x_i})' \\
&+ \big(\big(x_{it}x_{it}' - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)\big(\big(x_{is}x_{is}' - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \\
&+ \big(\big(x_{it}x_{it}' - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)\big(\big(x_{is}w_{is}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)' \\
&+ \big(\big(x_{it}w_{it}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)\big(\big(x_{is}x_{is}' - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \\
&+ \big(\big(x_{it}w_{it}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)\big(\big(x_{is}w_{is}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)'\big)\big(\hat{\delta}_i(h') - \delta_i(h')\big) \\
&+ \delta_i(h)'\big(-(u_{it}x_{it} - \overline{u_i x_i})\big(\big(x_{is}x_{is}' - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \\
&- (u_{it}x_{it} - \overline{u_i x_i})\big(\big(x_{is}w_{is}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)' \\
&- \big(\big(x_{it}x_{it}' - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)(u_{is}x_{is} - \overline{u_i x_i})' \\
&- \big(\big(x_{it}w_{it}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)(u_{is}x_{is} - \overline{u_i x_i})' \\
&+ \big(\big(x_{it}x_{it}' - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)\big(\big(x_{is}x_{is}' - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \\
&+ \big(\big(x_{it}x_{it}' - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)\big(\big(x_{is}w_{is}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)' \\
&+ \big(\big(x_{it}w_{it}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)\big(\big(x_{is}x_{is}' - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \\
&+ \big(\big(x_{it}w_{it}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)\big(\big(x_{is}w_{is}' - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)'\big)\delta_i(h') \\
&+ \big(\hat{\delta}_i(h) - \delta_i(h)\big)'(u_{it}x_{it} - \overline{u_i x_i})(u_{is}x_{is} - \overline{u_i x_i})'\big(\hat{\delta}_i(h') - \delta_i(h')\big) \\
&+ \big(\hat{\delta}_i(h) - \delta_i(h)\big)'(u_{it}x_{it} - \overline{u_i x_i})(u_{is}x_{is} - \overline{u_i x_i})'\delta_i(h') \\
&+ \delta_i(h)'(u_{it}x_{it} - \overline{u_i x_i})(u_{is}x_{is} - \overline{u_i x_i})'\big(\hat{\delta}_i(h') - \delta_i(h')\big).
\end{aligned}
$$

It thus holds that

$$
\begin{aligned}
&\widehat{\Xi}_i(h, h') - \widetilde{\Xi}_i(h, h') \\
&= -\big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i}\big) \\
&\quad \times \big(\big(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)'\big(\hat{\delta}_i(h) - \delta_i(h)\big)
\end{aligned}
$$

$$
\begin{aligned}
&-\big(\hat{\delta}_i(h)-\delta_i(h)\big)'\sum_{j=-T+1}^{T-1}K_N^{(j)}\frac{1}{T}\sum_{t=|j|+1}^{T}\big(u_{it}^{(+j)}x_{it}^{(+j)}-\overline{u_ix_i}\big)\\
&\times\Big(\big(x_{it}^{(-j)}w_{it}^{(-j)\prime}-\overline{x_iw_i}\big)\big(\hat{\theta}^w-\theta^w\big)\Big)'\big(\hat{\delta}_i(h)-\delta_i(h)\big)\\
&-\big(\hat{\delta}_i(h)-\delta_i(h)\big)'\sum_{j=-T+1}^{T-1}K_N^{(j)}\frac{1}{T}\sum_{t=|j|+1}^{T}\Big(\big(x_{it}^{(+j)}x_{it}^{(+j)\prime}-\overline{x_ix_i}\big)(\hat{\theta}_{g_i^0}-\theta_{g_i^0})\Big)\\
&\times\big(u_{i,t-\max(0,j)}x_{it}^{(-j)}-\overline{u_ix_i}\big)'\big(\hat{\delta}_i(h)-\delta_i(h)\big)\\
&-\big(\hat{\delta}_i(h)-\delta_i(h)\big)'\sum_{j=-T+1}^{T-1}K_N^{(j)}\frac{1}{T}\sum_{t=|j|+1}^{T}\Big(\big(x_{it}^{(+j)}w_{it}^{(+j)\prime}-\overline{x_iw_i}\big)\big(\hat{\theta}^w-\theta^w\big)\Big)\\
&\times\big(u_{it}^{(-j)}x_{it}^{(-j)}-\overline{u_ix_i}\big)'\big(\hat{\delta}_i(h)-\delta_i(h)\big)\\
&+\big(\hat{\delta}_i(h)-\delta_i(h)\big)'\sum_{j=-T+1}^{T-1}K_N^{(j)}\frac{1}{T}\sum_{t=|j|+1}^{T}\Big(\big(x_{it}^{(+j)}x_{it}^{(+j)\prime}-\overline{x_ix_i}\big)(\hat{\theta}_{g_i^0}-\theta_{g_i^0})\Big)\\
&\times\Big(\big(x_{it}^{(-j)}x_{it}^{(-j)\prime}-\overline{x_ix_i}\big)(\hat{\theta}_{g_i^0}-\theta_{g_i^0})\Big)'\big(\hat{\delta}_i(h)-\delta_i(h)\big)\\
&+\big(\hat{\delta}_i(h)-\delta_i(h)\big)'\sum_{j=-T+1}^{T-1}K_N^{(j)}\frac{1}{T}\sum_{t=|j|+1}^{T}\Big(\big(x_{it}^{(+j)}x_{it}^{(+j)\prime}-\overline{x_ix_i}\big)(\hat{\theta}_{g_i^0}-\theta_{g_i^0})\Big)\\
&\times\Big(\big(x_{it}^{(-j)}w_{it}^{(-j)\prime}-\overline{x_iw_i}\big)\big(\hat{\theta}^w-\theta^w\big)\Big)'\big(\hat{\delta}_i(h)-\delta_i(h)\big)\\
&+\big(\hat{\delta}_i(h)-\delta_i(h)\big)'\sum_{j=-T+1}^{T-1}K_N^{(j)}\frac{1}{T}\sum_{t=|j|+1}^{T}\Big(\big(x_{it}^{(+j)}w_{it}^{(+j)\prime}-\overline{x_iw_i}\big)\big(\hat{\theta}^w-\theta^w\big)\Big)\\
&\times\Big(\big(x_{it}^{(-j)}x_{it}^{(-j)\prime}-\overline{x_ix_i}\big)(\hat{\theta}_{g_i^0}-\theta_{g_i^0})\Big)'\big(\hat{\delta}_i(h)-\delta_i(h)\big)\\
&+\big(\hat{\delta}_i(h)-\delta_i(h)\big)'\sum_{j=-T+1}^{T-1}K_N^{(j)}\frac{1}{T}\sum_{t=|j|+1}^{T}\Big(\big(x_{it}^{(+j)}w_{it}^{(+j)\prime}-\overline{x_iw_i}\big)\big(\hat{\theta}^w-\theta^w\big)\Big)\\
&\times\Big(\big(x_{it}^{(-j)}w_{it}^{(-j)\prime}-\overline{x_iw_i}\big)\big(\hat{\theta}^w-\theta^w\big)\Big)')\big(\hat{\delta}_i(h')-\delta_i(h')\big)\\
&-\big(\hat{\delta}_i(h)-\delta_i(h)\big)'\sum_{j=-T+1}^{T-1}K_N^{(j)}\frac{1}{T}\sum_{t=|j|+1}^{T}\big(u_{it}^{(+j)}x_{it}^{(+j)}-\overline{u_ix_i}\big)\\
&\times\Big(\big(x_{it}^{(-j)}x_{it}^{(-j)\prime}-\overline{x_ix_i}\big)(\hat{\theta}_{g_i^0}-\theta_{g_i^0})\Big)'\delta_i(h)\\
&-\big(\hat{\delta}_i(h)-\delta_i(h)\big)'\sum_{j=-T+1}^{T-1}K_N^{(j)}\frac{1}{T}\sum_{t=|j|+1}^{T}\big(u_{it}^{(+j)}x_{it}^{(+j)}-\overline{u_ix_i}\big)\\
&\times\Big(\big(x_{it}^{(-j)}w_{it}^{(-j)\prime}-\overline{x_iw_i}\big)\big(\hat{\theta}^w-\theta^w\big)\Big)'\delta_i(h)
\end{aligned}
$$

$$
\begin{aligned}
&- \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big) \\
&\times \big(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\big)' \delta_i(h) \\
&- \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big) \\
&\times \big(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\big)' \delta_i(h) \\
&+ \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big) \\
&\times \big(\big(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \delta_i(h) \\
&+ \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big) \\
&\times \big(\big(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)' \delta_i(h) \\
&+ \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big) \\
&\times \big(\big(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \delta_i(h) \\
&+ \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big) \\
&\times \big(\big(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)')\delta_i\big(h'\big) \\
&- \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i}\big) \\
&\times \big(\big(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \big(\hat{\delta}_i(h) - \delta_i(h)\big) \\
&- \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i}\big) \\
&\times \big(\big(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)' \big(\hat{\delta}_i(h) - \delta_i(h)\big) \\
&- \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big) \\
&\times \big(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\big)' \big(\hat{\delta}_i(h) - \delta_i(h)\big)
\end{aligned}
$$

$$
\begin{aligned}
&- \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right) \\
&\times \left(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\right)' \left(\hat{\delta}_i(h) - \delta_i(h)\right) \\
&+ \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right) \\
&\times \left(\left(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \left(\hat{\delta}_i(h) - \delta_i(h)\right) \\
&+ \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right) \\
&\times \left(\left(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)' \left(\hat{\delta}_i(h) - \delta_i(h)\right) \\
&+ \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right) \\
&\times \left(\left(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \left(\hat{\delta}_i(h) - \delta_i(h)\right) \\
&+ \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right) \\
&\times \left(\left(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)') \left(\hat{\delta}_i(h') - \delta_i(h')\right) \\
&+ \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \left(- \left(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i}\right) \right. \\
&\times \left(\left(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \delta_i(h) \\
&- \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \left(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i}\right) \\
&\times \left(\left(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)' \delta_i(h) \\
&- \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right) \\
&\times \left(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\right)' \delta_i(h) \\
&- \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right) \\
&\times \left(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\right)' \delta_i(h)
\end{aligned}
$$

$$
\begin{aligned}
&+ \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big) \\
&\times \big(\big(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \delta_i(h) \\
&+ \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big) \\
&\times \big(\big(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)' \delta_i(h) \\
&+ \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big) \\
&\times \big(\big(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \delta_i(h) \\
&+ \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big) \\
&\times \big(\big(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)')\delta_i(h') \\
&+ \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i}\big) \\
&\times \big(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\big)' \big(\hat{\delta}_i(h') - \delta_i(h')\big) \\
&+ \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i}\big) \\
&\times \big(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\big)' \delta_i(h') \\
&+ \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i}\big) \\
&\times \big(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\big)' \big(\hat{\delta}_i(h') - \delta_i(h')\big).
\end{aligned}
$$

We examine each term. For vector a, let a_p denote the pth element of a. Let d_x be the dimension of x_{it}. With probability at least $a_{\theta,N}$, it holds

$$
\begin{aligned}
&\frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Bigg| \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i}\big) \\
&\quad \times \big(\big(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \big(\hat{\delta}_i(h) - \delta_i(h)\big) \Bigg|
\end{aligned}
$$

$$
\begin{aligned}
&\leq C r_{\theta,N}^{2} \iota_N^{-2} \frac{1}{\sigma_i} \Bigg| \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(v_{it}^{(j)} x_{it}^{(+j)} - \overline{v_i x_i}\big) \\
&\quad \times \big(\big(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \Bigg|_\infty \\
&\leq C r_{\theta,N}^{3} \iota_N^{-2} \Big(\min_{1\leq i\leq N} \sigma_i\Big)^{-1} \sum_{p=1}^{d_x} \Bigg| \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(v_{it}^{(j)} x_{it}^{(+j)} - \overline{v_i x_i}\big) \\
&\quad \times \left(x_{i,t-\max(0,j),p} x_{it}^{(-j)\prime} - \frac{1}{T} \sum_{u=1}^{T} x_{iu,p} x_{iu}' \right) \Bigg|_\infty \\
&\leq C r_{\theta,N}^{3} \iota_N^{-2} \sum_{p=1}^{d_x} \Big(\min_{1\leq i\leq N} \sigma_i\Big)^{-1} \Bigg| \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(v_{it}^{(j)} x_{it}^{(+j)} - \overline{v_i x_i}\big) \\
&\quad \times \left(x_{i,t-\max(0,j),p} x_{it}^{(-j)\prime} - \frac{1}{T} \sum_{u=1}^{T} x_{iu,p} x_{iu}' \right) \\
&\quad - \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}_P\big[v_{it} x_{it}\big(x_{is,p} x_{is} - \mathbb{E}_p(x_{is,p} x_{is})\big)\big] \Bigg|_\infty \\
&\quad + C r_{\theta,N}^{3} \iota_N^{-2} \Big(\min_{1\leq i\leq N} \sigma_i\Big)^{-1} \sum_{p=1}^{d_x} \Bigg| \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}_P\big[v_{it} x_{it}\big(x_{is,p} x_{is} - \mathbb{E}_p(x_{is,p} x_{is})\big)\big] \Bigg|_\infty
\end{aligned}
$$

Applying Lemma A.6 to

$$
\begin{aligned}
&\Bigg| \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \sum_{t=|j|+1}^{T} \big(v_{it}^{(j)} x_{it}^{(+j)} - \overline{v_i x_i}\big) \left(x_{i,t-\max(0,j),p} x_{it}^{(-j)\prime} - \frac{1}{T} \sum_{u=1}^{T} x_{iu,p} x_{iu}' \right) \\
&\quad - \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}_P\big[v_{it} x_{it}\big(x_{is,p} x_{is} - \mathbb{E}_p(x_{is,p} x_{is})\big)\big] \Bigg|_\infty
\end{aligned}
$$

and Lemma A.11 to

$$
\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}_P\big[v_{it} x_{it}\big(x_{is,p} x_{is} - \mathbb{E}_p(x_{is,p} x_{is})\big)\big]
$$

we obtain

$$
\begin{aligned}
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \big| \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T} \Bigg\{ \sum_{t=|j|+1}^{T} \big(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i}\big) \\
&\quad \times \big(\big(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \big(\hat{\delta}_i(h) - \delta_i(h)\big) \big| \Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
&\precsim_P r_{\theta,N}^3 \iota_N^{-2} T^{-c_1}\big(\log\big(N(G-1)\big)\big)^{c_2}\Big(\min_{1\le i\le N}\sigma_i\Big)^{-1} \\
&\quad + r_{\theta,N}^3 \iota_N^{-2} T^{-\rho}\Big(\min_{1\le i\le N}\sigma_i\Big)^{-1} \\
&\quad + r_{\theta,N}^3 \iota_N^{-2}\Big(\min_{1\le i\le N}\sigma_i\Big)^{-1} \\
&\precsim_P r_{\theta,N}^3 \iota_N^{-2}\Big(\min_{1\le i\le N}\sigma_i\Big)^{-1},
\end{aligned}
$$

where $\precsim_P$ signifies that for A_N and B_N, $A_N \precsim_P B_N$ if $\sup_{P\in\mathbb{P}_N} P(A_N > B_N\zeta_N) = o(1)$ for any $\zeta_N \to \infty$, and the last $\precsim_P$ follows by $b_N^{\mathrm{LV}} \to 0$. Following the same argument, we have

$$
\begin{aligned}
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|}\Bigg|\big(\hat\delta_i(h)-\delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \frac{1}{T}\Bigg\{\sum_{t=|j|+1}^{T}\big(u_{it}^{(+j)}x_{it}^{(+j)} - \overline{u_i x_i}\big) \\
&\quad \times \big(\big(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i}\big)\big(\hat\theta^w - \theta^w\big)\big)'\big(\hat\delta_i(h)-\delta_i(h)\big)\Bigg\}\Bigg| \\
&\precsim_P r_{\theta,N}^3 \iota_N^{-2}\Big(\min_{1\le i\le N}\sigma_i\Big)^{-1},
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|}\Bigg|\big(\hat\delta_i(h)-\delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \\
&\quad \times \frac{1}{T}\sum_{t=|j|+1}^{T}\Big\{\big(\big(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i}\big)\big(\hat\theta_{g_i^0} - \theta_{g_i^0}\big)\big) \\
&\quad \times \big(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\big)'\big(\hat\delta_i(h)-\delta_i(h)\big)\Big\}\Bigg| \\
&\precsim_P r_{\theta,N}^3 \iota_N^{-2}\Big(\min_{1\le i\le N}\sigma_i\Big)^{-1},
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|}\Bigg|\big(\hat\delta_i(h)-\delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \\
&\quad \times \frac{1}{T}\sum_{t=|j|+1}^{T}\big(\big(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i}\big)\big(\hat\theta^w - \theta^w\big)\big) \\
&\quad \times \big(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\big)'\big(\hat\delta_i(h)-\delta_i(h)\big)\Bigg| \\
&\precsim_P r_{\theta,N}^3 \iota_N^{-2}\Big(\min_{1\le i\le N}\sigma_i\Big)^{-1},
\end{aligned}
$$

$$
\begin{aligned}
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Big| \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \\
&\quad \times \frac{1}{T} \sum_{t=|j|+1}^{T} \Big(\big(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i} \big) (\hat{\theta}_{g_i^0} - \theta_{g_i^0}) \Big) \\
&\quad \times \Big(\big(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i} \big) (\hat{\theta}_{g_i^0} - \theta_{g_i^0}) \Big)' \big(\hat{\delta}_i(h) - \delta_i(h)\big) \Big| \\
&\precsim_P r_{\theta,N}^4 \iota_N^{-2} \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-2}, \\
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Big| \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \\
&\quad \times \frac{1}{T} \sum_{t=|j|+1}^{T} \Big(\big(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i} \big) (\hat{\theta}_{g_i^0} - \theta_{g_i^0}) \Big) \\
&\quad \times \Big(\big(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i} \big) \big(\hat{\theta}^w - \theta^w\big) \Big)' \big(\hat{\delta}_i(h) - \delta_i(h)\big) \Big| \\
&\precsim_P r_{\theta,N}^4 \iota_N^{-2} \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-2}, \\
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Big| \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \\
&\quad \times \frac{1}{T} \sum_{t=|j|+1}^{T} \Big(\big(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i} \big) \big(\hat{\theta}^w - \theta^w\big) \Big) \\
&\quad \times \Big(\big(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i} \big) (\hat{\theta}_{g_i^0} - \theta_{g_i^0}) \Big)' \big(\hat{\delta}_i(h) - \delta_i(h)\big) \Big| \\
&\precsim_P r_{\theta,N}^4 \iota_N^{-2} \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-2}, \\
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Big| \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \\
&\quad \times \frac{1}{T} \sum_{t=|j|+1}^{T} \Big(\big(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i} \big) \big(\hat{\theta}^w - \theta^w\big) \Big) \\
&\quad \times \Big(\big(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i} \big) \big(\hat{\theta}^w - \theta^w\big) \Big)') \big(\hat{\delta}_i(h') - \delta_i(h')\big) \Big|
\end{aligned}
$$

$$\lesssim_P r_{\theta,N}^4 \iota_N^{-2} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-2},$$

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Bigg| \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)}$$
$$\times \frac{1}{T} \sum_{t=|j|+1}^{T} \big(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i}\big)$$
$$\times \big(\big(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)' \delta_i(h) \Bigg|$$
$$\lesssim_P r_{\theta,N}^2 \iota_N^{-1} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-1},$$

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Bigg| \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)}$$
$$\times \frac{1}{T} \sum_{t=|j|+1}^{T} \big(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i}\big)$$
$$\times \big(\big(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)' \delta_i(h) \Bigg|$$
$$\lesssim_P r_{\theta,N}^2 \iota_N^{-1} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-1},$$

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Bigg| \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)}$$
$$\times \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i}\big)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\big)$$
$$\times \big(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\big)' \delta_i(h) \Bigg|$$
$$\lesssim_P r_{\theta,N}^2 \iota_N^{-1} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-1},$$

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Bigg| \big(\hat{\delta}_i(h) - \delta_i(h)\big)' \sum_{j=-T+1}^{T-1} K_N^{(j)}$$
$$\times \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i}\big)\big(\hat{\theta}^w - \theta^w\big)\big)$$

$$
\begin{aligned}
&\quad \times \left(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\right)' \delta_i(h) \Bigg| \\
&\lesssim_P r_{\theta,N}^2 \iota_N^{-1} \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-1}, \\
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Bigg| \left(\hat{\delta}_i(h) - \delta_i(h)\right)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \\
&\quad \times \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right) \\
&\quad \times \left(\left(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \delta_i(h) \Bigg| \\
&\lesssim_P r_{\theta,N}^3 \iota_N^{-1} \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-2}, \\
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Bigg| \left(\hat{\delta}_i(h) - \delta_i(h)\right)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \\
&\quad \times \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right) \\
&\quad \times \left(\left(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right)' \delta_i(h) \Bigg| \\
&\lesssim_P r_{\theta,N}^3 \iota_N^{-1} \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-2}, \\
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Bigg| \left(\hat{\delta}_i(h) - \delta_i(h)\right)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \\
&\quad \times \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i}\right)\left(\hat{\theta}^w - \theta^w\right)\right) \\
&\quad \times \left(\left(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \delta_i(h) \Bigg| \\
&\lesssim_P r_{\theta,N}^3 \iota_N^{-1} \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-2}, \\
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Bigg| \left(\hat{\delta}_i(h) - \delta_i(h)\right)' \sum_{j=-T+1}^{T-1} K_N^{(j)}
\end{aligned}
$$

$$\times \frac{1}{T}\sum_{t=|j|+1}^{T}\big(\big(x_{it}^{(+j)}w_{it}^{(+j)\prime}-\overline{x_iw_i}\big)\big(\hat{\theta}^w-\theta^w\big)\big)$$

$$\times \big(\big(x_{it}^{(-j)}w_{it}^{(-j)\prime}-\overline{x_iw_i}\big)\big(\hat{\theta}^w-\theta^w\big)\big)^{\prime}\big)\delta_i\big(h^{\prime}\big)\Bigg|$$

$$\precsim_P r_{\theta,N}^3\iota_N^{-1}\Big(\min_{1\le i\le N}\sigma_i\Big)^{-2},$$

$$\sup_{i,h,h^{\prime}}\frac{1}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h^{\prime})\|}\Bigg|\delta_i(h)^{\prime}\sum_{j=-T+1}^{T-1}K_N^{(j)}$$

$$\times\frac{1}{T}\sum_{t=|j|+1}^{T}\left(u_{it}^{(+j)}x_{it}^{(+j)}-\frac{1}{T}\sum_{u=1}^{T}u_{iu}x_{iu}\right)$$

$$\times\big(\big(x_{it}^{(-j)}x_{it}^{(-j)\prime}-\overline{x_ix_i}\big)(\hat{\theta}_{g_i^0}-\theta_{g_i^0})\big)^{\prime}\big(\hat{\delta}_i(h)-\delta_i(h)\big)\Bigg|$$

$$\precsim_P r_{\theta,N}^2\iota_N^{-1}\Big(\min_{1\le i\le N}\sigma_i\Big)^{-1},$$

$$\sup_{i,h,h^{\prime}}\frac{1}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h^{\prime})\|}\Bigg|\delta_i(h)^{\prime}\sum_{j=-T+1}^{T-1}K_N^{(j)}$$

$$\times\frac{1}{T}\sum_{t=|j|+1}^{T}\left(u_{it}^{(+j)}x_{it}^{(+j)}-\frac{1}{T}\sum_{u=1}^{T}u_{iu}x_{iu}\right)$$

$$\times\big(\big(x_{it}^{(-j)}w_{it}^{(-j)\prime}-\overline{x_iw_i}\big)\big(\hat{\theta}^w-\theta^w\big)\big)^{\prime}\big(\hat{\delta}_i(h)-\delta_i(h)\big)\Bigg|$$

$$\precsim_P r_{\theta,N}^2\iota_N^{-1}\Big(\min_{1\le i\le N}\sigma_i\Big)^{-1},$$

$$\sup_{i,h,h^{\prime}}\frac{1}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h^{\prime})\|}\Bigg|\delta_i(h)^{\prime}\sum_{j=-T+1}^{T-1}K_N^{(j)}$$

$$\times\frac{1}{T}\sum_{t=|j|+1}^{T}\big(\big(x_{it}^{(+j)}x_{it}^{(+j)\prime}-\overline{x_ix_i}\big)(\hat{\theta}_{g_i^0}-\theta_{g_i^0})\big)$$

$$\times\left(u_{it}^{(-j)}x_{it}^{(-j)}-\frac{1}{T}\sum_{u=1}^{T}u_{iu}x_{iu}\right)^{\prime}\big(\hat{\delta}_i(h)-\delta_i(h)\big)\Bigg|$$

$$\precsim_P r_{\theta,N}^2\iota_N^{-1}\Big(\min_{1\le i\le N}\sigma_i\Big)^{-1},$$

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Bigg| \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)}$$

$$\times \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i} \right) \left(\hat{\theta}^w - \theta^w \right) \right)$$

$$\times \left(u_{it}^{(-j)} x_{it}^{(-j)} - \frac{1}{T} \sum_{u=1}^{T} u_{iu} x_{iu} \right)' \left(\hat{\delta}_i(h) - \delta_i(h) \right) \Bigg|$$

$$\precsim_P r_{\theta,N}^2 \iota_N^{-1} \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-1},$$

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Bigg| \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)}$$

$$\times \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i} \right) (\hat{\theta}_{g_i^0} - \theta_{g_i^0}) \right)$$

$$\times \left(\left(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i} \right) (\hat{\theta}_{g_i^0} - \theta_{g_i^0}) \right)' \left(\hat{\delta}_i(h) - \delta_i(h) \right) \Bigg|$$

$$\precsim_P r_{\theta,N}^3 \iota_N^{-1} \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-2},$$

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Bigg| \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)}$$

$$\times \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i} \right) (\hat{\theta}_{g_i^0} - \theta_{g_i^0}) \right)$$

$$\times \left(\left(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i} \right) \left(\hat{\theta}^w - \theta^w \right) \right)' \left(\hat{\delta}_i(h) - \delta_i(h) \right) \Bigg|$$

$$\precsim_P r_{\theta,N}^3 \iota_N^{-1} \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-2},$$

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Bigg| \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)}$$

$$\times \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i} \right) \left(\hat{\theta}^w - \theta^w \right) \right)$$

$$\times \left(\left(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i} \right) (\hat{\theta}_{g_i^0} - \theta_{g_i^0}) \right)' \left(\hat{\delta}_i(h) - \delta_i(h) \right) \Bigg|$$

$$\lesssim_P r_{\theta,N}^3 \iota_N^{-1} \Big(\min_{1\le i\le N} \sigma_i \Big)^{-2},$$

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Big| \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)}$$

$$\times \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i} \big) \big(\hat{\theta}^w - \theta^w \big) \big)$$

$$\times \big(\big(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i} \big) \big(\hat{\theta}^w - \theta^w \big) \big)') \big(\hat{\delta}_i(h') - \delta_i(h') \big) \Big|$$

$$\lesssim_P r_{\theta,N}^3 \iota_N^{-1} \Big(\min_{1\le i\le N} \sigma_i \Big)^{-2},$$

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} |\delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)}$$

$$\times \frac{1}{T} \sum_{t=|j|+1}^{T} (- \left(u_{it}^{(+j)} x_{it}^{(+j)} - \frac{1}{T} \sum_{u=1}^{T} u_{iu} x_{iu} \right)$$

$$\times \big(\big(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i} \big) (\hat{\theta}_{g_i^0} - \theta_{g_i^0}) \big)' \delta_i(h)|$$

$$\lesssim_P r_{\theta,N} \Big(\min_{1\le i\le N} \sigma_i \Big)^{-1},$$

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Big| \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)}$$

$$\times \frac{1}{T} \sum_{t=|j|+1}^{T} \left(u_{it}^{(+j)} x_{it}^{(+j)} - \frac{1}{T} \sum_{u=1}^{T} u_{iu} x_{iu} \right)$$

$$\times \big(\big(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i} \big) \big(\hat{\theta}^w - \theta^w \big) \big)' \delta_i(h) \Big|$$

$$\lesssim_P r_{\theta,N} \Big(\min_{1\le i\le N} \sigma_i \Big)^{-1},$$

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \Big| \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)}$$

$$\times \frac{1}{T} \sum_{t=|j|+1}^{T} \big(\big(x_{it}^{(+j)} x_{it}^{(+j)\prime} - \overline{x_i x_i} \big) (\hat{\theta}_{g_i^0} - \theta_{g_i^0}) \big)$$

$$\times \left(u_{it}^{(-j)} x_{it}^{(-j)} - \frac{1}{T} \sum_{u=1}^{T} u_{iu} x_{iu} \right)' \delta_i(h) \Big|$$

$$\lesssim_P r_{\theta,N}\Big(\min_{1\le i\le N}\sigma_i\Big)^{-1},$$

$$\sup_{i,h,h'}\frac{1}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|}\Bigg|\delta_i(h)'\sum_{j=-T+1}^{T-1}K_N^{(j)}$$

$$\times\frac{1}{T}\sum_{t=|j|+1}^{T}\big((x_{it}^{(+j)}w_{it}^{(+j)\prime}-\overline{x_iw_i}\big)\big(\hat\theta^w-\theta^w)\big)$$

$$\times\left(u_{it}^{(-j)}x_{it}^{(-j)}-\frac{1}{T}\sum_{u=1}^{T}u_{iu}x_{iu}\right)'\delta_i(h)\Bigg|$$

$$\lesssim_P r_{\theta,N}\Big(\min_{1\le i\le N}\sigma_i\Big)^{-1},$$

$$\sup_{i,h,h'}\frac{1}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|}\Bigg|\delta_i(h)'\sum_{j=-T+1}^{T-1}K_N^{(j)}$$

$$\times\frac{1}{T}\sum_{t=|j|+1}^{T}\big((x_{it}^{(+j)}x_{it}^{(+j)\prime}-\overline{x_ix_i}\big)(\hat\theta_{g_i^0}-\theta_{g_i^0})\big)$$

$$\times\big((x_{it}^{(-j)}x_{it}^{(-j)\prime}-\overline{x_ix_i}\big)(\hat\theta_{g_i^0}-\theta_{g_i^0})\big)'\delta_i(h)\Bigg|$$

$$\lesssim_P r_{\theta,N}^2\Big(\min_{1\le i\le N}\sigma_i\Big)^{-2},$$

$$\sup_{i,h,h'}\frac{1}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|}\Bigg|\delta_i(h)'\sum_{j=-T+1}^{T-1}K_N^{(j)}$$

$$\times\frac{1}{T}\sum_{t=|j|+1}^{T}\big((x_{it}^{(+j)}x_{it}^{(+j)\prime}-\overline{x_ix_i}\big)(\hat\theta_{g_i^0}-\theta_{g_i^0})\big)$$

$$\times\big((x_{it}^{(-j)}w_{it}^{(-j)\prime}-\overline{x_iw_i}\big)\big(\hat\theta^w-\theta^w)\big)'\delta_i(h)\Bigg|$$

$$\lesssim_P r_{\theta,N}^2\Big(\min_{1\le i\le N}\sigma_i\Big)^{-2},$$

$$\sup_{i,h,h'}\frac{1}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|}\Bigg|\delta_i(h)'\sum_{j=-T+1}^{T-1}K_N^{(j)}$$

$$\times\frac{1}{T}\sum_{t=|j|+1}^{T}\big((x_{it}^{(+j)}w_{it}^{(+j)\prime}-\overline{x_iw_i}\big)\big(\hat\theta^w-\theta^w)\big)$$

$$
\begin{aligned}
&\times \left(\left(x_{it}^{(-j)} x_{it}^{(-j)\prime} - \overline{x_i x_i}\right)(\hat{\theta}_{g_i^0} - \theta_{g_i^0})\right)' \delta_i(h) \Bigg| \\
&\lesssim_P r_{\theta,N}^2 \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-2}, \\
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \| \delta_i(h) \| \| \delta_i(h') \|} \Bigg| \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \\
&\quad \times \frac{1}{T} \sum_{t=|j|+1}^{T} \left(\left(x_{it}^{(+j)} w_{it}^{(+j)\prime} - \overline{x_i w_i} \right) \left(\hat{\theta}^w - \theta^w \right) \right) \\
&\quad \times \left(\left(x_{it}^{(-j)} w_{it}^{(-j)\prime} - \overline{x_i w_i} \right) \left(\hat{\theta}^w - \theta^w \right) \right)') \delta_i(h') \Bigg| \\
&\lesssim_P r_{\theta,N}^2 \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-2}, \\
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \| \delta_i(h) \| \| \delta_i(h') \|} \Bigg| \left(\hat{\delta}_i(h) - \delta_i(h) \right)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \\
&\quad \times \frac{1}{T} \sum_{t=|j|+1}^{T} \left(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i} \right) \\
&\quad \times \left(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i} \right)' \left(\hat{\delta}_i(h') - \delta_i(h') \right) \Bigg| \\
&\lesssim_P r_{\theta,N}^2 \iota_N^{-2} \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-1}, \\
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \| \delta_i(h) \| \| \delta_i(h') \|} \Bigg| \left(\hat{\delta}_i(h) - \delta_i(h) \right)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \\
&\quad \times \frac{1}{T} \sum_{t=|j|+1}^{T} \left(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i} \right) \\
&\quad \times \left(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i} \right)' \delta_i(h') \Bigg| \\
&\lesssim_P r_{\theta,N} \iota_N^{-1} \Big(\min_{1 \le i \le N} \sigma_i \Big)^{-1},
\end{aligned}
$$

and

$$
\begin{aligned}
&\sup_{i,h,h'} \frac{1}{\sigma_i^2 \| \delta_i(h) \| \| \delta_i(h') \|} \Bigg| \delta_i(h)' \sum_{j=-T+1}^{T-1} K_N^{(j)} \\
&\quad \times \frac{1}{T} \sum_{t=|j|+1}^{T} \left(u_{it}^{(+j)} x_{it}^{(+j)} - \overline{u_i x_i} \right)
\end{aligned}
$$

$$\times \left(u_{it}^{(-j)} x_{it}^{(-j)} - \overline{u_i x_i}\right)' \left(\hat{\delta}_i(h') - \delta_i(h')\right)\Bigg|$$

$$\precsim_P r_{\theta,N} \iota_N^{-1} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-1}.$$

To sum up, we have

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \big|\widehat{\Xi}_i(h,h') - \widetilde{\Xi}_i(h,h')\big|$$

$$\precsim_P r_{\theta,N}^3 \iota_N^{-2} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-1} + r_{\theta,N}^4 \iota_N^{-2} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-2} + r_{\theta,N}^2 \iota_N^{-1} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-1}$$

$$+ r_{\theta,N}^3 \iota_N^{-1} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-2} + r_{\theta,N} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-1} + r_{\theta,N}^2 \Big(\min_{1\le i\le N} \sigma_i\Big)^{-2}$$

$$+ r_{\theta,N}^2 \iota_N^{-2} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-1} + r_{\theta,N} \iota_N^{-1} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-1}$$

$$\precsim_P r_{\theta,N} \iota_N^{-1} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-1},$$

where the last $\precsim_P$ follows by $b_N^{\mathrm{LV}} \to 0$. We conclude that we have

$$\sup_{i,h,h'} \frac{1}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \big|\widehat{\Xi}_i(h,h') - \Xi_i(h,h')\big|$$

$$\precsim_P r_{\theta,N} \iota_N^{-1} \Big(\min_{1\le i\le N} \sigma_i\Big)^{-1} + T^{-c_1} (\log N)^{c_2} + T^{-\rho}. \qquad \square$$

LEMMA A.6. *Let ξ_{it} be a random vector consisting of distinct elements of $v_{it} x_{it}$, $x_{it} x_{it}' - \mathbb{E}_P(x_{it} x_{it}')$, and $w_{it} x_{it}' - \mathbb{E}_P(w_{it} x_{it}')$. Let $\mathbb{P}_N$ be the set of probability measures which satisfy Assumption* 1.6 *and Assumption* 1.7 *with identical choices of a, b, d_1, and d_2. Let ζ_N be a sequence satisfying $\zeta_N \to \infty$ as $N \to \infty$. Assume $T^{\delta_1} \le N$ for some universal constant $\delta_1 > 0$. Let Assumption* 2 *hold and $\kappa_N \asymp T^\rho$ where $0 < \rho < (\vartheta - 1)/(3\vartheta - 2)$. Then there exist two constants $c_1 > 0$ depending only on (ρ, ϑ) and $c_2 > 0$ depending only on (d_2, ϑ), such that*

$$\sup_{P\in\mathbb{P}_N} P\Bigg(\sup_{1\le i\le N} \Bigg| \sum_{j=-T+1}^{T-1} K\Big(\frac{j}{\kappa_N}\Big)$$

$$\times \frac{1}{T} \sum_{t=|j|+1}^{T} \Bigg(\xi_{i,t+\min(0,j)} - \frac{1}{T}\sum_{u=1}^{T} \xi_{iu}\Bigg)\Bigg(\xi_{i,t-\max(0,j)} - \frac{1}{T}\sum_{u=1}^{T} \xi_{iu}\Bigg)'$$

$$- \Bigg(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathbb{E}_P\big[\xi_{it}\xi_{is}'\big]\Bigg)\Bigg|_\infty$$

$$> \zeta_N\big(T^{-c_1}(\log N)^{c_2} + T^{-\rho}\big)\Bigg) = o(1).$$

Proof. We apply Theorem 11(i) of Chang, Chen, and Wu (2024). In their theorem, B_n is the bound for the Orlicz norm. In our case, this bound depends only on K, a, and d_1 under Assumption 1.6 (by Lemma 8.1 in Kosorok (2008)) and Lemma A.7. Their γ_1 is equal to one in our case by Assumption 1.6. By Lemma A.9, γ_2 in Chang, Chen, and Wu (2024) depends only on d_2. The conditions for the kernel and the bandwidth are assumed. Thus, Theorem 11(i) in Chang, Chen, and Wu (2024) can be applied. Note that their results are stated in terms of stochastic order, but an inspection of their proof reveals that the constant terms hidden in the stochastic order depend only on the constants in the assumptions. □

Lemma A.7 (Tail bounds for functions). *Suppose that two random variances X_1 and X_2 satisfy $P(|X_a| > x) \leq C_a \exp(-b_a x^{d_a})$ for $a = 0, 1$, then $P(|X_1 X_2| > x) \leq C \exp(-bx^d)$ for some positive constants C, b, and d_2, and $P(|X_1 + X_2| > x) \leq C' \exp(-b' x^{d'})$ for some constants C', b', and d'.*

Proof. The first statement follows because

$$\begin{aligned} P\big(|X_1 X_2| > x\big) &\leq P\big(|X_1| > \sqrt{x}\big) + P\big(|X_2| > \sqrt{x}\big) \\ &\leq C_1 \exp\big(-b_1 x^{d_1/2}\big) + C_2 \exp\big(-b_2 x^{d_2/2}\big) \\ &\leq 2\max(C_1, C_2)\exp\big(-\min(b_1, b_2) z^{\min(d_1,d_2)/2}\big). \end{aligned}$$

For the second statement, we have

$$\begin{aligned} P\big(|X_1 + X_2| > x\big) &\leq P\big(|X_1| > x/2\big) + P\big(|X_2| > x/2\big) \\ &\leq C_1 \exp\big(-b_1/2^{d_1} x^{d_1}\big) + C_2 \exp\big(-b_2/2^{d_2} x^{d_2}\big) \\ &\leq 2\max(C_1, C_2)\exp\big(-\min\big(b_1/2^{d_1}, b_2/2^{d_2}\big) x^{\min(d_1,d_2)}\big). \end{aligned}$$ □

Lemma A.8 (Tail bounds for norms). *Let X_1 and X_2 denote two random vectors such that there are constants K, b, and d such that for any component Y of X_1 and X_2, $P(|Y| > x) \leq C\exp(-bx^d)$. Then there are constants C', b', and d' such that*

$$P\big(\|X_1\| > x\big) \leq C' \exp\big(-b' x^{d'}\big),$$

$$P\big(\|X_1\|^2 > x\big) \leq C' \exp\big(-b' x^{d'}\big),$$

$$P\big(\|X_1\|\|X_2\| > x\big) \leq C' \exp\big(-b' x^{d'}\big).$$

Proof. The second statement follows from the first statement. The third statement follows from the second statement of this lemma and the first statement of Lemma A.7. It remains to prove the first statement. Let $X_1 = (Y_1, \ldots, Y_p)$ and note that $P(|Y_j^2| > x) \leq C\exp(-bx^{d/2})$ for $j = 1, \ldots, p$. Now, the first statement follows from writing

$$\|X_1\|^2 = Y_1^2 + \cdots Y_p^2$$

and applying the second statement of Lemma A.7 repeatedly. □

LEMMA A.9 (Functions of mixing sequences). *Suppose that* (x_{it}, w_{it}, v_{it}) *is a strong mixing sequence over* t *with mixing coefficients* $\sup_i a_i[t] \leq C\exp(-at^d)$ *for constants* C, a, *and* d, *then so is* $g(x_{it}, w_{it}, v_{it})$ *where* g *is a measurable function.*

PROOF. The proof follows the argument in the proof of Theorem 14.1 in Davidson (1994). □

LEMMA A.10 (Large quantiles of the normal distribution). *Let* X *denote a standard normal vector with* $p \times p$ *correlation matrix* Ω *and let* $0 \leq d < 2$. *Let* $c_{\alpha,N}$ *denote the* $1 - \alpha/N$ *quantile of* X. *Then there is a constant* N_0 *that depends only on* $\underline{\alpha}$ *and* d *such that for* $\underline{\alpha} \leq \alpha \leq 1$ *and* $N \geq N_0$,

$$\sqrt{d\log(N/\alpha)} \leq c_{\alpha,N} \leq \sqrt{2\log p} + \sqrt{2\log(N/\alpha)}.$$

PROOF. The upper bound is given in Lemma D.4 in Chernozhukov, Chetverikov, and Kato (2019). To prove the lower bound put $a_N = \sqrt{d\log(N/\alpha)}$. Let Φ denote the cumulative distribution function of a standard normal random variable, and let ϕ denote its probability density function. Gordon's lower bound (see, e.g., Duembgen (2010)) states that

$$1 - \Phi(x) > \frac{\phi(x)}{x\bigl(1 - 1/x^2\bigr)}$$

for $x > 0$, and thus $1 - \Phi(x) > \frac{1}{2}\phi(x)/x$ for $x > \sqrt{2}$. Therefore,

$$\begin{aligned} P\Bigl(\max_{j=1,\dots,p} X_j > a_N\Bigr) &\geq P(X_1 > a_N) \\ &= 1 - \Phi(a_N) \\ &> \frac{\phi(a_N)}{2a_N} = \frac{\exp\Bigl(-\frac{a_N^2}{2}\Bigr)}{a_N\sqrt{8\pi}} = \frac{(\alpha/N)^{d/2}}{a_N\sqrt{8\pi}} = \alpha/N\left(\frac{(N/\alpha)^{1-d/2}}{a_N\sqrt{8\pi}}\right) \\ &\geq \alpha/N\left(\frac{N^{1-d/2}}{\sqrt{8d\pi\log(N/\underline{\alpha})}}\right) \geq \alpha/N, \end{aligned}$$

where the last inequality holds for $N \geq N_0$ and N_0 is chosen such that $N \geq N_0$ implies that $N^{1-d/2}/\sqrt{8d\pi\log(N/\underline{\alpha})} \geq 1$. Such an N_0 can be found since $N^{1-d/2}/\sqrt{8d\pi\log(N/\underline{\alpha})} \to \infty$. The inequality $P(\max_{j=1,\dots,p} X_j > a_N) > \alpha/N$ implies $c_{\alpha,N} \geq a_N$. □

LEMMA A.11 (Long-run variance is finite). *Let* ξ_{it} *denote any element of the vectors* $v_{it}x_{it}$, $\mathrm{vec}(x_{it}x_{it}') - \mathbb{E}_P\mathrm{vec}(x_{it}x_{it}')$, $\mathrm{vec}(w_{it}x_{it}') - \mathbb{E}_P\mathrm{vec}(w_{it}x_{it}')$, *and* $\mathrm{vec}(v_{it}^2x_{it}x_{it}') - \mathbb{E}_P\mathrm{vec}(v_{it}^2x_{it}x_{it}')$, *or any of the random variables* $\|x_{it}\|^2$, $\|x_{it}\|\|w_{it}\|$, $|v_{it}|\|x_{it}\|$, $\|x_{it}\|^4$, $\|x_{it}^2\|\|w_{it}^2\|$, *and* $|v_{it}^2|\|x_{it}^2\|$. *Let* $\mathbb{P}_N$ *be a set of probability measures, which satisfy Assumptions* 1.6–1.8 *with identical choices of* a, b, d_1, *and* d_2. *Let*

$$s_{i,T}^2(P) = \max_{1\leq t\leq T}\Bigl(\mathbb{E}_P\bigl(\xi_{it}^2\bigr) + 2\sum_{s>t}\bigl|\mathbb{E}(\xi_{it}\xi_{is})\bigr|\Bigr).$$

Then there exists a constant $C_\xi < \infty$ such that

$$\limsup_{N,T\to\infty} \sup_{P\in\mathbb{P}_N} \max_{1\le i\le N} s^2_{i,T}(P) < C_\xi.$$

In particular,

$$\limsup_{N,T\to\infty} \sup_{P\in\mathbb{P}_N} \max_{1\le i\le N} \left| \mathrm{var}_P\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_{it}\right)\right| < C_\xi.$$

Proof. Lemma A.7 and Lemma A.8 imply $P(|\xi_{it}| > z) < \exp(-(z/a_\xi)^{d_{1,\xi}})$ for some a_ξ and $d_{1,\xi}$, which in turns implies that $\mathbb{E}_P(\xi_{it}^m) < M_{\xi_p}$ for some universal constant $M_\xi < \infty$ for any integer m by Lemma A.12. Moreover, Lemma A.9 implies that ξ_{it} is an α-mixing sequence with mixing coefficient $\sup_i \alpha_{i,\xi}[k] \le \exp(1 - b_\xi k^{d_{2,\xi}})$. Thus, by the argument in Galvao and Kato (2014, Section C.1), which is an application of Davydov (1968), it holds that, for any $s > t$ and any integer m,

$$\left|\mathbb{E}_P(\xi_{it}\xi_{is})\right| \le 12\left(\mathbb{E}_P\left(|\xi_{it}|^m\right)\right)^{2/m}\left(\alpha_{i,\xi}[s-t]\right)^{1-2/m}.$$

In particular,

$$2\sum_{s>T}\left|\mathbb{E}_P(\xi_{it}\xi_{is})\right| \le 24\left(\mathbb{E}_P\left(|\xi_{it}|^m\right)\right)^{2/m}\sum_{s>T}\left(\alpha_{i,\xi}[s-t]\right)^{1-2/m}.$$

The right-hand side is bounded by a constant C'_ξ that depends only on a, b, d_1, and d_2. This follows from the existence of moments and the mixing property of ξ_{it}. Note that the stationarity assumption is used to apply the result of Davydov (1968). □

Lemma A.12 (Exponential tail bound implies existence of moments). *Suppose that a random variable X satisfies that $P(|X| > x) < C\exp(-(x/a)^d)$ for some $C, a > 0$ and $d > 1$. Then, for any integer p, $\mathbb{E}|X|^p < M$ for M depending only on C, a, d, and p.*

Proof. By the argument given in Kosorok (2008, page 129), which is based on the series expansion of the exponential function, we have $(\mathbb{E}(|X|^p))^{1/p} \le p!\|X\|_{\psi_1}$ where $\|\cdot\|_{\psi_a}$ is the Orlicz norm with $\psi_a(x) = \exp(x^a) - 1$ as defined in the proof of Lemma A.13. By Kosorok (2008, Lemma 8.1), $\|X\|_{\psi_1}$ is bounded by a constant, which depends on C, a, and d. □

Lemma A.13 (Large CLT for mixing sequences). *Suppose that $\{\{X_{jt}\}_{j=1}^J\}_{t=1}^T$ is an α-mixing sequence (as a sequence indexed by t) with mixing coefficients $\alpha(k)$. Suppose that $T^{\delta_1} \le J$ for some $\delta_1 > 0$. Let $S_J = T^{-1/2}\sum_{t=1}^T (X_{1t}, \dots, X_{Jt})'$. Let $G \sim N(0, \Xi)$, where Ξ is the long-run covariance matrix of $(X_{1t}, \dots, X_{Jt})$. Assume the following three conditions:*

1. *There exist some universal constants $C_1 > 0$, $a > 0$, and $d_1 > 0$ such that $P(|X_{jt}| > x) < C_1\exp(-(1/a)^{d_1}x^{d_1})$ for all $t \in \{1, \dots, T\}$ and $j \in \{1, \dots, J\}$.*
2. *There exist some universal constants $C_2 > 1$, $b > 0$, and $d_2 > 0$ such that $\alpha(k) \le C_2\exp(-bk^{d_2})$ for any $k \ge 1$.*

3. *There exists a universal constant* $C_3 > 0$ *such that* $V_{T,j} \geq C_3$ *for any* $j \in \{1, \ldots, J\}$, *where* $V_{T,j} = \mathrm{var}(\sum_{t=1}^{T} X_{jt}/\sqrt{T})$.

For $a \in \mathbb{R}^N$, *define* $A(a) = \{x \in \mathbb{R}^N : x_j \leq a_j$ *for* $j = 1, \ldots, J\}$. *Let* $\mathcal{A} = \bigcup_{a \in \mathbb{R}^J} A(a)$. *Then it holds that*

$$\sup_{P \in \mathbb{P}} \sup_{A \in \mathcal{A}} \left| P(S_J \in A) - P(G \in A) \right| \lesssim \frac{(\log J)^{(1+2d_2)/(3d_2)}}{T^{1/9}} + \frac{(\log J)^{7/6}}{T^{1/9}}$$

provided that $(\log J)^{3-d_2} = o(T^{d_2/3})$ *and* $\mathbb{P}$ *is a collection of probabilities measures under which the above three conditions are satisfied with identical choices of* C_1, C_2, C_3, a, b, d_1, *and* d_2.

PROOF. The lemma follows by Theorem 1 of Chang, Chen, and Wu (2024), noting the remark at the beginning of Section 2.1 of Chang, Chen, and Wu (2024). Theorem 1 of Chang, Chen, and Wu (2024) has three conditions, and the second and third conditions are given in the statement of the lemma. The first condition is "there exist a sequence of constants $B_J \geq 1$ and a universal constant $d_1 \geq 1$ such that $\|X_{jt}\|_{\psi_{d_1}} \leq B_J$ for all $t \in \{1, \ldots, T\}$ and $j \in \{1, \ldots, J\}$," where $\|\xi\|_{\psi_\alpha} = \inf[\lambda > 0 : E(\psi_\alpha(|\xi|/\lambda)) \leq 1]$ for $\psi_\alpha(x) = \exp(x^\alpha) - 1$ (the Orlicz norm with ψ_α). By Lemma 8.1 of Kosorok (2008), $P(|X_{jt}| > x) < C_1 \exp(-(1/a)^{d_1} x^{d_1})$ implies this condition by taking $B_J = ((1 + C_1/(1/a)^{d_1}))^{1/d_1}$, which is constant if C_1, a, and d_1 are constant. □

LEMMA A.14. *Let* $\mathbb{P}_N$ *be the set of probability measures, which satisfy Assumption* 1 *with identical choices of* a, b, d_1, *and* d_2. *Assume that there are finite constants* $0 < \delta_1 < \delta_2$ *such that* $T^{\delta_1} \leq N \leq o(1)T^{\delta_2}$. *Let Assumption* 2 *hold with* $\kappa_N \asymp T^\rho$ *where* $0 < \rho < (\vartheta - 1)/(3\vartheta - 2)$. *Let*

$$b_N^{\mathrm{LV}*} = \frac{(\sqrt{T} \vee \log N) r_{\theta,N}}{\iota_N \min_{1 \leq i \leq N} \sigma_i} + T^{-c_1} (\log N)^{c_2} + T^{-\rho},$$

where $c_1 > 0$ *and* $c_2 > 0$ *are two constants defined in Lemma* A.5 *with* c_1 *depending only on* (ρ, ϑ) *and* c_2 *depending only on* (d_2, ϑ). *Assume that* $b_N^{\mathrm{LV}*} \to 0$. *Then for any sequence* ζ_N *such that* $\zeta_N \to \infty$ *as* $N, T \to \infty$,

$$\sup_{P \in \mathbb{P}_N} P\Big(\max_{1 \leq i \leq N} \max_{(h, h^*) \in \mathbb{G}^2} \big| \widehat{D}_i(g_i^0, h) - D_i(g_i^0, h) \big| > \zeta_N b_N^{\mathrm{LV}*} \Big) = o(1).$$

PROOF. Throughout the proof, let C denote a generic constant that does not depend on $P \in \mathbb{P}$ and whose value may change between different equations. Let $\delta_i(h) = \theta_{g_i^0} - \theta_h$ and $\hat{\delta}_i(h) = \hat{\theta}_{g_i^0} - \hat{\theta}_h$. Let $\widehat{\xi}_i(h) = \sqrt{\widehat{\Xi}_i(h, h)}$ and $\xi(h) = \sqrt{\Xi_i(h, h)}$. Let

$$b_N^{\mathrm{LV}} = \frac{r_{\theta,N}}{\iota_N \min_{1 \leq i \leq N} \sigma_i} + T^{-c_1} (\log N)^{c_2} + T^{-\rho}.$$

By the inequality $|1-\sqrt{a}| \leq |1-a|$,

$$\left|1 - \frac{\hat{\xi}_i(h)}{\xi_i(h)}\right| \leq \frac{\left|\hat{\xi}_i(h,h') - \xi_i(h,h')\right|}{\xi_i(h,h')}$$
$$\leq \frac{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|}{\xi_i(h,h')} \frac{\left|\hat{\xi}_i(h,h') - \xi_i(h,h')\right|}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|}.$$

By Assumption 1.4, $\xi_i(h,h')/(\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|)$ is bounded away from zero. Moreover, with probability approaching one,

$$\frac{\left|\hat{\xi}_i(h,h') - \xi_i(h,h')\right|}{\sigma_i^2 \|\delta_i(h)\| \|\delta_i(h')\|} \leq \zeta_N b_N^{\mathrm{LV}}$$

uniformly over $i = 1, \dots, N$ and $h, h' \in \mathbb{G} \setminus \{g_i^0\}$. Therefore, we can take

$$\left|1 - \hat{\xi}_i(h)/\xi_i(h)\right| \leq \zeta_N b_N^{LR}. \tag{A.11}$$

Next, we consider $\|T^{-1/2}\sum_{t=1}^T v_{it}x_{it}\|$. Consider any component $x_{it,p}$ of x_{it}. Set $\xi_{it} = v_{it}x_{it,p}$ in Lemma A.11 and conclude that

$$s_T^2 = \max_{1\leq i\leq N}\max_{1\leq t\leq T}\left(\mathbb{E}\left(\xi_{it}^2\right) + 2\sum_{s>t}\left|\mathbb{E}(\xi_{it}\xi_{is})\right|\right)$$

is bounded and fulfills the condition in Lemma A.15. By Lemma A.7 and Lemma A.9, ξ_{it} satisfies the tail and mixing conditions for X_{it} in Lemma A.15. Now, applying Lemma A.15

$$\max_{1\leq i\leq N}\left\|T^{-1/2}\sum_{t=1}^T v_{it}x_{it}\right\| = O_p(\log N). \tag{A.12}$$

Similarly, it can be argued that

$$\max_{1\leq i\leq N}\left|T^{-1/2}\sum_{t=1}^T\left(\|x_{it}\|^2 - \mathbb{E}\|x_{it}\|\right)\right| = O_p(\log N),$$
$$\max_{1\leq i\leq N}\left|T^{-1/2}\sum_{t=1}^T\left(\|x_{it}\|\|w_{it}\| - \mathbb{E}\left\|x_{it}\|w_{it}\|\right\|\right)\right| = O_p(\log N), \tag{A.13}$$
$$\max_{1\leq i\leq N}\left|T^{-1/2}\sum_{t=1}^T\left(|v_{it}|\|x_{it}\| - \mathbb{E}|v_{it}|\|x_{it}\|\right)\right| = O_p(\log N).$$

We write

$$\widehat{D}_i(h) - D_i(h) = \left(\frac{\xi_i(h)}{\hat{\xi}_i(h)} - 1\right)D_i(h) + \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^T\left(\hat{d}_{it}(h) - d_{it}(h)\right)/\left(\sigma_i\left\|\delta_i(h)\right\|\right)}{\hat{\xi}_i(h)/\left(\sigma_i\left\|\delta_i(h)\right\|\right)} \equiv J_1 + J_2.$$

We bound J_1 by writing

$$\left|\left(\frac{\xi_i(h)}{\hat{\xi}_i(h)}-1\right)D_i(h)\right| = \left|\frac{\xi_i(h)}{\hat{\xi}_i(h)}\left(1-\frac{\hat{\xi}_i(h)}{\xi_i(h)}\right)\frac{\frac{1}{T}\sum_{t=1}^{T} v_i x_i' \delta_i(h)/\|\delta_i(h)\|}{\xi_i(h)/(\sigma_i\|\delta_i(h)\|)}\right|$$

$$\leq \frac{\xi_i(h)}{\hat{\xi}_i(h)}\left|1-\frac{\hat{\xi}_i(h)}{\xi_i(h)}\right|\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{it}x_i\right\|\left(\frac{\xi_i(h)}{\sigma_i\|\delta_i(h)\|}\right)^{-1}.$$

The right-hand side is bounded by (A.11), (A.12), noting that $\xi_i(h)/(\sigma_i\|\delta_i(h)\|)$ is bounded away from zero by Assumption 1.4, and observing that

$$\frac{\hat{\xi}_i(h)}{\xi_i(h)} \geq 1-\left|\frac{\hat{\xi}_i(h)}{\xi_i(h)}-1\right|$$

in conjunction with (A.11) implies a lower bound on $\hat{\xi}_i(h)/\xi_i(h)$. Hence, J_1 is bounded by $C\zeta_N b_N^{\mathrm{LV}}\log N$ with probability approaching one.

To bound J_2, we derive a lower bound on its denominator from

$$\frac{\hat{\xi}_i(h)}{\sigma_i\|\delta_i(h)\|} = \frac{\xi_i(h)}{\sigma_i\|\delta_i(h)\|}\left\{\left(\frac{\hat{\xi}_i(h)}{\xi_i(h)}-1\right)+1\right\}$$

in conjunction with (A.11) and noting that $\xi_i(h)/(\sigma_i\|\delta_i(h)\|)$ is bounded away from zero by Assumption 1.4. For the numerator in J_2, we observe the following decomposition:

$$\frac{\hat{d}_{it}(h)-d_{it}(h)}{\sigma_i\|\delta_i(h)\|} = \frac{1}{2}\frac{x_{it}'(\hat{\theta}_{g_i^0}-\theta_{g_i^0})+w_{it}'(\hat{\theta}^w-\theta^w)}{\sigma_i}x_{it}'\left(\frac{\hat{\delta}_i(h)}{\|\delta_i(h)\|}\right)$$
$$-\frac{1}{2}v_{it}x_{it}'\frac{(\hat{\delta}_i(h)-\delta_i(h))}{\|\delta_i(h)\|}.$$

In the following arguments, we use that

$$\|\hat{\delta}_i(h)\|/\|\delta_i(h)\| \leq 1+\frac{\|\hat{\delta}_i(h)-\delta_i(h)\|}{\|\delta_i(h)\|}$$

is bounded by the fact that $r_{\theta,N}=o(1\wedge\iota_N)$. With probability at least $1-a_{\theta,N}$, we bound

$$\left|\frac{\hat{d}_{it}(h)-d_{it}(h)}{\sigma_i\|\delta_i(h)\|}\right| \leq C\left(\frac{\|x_{it}\|^2+\|x_{it}\|\|w_{it}\|}{\min_{1\leq i\leq N}\sigma_i}+\frac{|v_{it}|\|x_{it}\|}{\iota_N}\right)r_{\theta,N}$$
$$\leq r_{\theta,N}C\left(\frac{\mathbb{E}\|x_{it}\|^2+\mathbb{E}\|x_{it}\|\|w_{it}\|}{\min_{1\leq i\leq N}\sigma_i}+\frac{\mathbb{E}|v_{it}|\|x_{it}\|}{\iota_N}\right)$$

$$\leq r_{\theta,N} C\left(\frac{\|x_{it}\|^2 + \|x_{it}\|\|w_{it}\| - \left(\mathbb{E}\|x_{it}\|\|w_{it}\| + \mathbb{E}\|x_{it}\|\|w_{it}\|\right)}{\min\limits_{1\leq i\leq N} \sigma_i} + \frac{|v_{it}|\|x_{it}\| - \mathbb{E}|v_{it}|\|x_{it}\|}{\iota_N}\right).$$

Noting that $\mathbb{E}\|x_{it}\|^2$, $\mathbb{E}\|x_{it}\|\|w_{it}\|$ and $\mathbb{E}|v_{it}|\|x_{it}\|$ are bounded uniformly over i and t by Assumption 1.6, (A.13) implies

$$\max_{1\leq i\leq T}\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\hat{d}_{it}(h) - d_{it}(h)}{\sigma_i\|\delta_i(h)\|}\right| \leq r_{\theta,N} C\Big(\min_{1\leq i\leq N}\sigma_i \wedge \iota_N\Big)^{-1}(\sqrt{T} + \log N)$$
$$\leq r_{\theta,N} C\Big(\min_{1\leq i\leq N}\sigma_i \wedge \iota_N\Big)^{-1}\sqrt{T},$$

where the last inequality follows since $N \leq o(1)T^{\delta_2}$. The bounds on J_1 and J_2 yield the desired result. □

Lemma A.15 (Fuk–Nagaev-type inequality for mixing sequences). *Suppose that X_{it} is a strongly mixing process with zero mean for each $i = 1, \dots, N$ with tail probabilities $\sup_{i=1,\dots,N} P(|X_{it}| > x) \leq \exp(1 - (x/a)^{d_1})$ and with strong mixing coefficients $\sup_{i=1,\dots,N} a_i[t] \leq \exp(-bt^{d_2})$, where a, b, d_1, and d_2 are positive constants. Let $\mathbb{P}_N$ denote a sequence of sets of probability measures that satisfy the above conditions with given values of a, b, d_1, and d_2. Let*

$$s_T^2 = \max_{1\leq i\leq N}\max_{1\leq t\leq T}\Big(\mathbb{E}(X_{it}^2) + 2\sum_{s>t}\big|\mathbb{E}(X_{it}X_{is})\big|\Big).$$

Assume that $s_T^2 < C_s \log^{a_s} N$ for constants C_s and $0 \leq a_s \leq 1$, which do not depend on N, T, nor P. Then it holds that for any constant $C > 0$, as $N, T \to \infty$ with $NT^{-\delta_2} \to 0$ for some $\delta_2 > 0$,

$$\sup_{P\in\mathbb{P}_N} P\left(\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}X_{it}\right| \geq CT^{-1/2}\log N\right) \to 0.$$

Proof. By the Bonferroni inequality and inequality (1.7) in Merlevède, Peligrad, and Rio (2011), which is an application of Rio (2017, Theorem 6.2) (the original French version was published in 2000), we have

$$\sup_{P\in\mathbb{P}} P\left(\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T}X_{it}\right| \geq x\right) \leq \sup_{P\in\mathbb{P}}\sum_{i=1}^{N}P\left(\left|\frac{1}{T}\sum_{t=1}^{T}X_{it}\right| \geq x\right)$$
$$\leq 4N\left(1 + \frac{T(x/4)^2}{rs_T^2}\right)^{-r/2} + 4CN(x/4)^{-1}\exp\left(-a\frac{(Tx/4)^d}{b^d r^d}\right),$$

where $r \geq 1$, $d = (d_1^{-1} + d_2^{-1})^{-1}$ and C' is a positive constant. Thus, for $x = CT^{-1/2} \log N$, it holds that

$$\begin{aligned}
&\sup_{P \in \mathbb{P}} P\left(\max_{i=1,\dots,N} \left| \frac{1}{T} \sum_{t=1}^{T} X_{it} \right| \geq CT^{-1/2} \log N \right) \\
&\leq 4N\left(1 + \frac{C^2 \log^2 N}{16 r s_T^2}\right)^{-r/2} + 16(C'/C) N T^{1/2} \log^{-1} N \exp\left(-a \frac{(CT^{1/2} \log N)^d}{4^d b^d r^d}\right) \\
&= 4N \exp\left(-\frac{r}{2} \log\left(1 + \frac{C^2 \log^2 N}{16 r s_T^2}\right)\right) \\
&\quad + 16(C'/C) \frac{N T^{1/2}}{\log N} \exp\left(-a \frac{c^d}{4^d b^d} \left(\frac{T^{1/2} \log N}{r}\right)^d\right).
\end{aligned}$$

We take $r = T^{1/2-c}$ for $0 < c < 1/2$. The second term on the last line in the above display converges to zero because $T^{1/2} \log N / r = T^c \log N$ and $NT^{-\delta_2} \to 0$. We now argue that the first term vanishes as well. For a close to zero, a second-order Taylor expansion of the natural logarithm function yields

$$\log(1 + a) = a - \frac{1}{(1 + a^*)^2} a^2,$$

where a^* is an intermediate value between zero and a. For a close to zero, $1/(1 + a^*)$ is bounded and, therefore,

$$\log(1 + a) \geq a + O(a^2).$$

In particular, $\log(1 - a) \geq a + O(a^2)$. We set $a = C^2 \log^2 N / (16 T^{1/2-c} s_T^2)$. Under the assumption of the lemma, $a \to 0$. Now, the term in the exponential function can be bounded by

$$\begin{aligned}
-\frac{T^{1/2-c}}{2} \log\left(1 + \frac{C^2 \log^2 N}{16 T^{1/2-c} s_T^2}\right) &\leq -\frac{T^{1/2-c}}{2} \left\{ \frac{C^2 \log^2 N}{16 T^{1/2-c} s_T^2} + O\left(\left[\frac{C^2 \log^2 N}{16 T^{1/2-c} s_T^2}\right]^2\right)\right\} \\
&\leq -\frac{T^{1/2-c} C^2 \log^2 N}{32 T^{1/2-c} s_T^2} (1 + o(1)) \\
&\leq -\frac{C^2 \log^2 N}{64 s_T^2}.
\end{aligned}$$

Thus, under $s_T^2 \leq K \log^{a_s} N$ with $0 < a_s < 1$,

$$4N \exp\left(-\frac{r}{2} \log\left(1 + \frac{C^2 \log^2 N}{16 r s_T^2}\right)\right) \leq 4N \exp\left(-\frac{C^2 \log^2 N}{64 s_T^2}\right) \to 0. \qquad \square$$

LEMMA A.16. *Let $\mathbb{P}_N$ be the set of probability measures, which satisfy Assumptions* 1.1–1.7 *and Assumption* 1.9 *with identical choices of a, b, d_1, and d_2. Assume that there are finite* *constants* $0 < \delta_1 < \delta_2$ *such that* $T^{\delta_1} \le N \le o(1)T^{\delta_2}$ *and that*

$$\frac{r_{\theta,N}}{\iota_N \wedge \min\limits_{1\le i\le N} \sigma_i} \to 0.$$

Suppose that Assumption 3 *holds and* $\kappa_N = 0$. *Then there is a constant* C *such that*

$$\sup_{P\in\mathbb{P}_N} P\Big(\max_{1\le i\le N} \max_{(h,h^*)\in\mathbb{G}^2} \big|\widehat{D}_i(g_i^0, h) - D_i(g_i^0, h)\big| > C\frac{r_{\theta,N}\sqrt{T}}{\iota_N \wedge \min\limits_{1\le i\le N} \sigma_i}\Big) = o(1)$$

and for all $c > 0$

$$\sup_{P\in\mathbb{P}_N} P\Big(\max_{1\le i\le N} \max_{(h,h^*)\in\mathbb{G}^2} \big\|\widehat{\Omega}_i(g_i^0) - \Omega_i(g_i^0)\big\| > c\Big) = o(1).$$

Proof. Following the arguments in Lemma A.14, we bound, with probability at least $1 - a_{\theta,N}$,

$$\left|\frac{\hat{d}_{it}(h) - d_{it}(h)}{\sigma_i\|\delta_i(h)\|}\right| \le C\left(\frac{\|x_{it}\|^2 + \|x_{it}\|\|w_{it}\|}{\min\limits_{1\le i\le N} \sigma_i} + \frac{|v_{it}|\|x_{it}\|}{\iota_N}\right) r_{\theta,N},$$

and hence

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} \frac{\hat{d}_{it}(h) - d_{it}(h)}{\sigma_i\|\delta_i(h)\|} &= O_p\left(\frac{r_{\theta,N}}{\iota_N \wedge \min\limits_{1\le i\le N} \sigma_i}\right), \\ \frac{1}{T}\sum_{t=1}^{T} \frac{\big(\hat{d}_{it}(h) - d_{it}(h)\big)\big(\hat{d}_{it}(h') - d_{it}(h')\big)}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|} &= O_p\left(\frac{r_{\theta,N}^2}{\iota_N^2 \wedge \min\limits_{1\le i\le N} \sigma_i^2}\right) \end{aligned} \tag{A.14}$$

uniformly in unit i and probability measure $P \in \mathbb{P}_N$. Following similar arguments as in the proof of Lemma A.14, Lemma A.15 yields

$$\begin{aligned} &\left|\frac{1}{T}\sum_{t=1}^{T} \frac{d_{it}(h)d_{it}(h')}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|} - \frac{\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_P\big[d_{it}(h)d_{it}(h')\big]}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|}\right| \\ &\le C\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\|x_{it}\|^2 + \|x_{it}\|\|w_{it}\|}{\min\limits_{1\le i\le N} \sigma_i} + \frac{|v_{it}|\|x_{it}\|}{\iota_N}\right)^2 r_{\theta,N}^2 \\ &\le \frac{Cr_{\theta,N}^2}{\iota_N^2 \wedge \min\limits_{1\le i\le N} \sigma_i^2} + o_p(1), \end{aligned} \tag{A.15}$$

where the $o_p(1)$ term is uniform over units i and probability measures $P \in \mathbb{P}_N$. Noting that

$$\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{d_{it}(h)}{\sigma_i\|\delta_i(h)\|}\right| = \left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}v_{it}x_{it}\right\|,$$

Lemma A.15 implies

$$\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{d_{it}(h)}{\sigma_i\|\delta_i(h)\|}\right| = O_p(\log N) \tag{A.16}$$

uniformly in i and P. Since Assumption 1.6 bounds the expectation $\mathbb{E}[v_{it}^2 x_{it}x_{it}']$ by a finite constant, we have

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\frac{d_{it}(h)d_{it}(h')}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|} &= \frac{\delta_i(h)'\frac{1}{T}\sum_{t=1}^{T}v_{it}^2 x_{it}x_{it}'\delta_i(h')}{\|\delta_i(h)\|\|\delta_i(h')\|}. \\
&\le \left\|\frac{1}{T}\sum_{t=1}^{T}\left(v_{it}^2 x_{it}x_{it}' - \mathbb{E}[v_{it}^2 x_{it}x_{it}']\right)\right\| \\
&\quad + \left\|\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[v_{it}^2 x_{it}x_{it}']\right\| \\
&= O_p(\log N/\sqrt{T}) = o_p(1)
\end{aligned} \tag{A.17}$$

uniformly in i and P.

Now, combining the decomposition

$$\begin{aligned}
&\frac{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{d}_{it}(h) - \bar{\hat{d}}_{it}(h)\right)\left(\hat{d}_{it}(h') - \bar{\hat{d}}_{it}(h')\right)}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|} - \frac{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_P[d_{it}(h)d_{it}(h')]}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\hat{d}_{it}(h) - d_{it}(h)}{\sigma_i\|\delta_i(h)\|}\right)\left(\frac{\hat{d}_{it}(h') - d_{it}(h')}{\sigma_i\|\delta_i(h')\|}\right) \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\frac{d_{it}(h)}{\sigma_i\|\delta_i(h)\|}\left(\frac{\hat{d}_{it}(h') - d_{it}(h')}{\sigma_i\|\delta_i(h')\|}\right) + \frac{1}{T}\sum_{t=1}^{T}\frac{d_{it}(h')}{\sigma_i\|\delta_i(h')\|}\left(\frac{\hat{d}_{it}(h) - d_{it}(h)}{\sigma_i\|\delta_i(h)\|}\right). \\
&\quad - \left(\frac{1}{T}\sum_{t=1}^{T}\frac{\hat{d}_{it}(h)}{\sigma_i\|\delta_i(h)\|}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\hat{d}_{it}(h')}{\sigma_i\|\delta_i(h')\|}\right) \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\frac{d_{it}(h)d_{it}(h')}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|} - \frac{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_P[d_{it}(h)d_{it}(h')]}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|}
\end{aligned}$$

with (A.14), (A.15), (A.16), and (A.17) yields a constant C such that

$$\sup_{P\in\mathbb{P}} P\left(\max_{1\le i\le N}\left|\frac{\frac{1}{T}\sum_{t=1}^{T}\big(\hat{d}_{it}(h)-\bar{\hat{d}}_{it}(h)\big)\big(\hat{d}_{it}(h')-\bar{\hat{d}}_{it}(h')\big)}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|}-\frac{1}{T}\sum_{t=1}^{T}\frac{d_{it}(h)d_{it}(h')}{\sigma_i^2\|\delta_i(h)\|\|\delta_i(h')\|}\right|\right.$$
$$\left.\ge C\frac{r_{\theta,N}}{\iota_N\wedge\min_{1\le i\le N}\sigma_i}\right)=o(1).$$

This implies the first statement of the lemma. The proof of the second statement of the lemma is similar to the proof of Lemma A.14, but replacing all references to Lemma A.5 by a reference to the result in the previous display. □

Lemma A.17. *Let $\nu(N)\ge 1$ denote a sequence that converge to infinity and let $c_N(\alpha)$ denote the $(1-\alpha/N)$-quantile of the t-distribution with $\nu(N)$ degrees of freedom. Suppose that $(\log N)/\nu(N)\to 0$. For each $\epsilon>0$ and $0<\underline{\alpha}<1$, there is a threshold N_0 such that for $N\ge N_0$,*

$$\sup_{\underline{\alpha}\le\alpha<1} c_N(\alpha)\le\sqrt{2(1+\epsilon)\log(N/\alpha)}.$$

Proof. For notational convenience, write $\nu=\nu(N)$. We prove the bound for $\alpha=\underline{\alpha}$ and write $c_N=c_N(\alpha)$. The uniformity then follows from the monotonicity of the distribution function. Clearly, $c_N\to\infty$ so we can take $c_N\ge 1$, provided that N is large enough. The density function of the t-distribution with ν degrees of freedom is given by $f_\nu(x)=c(\nu)(1+x^2/\nu)^{-\frac{\nu+1}{2}}$, where

$$c(\nu)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\to\frac{1}{\sqrt{2\pi}}$$

as $\nu\to\infty$. It follows that there is a universal constant C such that $c(\nu)\le C$. We first show that $c_N^2/\nu=O(1)$. The proof is by contradiction. Suppose that $\limsup_{N\to\infty}c_N^2/\nu=\infty$. Applying Theorem 1 in Soms (1976) with $n=1$ yields

$$1-F_\nu(c_N)\le f_\nu(c_N)\frac{1}{c_N}\left(1+\frac{c_N^2}{\nu}\right). \tag{A.18}$$

This implies that

$$\frac{\alpha}{N}\le c(\nu)\left(1+\frac{c_N^2}{\nu}\right)^{-\frac{\nu+1}{2}}\left(1+\frac{c_N^2}{\nu}\right)\le C\left(1+\frac{c_N^2}{\nu}\right)^{-\frac{\nu-1}{2}}.$$

Taking logs and rearranging give

$$\frac{\log(N/\alpha)}{\nu}\ge\frac{1}{2}\frac{\nu-1}{\nu}\left(\log\left(1+\frac{c_N^2}{\nu}\right)-C\right).$$

The left-hand side of the inequality vanishes under the assumptions of the lemma, whereas a subsequence of the right-hand side diverges to infinity. This establishes that the inequality is impossible and, therefore, $c_N^2/\nu = O(1)$. This implies that there exists a constant b such that

$$1 < b \le \left(1 + \frac{c_N^2}{\nu}\right)^{\frac{\nu}{c_N^2}} \le e,$$

so that we can take

$$\left(\left(1 + \frac{c_N^2}{\nu}\right)^{\frac{\nu}{c_N^2}}\right)^{-1} \le e^{-\frac{\nu}{\nu+1}(1+\epsilon^*/2)^{-1}}$$

for a positive ϵ^*. Then

$$f_\nu(c_N) \le C\left[\left(1 + \frac{c_N^2}{\nu}\right)^{\frac{\nu}{c_N^2}}\right]^{-\frac{c_N^2}{2}[\frac{\nu+1}{\nu}]} \le C\exp\left(-\frac{c_N^2}{2}(1+\epsilon^*/2)^{-1}\right).$$

Take N large enough that

$$\frac{1}{1+\epsilon^*/2} - \frac{4\log c_N}{c_N^2} > \frac{1}{1+\epsilon^*}.$$

Then the right-hand side of (A.18) can be bounded by

$$\begin{aligned}
&C\exp\left(-\frac{c_N^2}{2}(1-\epsilon^*/2)^{-1}\right)\left(1+\frac{c_N^2}{\nu}\right)\\
&\le 2C\exp\left(-\frac{c_N^2}{2}\left((1+\epsilon^*/2)^{-1} - \frac{4\log c_N}{c_N^2}\right)\right)\\
&\le 2C\exp\left(-\frac{c_N^2}{2}(1+\epsilon^*)^{-1}\right).
\end{aligned}$$

Plugging in $1 - F_\nu(c_N) = \alpha/N$ and taking logs give

$$\begin{aligned}
c_N^2 &\le \big(1+\epsilon^*\big)\log(N/\alpha) + \log(2C)\\
&\le 2\big(1+\epsilon^*\big)\log(N/\alpha)\left(1 + \frac{1}{2\big(1+\epsilon^*\big)}\frac{\log(2C)}{\log(N/\alpha)}\right).
\end{aligned}$$

Hence, there is a constant C such that $c_N^2 \le C\log(N/\alpha)$. Using this inequality, we can now verify that $c_N^2/\nu \to 0$ so that

$$\left(1 + \frac{c_N^2}{\nu}\right)^{\frac{\nu}{c_N^2}} \to e,$$

allowing us to take $\epsilon^* = \epsilon/2$ for sufficiently large N. Taking N large enough that

$$(1+\epsilon/2)\left(1 + \frac{1}{2(1+\epsilon/2)}\frac{\log(2C)}{\log(N)}\right) \le 1+\epsilon$$

yields $c_N^2 \le 2(1+\epsilon)\log(N/\alpha)$. □

Lemma A.18. *For $\nu \geq 1$, let F_ν and f_ν denote the distribution and density function of a t-distributed random variable with ν degrees of freedom. For $x^2 > 2$,*

$$f_\nu(x) < 2x\big(1 - F_\nu(x)\big).$$

Proof. Applying Theorem 1 in Soms (1976) with $n = 2$ yields the inequality

$$1 - F_\nu(x) \geq \big(1 + x^2/\nu\big)\left(1 - \frac{\nu}{(\nu+2)x^2}\right) f_\nu(x)/x.$$

Now, $x^2 > 2$ implies

$$1 - F_\nu(x) > \left(1 - \frac{1}{2}\right) f_\nu(x)/x.$$

□

Lemma A.19. *Let $\mathbb{P}_N$ denote a family of probability measures satisfying Assumptions 1 with parameters satisfying $\|\theta_h\| < M$ for some finite M for any h. Assume $r_{\theta,N} = o(1 \wedge \iota_N)$ and $r_{\theta,N}(\sqrt{T} + \sqrt{\log N})(\iota_N + \min_{1\leq i\leq N} \sigma_i) = o(1)$. There are constants C and C' such that*

$$\sup_{P\in\mathbb{P}_N} P\left(\max_{1\leq i\leq N}\left|\frac{1}{T}\sum_{t=1}^{T} \frac{\hat{d}_{it}^U\big(g_i^0, h\big) - d_{it}^U\big(g_i^0, h\big)}{s_i^U(h)}\right| \geq C \frac{r_{\theta,N}}{\iota_N + \min\limits_{1\leq i\leq N} \sigma_i}\right) = o(1). \tag{A.19}$$

$$\sup_{P\in\mathbb{P}_N} P\left(\max_{1\leq i\leq N}\big|\widehat{D}_i^U\big(g_i^0, h\big) - \tilde{D}_i^U\big(g_i^0, h\big)\big| \geq C' \frac{r_{\theta,N}(\sqrt{T} + \sqrt{\log N})}{\iota_N + \min\limits_{1\leq i\leq N} \sigma_i}\right) = o(1). \tag{A.20}$$

Proof. Throughout the proof, let C denote a generic constant that does not depend on $P \in \mathbb{P}$. Let $\delta_i(h) = \theta_{g_i^0} - \theta_h$ and $\hat{\delta}_i(h) = \hat{\theta}_{g_i^0} - \hat{\theta}_h$. Note that

$$\|\hat{\delta}_i(h)\| / \|\delta_i(h)\| \leq 1 + \frac{\|\hat{\delta}_i(h) - \delta_i(h)\|}{\|\delta_i(h)\|}$$

is bounded by the fact that $r_{\theta,N} = o(1 \wedge \iota_N)$.

We observe

$$\begin{aligned}
&\hat{d}_{it}^U(h) - d_{it}^U(h) \\
&\quad = \frac{1}{2}\Big(\big(y_{it} - w_{it}'\hat{\theta}^w - x_{it}'\hat{\theta}_{g_i^0}\big)^2 - \big(y_{it} - w_{it}'\hat{\theta}^w - x_{it}'\hat{\theta}_h\big)^2\Big) \\
&\qquad - \frac{1}{2}\Big(\big(y_{it} - w_{it}'\theta^w - x_{it}'\theta_{g_i^0}\big)^2 - \big(y_{it} - w_{it}'\theta^w - x_{it}'\theta_h\big)^2\Big) \\
&\quad = -u_{it}x_{it}'\big(\hat{\delta}_i(h) - \delta_i(h)\big) + w_{it}'\big(\hat{\theta}^w - \theta^w\big)x_{it}'\hat{\delta}_i(h) + x_{it}'\hat{\delta}_i(h)x_{it}'(\hat{\theta}_{g_i^0} - \theta_{g_i^0}) \\
&\qquad - \frac{1}{2}\big(x_{it}'\hat{\delta}_i(h)\big)^2 + \frac{1}{2}\big(x_{it}'\delta_i(h)\big)^2.
\end{aligned}$$

Thus, we have, with probability at least $1 - a_{\theta,N}$,

$$\left|\frac{\hat{d}_{it}^U(h) - d_{it}^U(h)}{\sigma_i\|\delta_i(h)\|}\right| \leq C\left(\frac{\|x_{it}\|^2 + \|x_{it}\|\|w_{it}\|}{\min\limits_{1\leq i\leq N} \sigma_i} + \frac{|v_{it}|\|x_{it}\|}{\iota_N}\right) r_{\theta,N}, \tag{A.21}$$

$$\left(\frac{\hat{d}_{it}^U(h) - d_{it}^U(h)}{\sigma_i \|\delta_i(h)\|}\right)^2 \le C\left(\frac{\|x_{it}\|^4 + \|x_{it}\|^2\|w_{it}\|^2}{\min\limits_{1\le i\le N}\sigma_i^2} + \frac{|v_{it}|^2\|x_{it}\|^2}{\iota_N^2}\right) r_{\theta,N}^2 \tag{A.22}$$

and

$$\begin{aligned}&\left|\frac{d_{it}^U(h)}{\sigma_i\|\delta_i(h)\|}\left(\frac{\hat{d}_{it}^U(h) - d_{it}^U(h)}{\sigma_i\|\delta_i(h)\|}\right)\right| \\ &\quad \le C\left(\frac{\|x_{it}\|^4 + |v_i|\|x_{it}\|^3 + |v_i|\|x_{it}\|^2\|w_{it}\| + \|x_{it}\|^3\|w_{it}\|}{\min\limits_{1\le i\le N}\sigma_i}\right. \\ &\qquad \left. + \frac{|v_i|^2\|x_{it}\|^2 + |v_i|\|x_{it}\|^3}{\iota_N}\right) r_{\theta,N},\end{aligned} \tag{A.23}$$

where (A.23) also rely on the compactness assumption in the theorem.

Combining (A.22), Lemmas A.12, and A.15, the fact that $s_i^U(h) = \sigma_i\|\delta_i(h)\|\mathbb{E}\|x_{it}\| + \mathbb{E}(\|x_{it}\|^2)\|\delta_i(h)\|^2 \ge C\sigma_i\|\delta_i(h)\|$ yields (A.19).

Let

$$\hat{s}_i^U(h)^2 = \frac{1}{T}\sum_{t=1}^T\left(\hat{d}_{it}^U(h) - \bar{\hat{d}}_{it}^U(h)\right)^2, \qquad \tilde{s}_i^U(h)^2 = \frac{1}{T}\sum_{t=1}^T\left(d_{it}^U(h) - \bar{d}_{it}^U(h)\right)^2$$

We observe that

$$\begin{aligned}&\frac{\hat{s}_i^U(h)^2 - \tilde{s}_i^U(h)^2}{\sigma_i^2\|\delta_i(h)\|^2} \\ &= \frac{1}{T}\sum_{t=1}^T\left(\frac{\hat{d}_{it}^U(h) - d_{it}^U(h)}{\sigma_i\|\delta_i(h)\|}\right)^2 + 2\frac{1}{T}\sum_{t=1}^T\frac{d_{it}^U(h)}{\sigma_i\|\delta_i(h)\|}\frac{\left(\hat{d}_{it}^U(h) - d_{it}^U(h)\right)}{\sigma_i\|\delta_i(h)\|} \\ &\quad - T\left(\frac{1}{T}\sum_{t=1}^T\frac{\hat{d}_{it}^U(h) - d_{it}^U(h)}{\sigma_i\|\delta_i(h)\|}\right)\left(\frac{1}{T}\sum_{t=1}^T\frac{\hat{d}_{it}^U(h) - d_{it}^U(h)}{\sigma_i\|\delta_i(h)\|} + \frac{2}{T}\sum_{t=1}^T\frac{d_{it}^U(h)}{\sigma_i\|\delta_i(h)\|}\right).\end{aligned}$$

Note that

$$\left|\frac{1}{T}\sum_{t=1}^T\frac{d_{it}^U(h)}{\sigma_i\|\delta_i(h)\|}\right| \le \left\|\frac{1}{T}\sum_{t=1}^T v_{it}x_{it}\right\| + \|\delta_i(h)\|\frac{1}{T}\sum_{t=1}^T\|x_{it}\|^2$$

By Lemmas A.12 and A.15, and the compactness condition, we have

$$\sup_{P\in\mathbb{P}} P\left(\max_{1\le i\le N}\left|\frac{1}{T}\sum_{t=1}^T\frac{d_{it}^U(h)}{\sigma_i\|\delta_i(h)\|}\right| \ge C\left(1 + T^{-1/2}\sqrt{\log N}\right)\right) = o(1), \tag{A.24}$$

Combining Lemma A.15, (A.21), (A.22), (A.23), and (A.24) yield

$$\sup_{P\in\mathbb{P}}\left(\max_{1\le i\le N}\left|\frac{\hat{s}_i^U(h)^2 - \tilde{s}_i^U(h)^2}{\sigma_i^2\|\delta_i(h)\|^2}\right| \ge C\frac{r_{\theta,N}}{\iota_N \wedge \min\limits_{1\le i\le N}\sigma_i}\right) = o(1). \tag{A.25}$$

Observing that $s_i^U(h) > s_i(h)$, which holds under $\mathbb{E}(u_{it}x_{itk_1}x_{itk_2}x_{itk_3}) = 0$, $\sigma_i\|\delta_i(h)\|/s_i^U(h)$ is bounded away from infinity by Assumption 1.4. This in turn implies that $\sigma_i\|\delta_i(h)\|/\tilde{s}_i^U(h)$ is bounded away from infinity. We thus have

$$\sup_{P\in\mathbb{P}}\left(\max_{1\le i\le N}\left|\frac{\hat{s}_i^U(h)}{\tilde{s}_i^U(h)} - 1\right| \ge C\left(\frac{r_{\theta,N}}{\iota_N \wedge \min_{1\le i\le N}\sigma_i}\right)\right) = o(1). \tag{A.26}$$

Lastly, we consider

$$\begin{aligned}\widehat{D}_i^U(h) - \tilde{D}_i^U(h) &= \left(\frac{\tilde{s}_i^U(h)}{\hat{s}_i^U(h)} - 1\right)D_i(h) + \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^T\big(\hat{d}_{it}^U(h) - d_{it}^U(h)\big)/\big(\sigma_i\|\delta_i(h)\|\big)}{\hat{s}_i^U(h)/\big(\sigma_i\|\delta_i(h)\|\big)} \\ &= J_1^U + J_2^U.\end{aligned}$$

We bound J_1 by writing

$$\begin{aligned}&\left|\left(\frac{\tilde{s}_i^U(h)}{\hat{s}_i^U(h)} - 1\right)\tilde{D}_i(h)\right| \\ &= \left|\frac{\tilde{s}_i^U(h)}{\hat{s}_i^U(h)}\left(1 - \frac{\hat{s}_i^U(h)}{\tilde{s}_i^U(h)}\right)\frac{\frac{1}{T}\sum_{t=1}^T d_{it}^U(h)/\big(\sigma_i\|\delta_i(h)\|\big)}{\tilde{s}_i^U(h)/\big(\sigma_i\|\delta_i(h)\|\big)}\right| \\ &\le \frac{\tilde{s}_i^U(h)}{\hat{s}_i^U(h)}\left|1 - \frac{\hat{s}_i^U(h)}{\tilde{s}_i^U(h)}\right|\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T v_{it}x_i\right\| + \frac{1}{\sigma_i\sqrt{T}}\sum_{t=1}^T\|x_{it}\|^2\|\delta_i(h)\|\right)\left(\frac{\tilde{s}_i^U(h)}{\sigma_i\|\delta_i(h)\|}\right)^{-1}\end{aligned}$$

and applying Lemma A.15 to bound $\|T^{-1/2}\sum_{t=1}^T v_{it}x_{it}\|$, (A.26) to bound $|1-\hat{s}_i(h)/\tilde{s}_i(h)|$, the discussion above (A.26) for a lower bound on $\tilde{s}_i^U(h)/(\sigma_i\|\delta_i(h)\|)$ and

$$\frac{\hat{s}_i^U(h)}{\tilde{s}_i^U(h)} \ge 1 - \left|\frac{\hat{s}_i^U(h)}{\tilde{s}_i^U(h)} - 1\right|$$

in conjunction with (A.26) to derive a lower bound on $\hat{s}_i^U(h)/\tilde{s}_i^U(h)$. To bound J_2^U, we derive a lower bound on its denominator from

$$\frac{\hat{s}_i^U(h)}{\sigma_i\|\delta_i(h)\|} = \frac{\tilde{s}_i^U(h)}{\sigma_i\|\delta_i(h)\|}\left\{\left(\frac{\hat{s}_i^U(h)}{\tilde{s}_i^U(h)} - 1\right) + 1\right\}$$

in conjunction with (A.26) and the lower bound on $\tilde{s}_i^U(h)/(\sigma_i\|\delta_i(h)\|)$ discussed above. The numerator in J_2^U is bounded by the argument used to prove (A.19). The bounds on J_1^U and J_2^U yield (A.20). □

Appendix B: Additional results for empirical application

B.1 *One-step confidence set*

Table B.1. Marginal confidence set at level $1-\alpha=0.95$. "p-val $\hat{g}_i$" is the p-value for the significance of the estimated group membership. "card" is the cardinality of the marginal confidence set for the state. "CS" is the marginal confidence set. "baseline" refers to the procedure with critical values defined in Section 2.4. "SNS" refers to the procedure with critical values defined in Section 4.1.

		baseline			SNS		
State	$\hat{g}_i$	p-val $\hat{g}_i$	card	CS	p-val $\hat{g}_i$	card	CS
Alabama	1	0.000	1	1	0.000	1	1
Alaska	3	1.000	3	2, 3, 4	1.000	3	2, 3, 4
Arizona	3	1.000	3	2, 3, 4	1.000	3	2, 3, 4
Arkansas	2	0.027	1	2	0.041	1	2
California	3	0.000	1	3	0.000	1	3
Colorado	4	0.893	2	3, 4	1.000	2	3, 4
Connecticut	3	0.631	2	2, 3	0.715	2	2, 3
Delaware	2	1.000	2	2, 3	1.000	2	2, 3
D.C.	2	1.000	4	1, 2, 3, 4	1.000	4	1, 2, 3, 4
Florida	4	0.049	1	4	0.074	2	3, 4
Georgia	1	0.000	1	1	0.000	1	1
Hawaii	2	1.000	4	1, 2, 3, 4	1.000	4	1, 2, 3, 4
Idaho	3	0.664	2	3, 4	0.852	2	3, 4
Illinois	3	0.003	1	3	0.004	1	3
Indiana	3	0.024	1	3	0.038	1	3
Iowa	4	0.000	1	4	0.000	1	4
Kansas	4	0.010	1	4	0.015	1	4
Kentucky	3	0.093	2	3, 4	0.151	3	2, 3, 4
Louisiana	1	0.000	1	1	0.000	1	1
Maine	2	0.015	1	2	0.023	1	2
Maryland	2	0.000	1	2	0.000	1	2
Massachusetts	2	1.000	2	2, 3	1.000	2	2, 3
Michigan	2	0.001	1	2	0.001	1	2
Minnesota	2	0.000	1	2	0.000	1	2
Mississippi	1	0.001	1	1	0.001	1	1
Missouri	2	0.002	1	2	0.003	1	2
Montana	3	1.000	3	2, 3, 4	1.000	3	2, 3, 4
Nebraska	4	1.000	3	2, 3, 4	1.000	3	2, 3, 4
Nevada	2	1.000	4	1, 2, 3, 4	1.000	4	1, 2, 3, 4
New Hampshire	3	0.080	2	3, 4	0.130	2	3, 4
New Jersey	2	1.000	2	2, 3	1.000	2	2, 3
New Mexico	3	0.094	3	2, 3, 4	0.121	3	2, 3, 4
New York	2	0.000	1	2	0.000	1	2
North Carolina	2	0.000	1	2	0.000	1	2
North Dakota	4	0.010	1	4	0.015	1	4

(*Continues*)

Table B.1. *Continued.*

State	$\hat{g}_i$	baseline p-val $\hat{g}_i$	baseline card	baseline CS	SNS p-val $\hat{g}_i$	SNS card	SNS CS
Ohio	1	0.000	1	1	0.000	1	1
Oklahoma	3	0.168	2	2, 3	0.207	2	2, 3
Oregon	4	0.000	1	4	0.000	1	4
Pennsylvania	3	0.021	1	3	0.030	1	3
Rhode Island	2	0.000	1	2	0.001	1	2
South Carolina	1	0.000	1	1	0.000	1	1
South Dakota	4	1.000	3	2, 3, 4	1.000	3	2, 3, 4
Tennessee	2	0.000	1	2	0.000	1	2
Texas	1	0.000	1	1	0.000	1	1
Utah	4	1.000	2	3, 4	1.000	2	3, 4
Vermont	4	0.000	1	4	0.000	1	4
Virginia	2	1.000	2	2, 3	1.000	2	2, 3
Washington	4	0.000	1	4	0.000	1	4
West Virginia	3	0.014	1	3	0.041	1	3
Wisconsin	3	0.392	2	2, 3	0.455	2	2, 3
Wyoming	3	1.000	3	2, 3, 4	1.000	3	2, 3, 4

B.2 *Two-step confidence set assuming no serial correlation*

Under the assumption of no serial correlation, we can use the non-HAC variance estimator from Section 4.2 in the main paper and eliminate nine units in the first step.

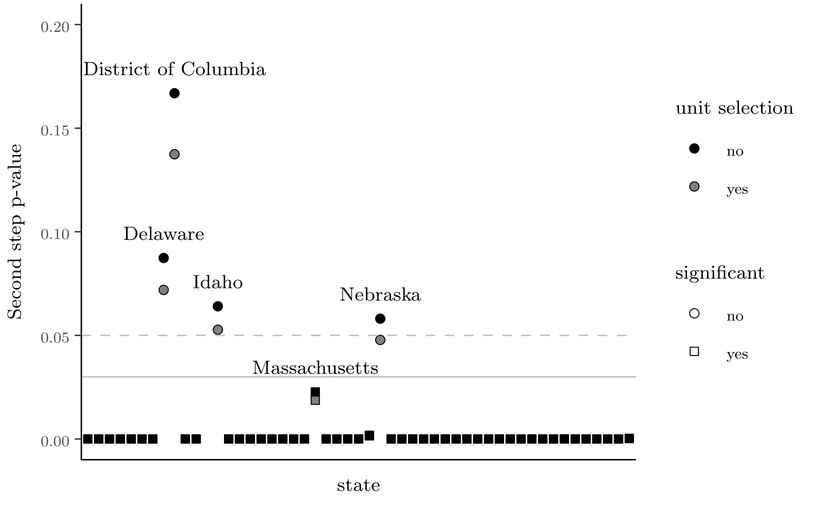

Figure B.1. Two-step procedure assuming no serial correlation. Second-step p-values for the significance of the estimated group memberships with and without unit selection ($\alpha = 0.05$, $\beta = 0.01$). The dashed horizontal line indicates the threshold for significance without unit selection. The solid horizontal line indicates the threshold for significance with unit selection.

As illustrated in Figure B.1, the elimination at the first stage decreases the second-step p-values since the Bonferroni adjustment is over a smaller number of simultaneous tests. On the other hand, turning on unit selection lowers the threshold p-value at which we can conclude significance from $\alpha = 0.05$ to $\alpha - 2\beta = 0.03$. In this example, the two-step procedure reduces p-values in the second step but does not produce a smaller confidence set.

Appendix C: Additional simulation results

C.1 *Choice of the regularization sequence* ϵ_N

C.1.1 *Benchmark simulation designs from Section 6* For our simulation results in Section 6 in the main text, we set the regularization sequence ϵ_N equal to the constant sequence $\epsilon_N = 0.01$. In this appendix, we investigate the robustness of our simulation results to this choice of regularization sequence.

The simulation design is identical to the specification simulated in Section 6. For all simulation results presented in this section, we estimate the group-specific model parameters by the *k-means* estimator and use the HAC-type estimator of the long-run variance with a data-driven bandwidth. Simulation results are based on 500 replications.

We first consider constant sequences $\epsilon_N = 0, 0.01, 0.05$. Here, $\epsilon_N = 0.01$ is the value used in the simulations in the main text, and $\epsilon_N = 0$ turns off the regularization of the variance matrix. Table C.1 reports the simulation results.

The results are not very sensitive to the choice of ϵ_N. Notably, the performance of our method is not affected substantially by turning off regularization completely (i.e., choosing $\epsilon_N = 0$). In the next section, we show that regularization plays a greater, though still limited, role in an alternative design that is tailored to make regularization relevant.

Table C.1. Simulation results for $\epsilon_N = 0, 0.01, 0.05$. Nominal level $1 - \alpha = 0.95$. "coverage" is the empirical coverage probability of the joint confidence set. "average cardinality" is the expected average (over all units) cardinality of the marginal unitwise confidence sets.

			coverage			average cardinality		
ρ	N	T	$\epsilon = 0$	$\epsilon = 0.01$	$\epsilon = 0.05$	$\epsilon = 0$	$\epsilon = 0.01$	$\epsilon = 0.05$
0.0	50	60	0.92	0.92	0.93	1.94	1.97	2.00
		120	1.00	1.00	1.00	1.14	1.16	1.18
	100	60	0.95	0.95	0.95	2.10	2.13	2.19
		120	0.99	1.00	1.00	1.19	1.21	1.26
	200	60	0.97	0.96	0.95	2.28	2.30	2.37
		120	1.00	0.98	0.99	1.27	1.28	1.34
0.5	50	60	0.69	0.69	0.66	3.39	3.42	3.49
		120	0.98	0.98	0.98	3.67	3.72	3.78
	100	60	0.67	0.69	0.73	3.60	3.62	3.66
		120	0.97	0.98	0.97	3.83	3.84	3.85
	200	60	0.68	0.70	0.68	3.73	3.74	3.76
		120	0.97	0.97	0.98	3.86	3.87	3.87

TABLE C.2. Simulation results for $\epsilon_N = \log^{-2} N, \log^{-3/2} N$. Nominal level $1-\alpha = 0.95$. "coverage" is the empirical coverage probability of the joint confidence set. "average cardinality" is the expected average (over all units) cardinality of the marginal unitwise confidence sets.

			ϵ		coverage		average cardinality	
ρ	N	T	$\epsilon = \log^{-2} N$	$\epsilon = \log^{-3/2} N$	$\epsilon = \log^{-2} N$	$\epsilon = \log^{-3/2} N$	$\epsilon = \log^{-2} N$	$\epsilon = \log^{-3/2} N$
0.0	50	60	0.07	0.13	0.92	0.93	2.03	2.04
		120	0.07	0.13	1.00	0.99	1.19	1.18
	100	60	0.05	0.10	0.96	0.96	2.19	2.25
		120	0.05	0.10	1.00	1.00	1.24	1.26
	200	60	0.04	0.08	0.95	0.97	2.37	2.44
		120	0.04	0.08	1.00	1.00	1.34	1.35
0.5	50	60	0.07	0.13	0.72	0.74	3.49	3.53
		120	0.07	0.13	0.98	0.97	3.79	3.80
	100	60	0.05	0.10	0.75	0.71	3.67	3.68
		120	0.05	0.10	0.99	0.98	3.85	3.86
	200	60	0.04	0.08	0.67	0.68	3.75	3.77
		120	0.04	0.08	0.98	0.98	3.87	3.87

We also simulate vanishing sequences $\epsilon_N = \log^{-2} N, \log^{-3/2} N$. These sequences satisfy the rate condition imposed in Theorem 2. Table C.2 reports the simulation results.

Again, we find that the simulation results are not sensitive to the choice of regularization sequence.

C.1.2 *A design where regularization matters* In our benchmark designs, the choice of regularization parameter hardly affects the performance of our procedure, raising the question of whether regularization is indeed necessary. It seems possible that regularization may be a purely technical device to facilitate the mathematical proof of the validity of our procedure, but that it may not have any practical relevance.

We address this concern by presenting an alternative simulation design where regularization affects the finite-sample performance of our procedure.

The design is very stylized and exhibits close-to-perfect correlations among the moment inequalities. For such correlations, our comparison bound relies on regularization to bound the estimation error in the critical values (see proof of Lemma A.3).

Similar to the simulation designs in Section 6, the data generating process is given by

$$\widetilde{\texttt{lemp}}_{it} = \theta_{g_i^0,1}\widetilde{\texttt{lmw}}_{it} + \theta_{g_i^0,2}\widetilde{\texttt{lpop}}_{it} + \theta_{g_i^0,3}\widetilde{\texttt{lemp}}_{it}^{\mathrm{TOT}} + \sigma_i v_{it}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. We simplify the generating process of the covariates and obtain $x_{it} = (\widetilde{\texttt{lmw}}_{it}, \widetilde{\texttt{lpop}}_{it}, \widetilde{\texttt{lemp}}_{it}^{\mathrm{TOT}})$ by sampling independently three times from the empirical distribution of $\widetilde{\texttt{lmw}}_{it}$ observed in the data for our application. This guarantees that the components of x_{it} have identical and independent marginal distributions, which makes it easier to parameterize the correlation structure of the moment inequalities with the parameter κ below.

For group $g = 1$, we set the group-specific coefficient $\theta_1 = (\theta_{1,1}, \theta_{1,2}, \theta_{1,3})$ equal to $(0.5, 0.5, 0.5)$. For the remaining groups, the coefficients are a convex combination of a design with parallel groups and a design with orthogonal groups. For $g = 2, 3, 4$, the coefficients with parallel groups are $\theta_g^{\text{parallel}} = c_g\theta_1$, with $c_2 = 0.7$, $c_3 = 0.4$, and $c_4 = 0.1$. The coefficients with orthogonal groups are $\theta_2^{\text{orthogonal}} = (0.5, 0, 0)$, $\theta_3^{\text{orthogonal}} = (0, 0.5, 0)$, $\theta_4^{\text{orthogonal}} = (0, 0, 0.5)$. For $g = 2, 3, 4$, the group-specific coefficients are given by the convex combination $\theta_g = (1-\kappa)\theta_g^{\text{parallel}} + \kappa\theta_g^{\text{orthogonal}}$, where $\kappa = 0, 0.05, 0.1, 0.2$. For $\kappa = 0$, groups are parallel, and all off-diagonal entries of the population correlation matrix $\Omega_i(g)$ are perfect correlations. For $\kappa = 1$, groups are orthogonal, and the matrix $\Omega_i(g)$ is diagonal.

As in the designs in Section 6, each unit i is assigned to one of the four groups with equal probability and exhibits a random heteroscedasticity parameter $\sigma_i = 0.1 \times \chi^2(4)/4$, where $\chi^2(df)$ is a random draw from a χ^2-distribution with df degrees of freedom.

To establish a conjecture about the role of regularization in this design, we briefly review where regularization enters our theoretical arguments. Regularization is part of our strategy to control estimation errors in the critical values. Estimation error is a concern because group-specific critical values are estimated from data. It is a greater concern in small panels (T and N small) than in large panels (T or N large). Consider a positive off-diagonal entry in $\Omega_i(g_i^0)$. From the proof of Lemma A.3, it is apparent that estimation error is easily controlled if the entry is bounded away from unity. We apply an argument that relies on our regularization scheme to control estimation error if the entry is close to unity. In summary, regularization is expected to be relevant if T and/or N are small and κ is small.

This conjecture is confirmed by the simulation results in Table C.3 For $\kappa = 0, 0.1$, regularization improves size control in the designs with small samples. In particular, we see improvements if $N = 50$. For $\kappa = 0.2$, regularization leads to slightly worse size control. We interpret this as a sign that for $\kappa = 0.2$, the cost of regularization in terms of a biased variance estimator is not outweighed by the benefit of guarding against underestimating close-to-perfect positive correlations.

Simulation designs that investigate the role of regularization are by necessity designs with substantial sampling error in the the group-specific coefficients. Without sampling error, there is no uncertainty about the critical values and regularization is not needed. The overall noisiness that makes the designs presented here informative about regularization also affects the performance of our procedure directly, leading to a confidence set that is underpowered independently of imprecisely estimated critical values. This can be seen by comparing the performance of the regularized procedure with MVT critical values to the procedure with data-independent SNS critical values. The coverage under SNS critical values provides an upper bound on the coverage that can be achieved by eliminating estimation error in the critical values, that is, an upper bound on what better regularization can achieve. This has to be considered when interpreting the improvements in size control from regularization. For example, for $\kappa = 0.1$, $N = 50$, and

TABLE C.3. Simulation results for a stylized design with strongly correlated moment inequalities. Nominal level $1-\alpha=0.95$. "coverage" is the empirical coverage probability of the joint confidence set. "average cardinality" is the expected average (over all units) cardinality of the marginal unitwise confidence sets. MVT = use MVT critical values. SNS = use SNS critical values.

			coverage				average cardinality			
			MVT			SNS	MVT			SNS
κ	N	T	$\epsilon=0$	$\epsilon=0.01$	$\epsilon=0.05$		$\epsilon=0$	$\epsilon=0.01$	$\epsilon=0.05$	
0.0	50	60	0.82	0.80	0.84	0.86	2.19	2.18	2.22	2.30
		120	0.91	0.94	0.94	0.95	1.66	1.65	1.69	1.74
	100	60	0.92	0.91	0.93	0.95	2.32	2.32	2.36	2.47
		120	0.93	0.96	0.95	0.96	1.72	1.73	1.77	1.83
0.1	50	60	0.69	0.74	0.73	0.75	2.32	2.32	2.35	2.42
		120	0.86	0.87	0.89	0.91	1.80	1.79	1.80	1.85
	100	60	0.84	0.82	0.85	0.87	2.46	2.46	2.50	2.58
		120	0.93	0.92	0.92	0.95	1.87	1.87	1.89	1.95
0.2	50	60	0.65	0.62	0.66	0.68	2.42	2.42	2.43	2.50
		120	0.87	0.86	0.85	0.88	1.88	1.90	1.89	1.93
	100	60	0.78	0.78	0.76	0.81	2.59	2.60	2.59	2.67
		120	0.91	0.90	0.91	0.94	1.99	1.98	1.99	2.03

$T=60$, regularization improves the size by about five percentage points, bringing the size within a percentage point of the size under SNS critical values.

The simulation results offer some evidence that the theoretical considerations that motivate our regularization approach have practical relevance. This suggests that it may not be possible to rigorously justify a version of our procedure that does not use regularization. On the other hand, even in this highly stylized design, gains from regularization are limited. From a practical perspective, correct regularization may not be a key concern.

C.2 *Testing the estimated group membership* $\hat{g}_i$

In the definition of $\widehat{C}_{\alpha,N,i}$, we explicitly add $\hat{g}_i$ to the confidence set. Not doing this changes the marginal confidence set of unit i only if $\hat{g}_i$ is not already included anyway, that is, if

$$\widehat{T}_i(\hat{g}_i) > \hat{c}_{\alpha,N,i}(\hat{g}_i). \tag{C.1}$$

We simulate the finite sample probability of this happening in our simulation designs from Section 6 in the main text. The simulation results are summarized in Table C.4. The column "$\hat{g}_i$ not rej" gives the simulated probability of our group membership not rejecting the estimated group membership (i.e., one minus the probability of the event defined in equation (C.1)). We find that our test for group membership does not reject the estimated group membership with probability close to, but not equal to, one.

TABLE C.4. Simulated probability of the event (C.1).

ρ	σ	N	T	coverage	cardinality	$\hat{g}_i$ not rej
0.0	0.1	50	60	0.99	1.60	0.99
			120	1.00	1.07	0.99
		100	60	0.99	1.76	1.00
			120	1.00	1.12	1.00
		200	60	0.99	1.95	1.00
			120	1.00	1.17	1.00
	0.2	50	60	0.92	1.97	0.99
			120	1.00	1.16	0.99
		100	60	0.95	2.13	0.99
			120	1.00	1.21	1.00
		200	60	0.96	2.30	1.00
			120	0.98	1.28	1.00
0.5	0.1	50	60	0.87	3.35	0.99
			120	0.99	3.71	1.00
		100	60	0.86	3.52	1.00
			120	0.99	3.80	1.00
		200	60	0.82	3.67	1.00
			120	1.00	3.84	1.00
	0.2	50	60	0.69	3.42	0.99
			120	0.98	3.72	1.00
		100	60	0.69	3.62	0.99
			120	0.98	3.84	1.00
		200	60	0.70	3.74	1.00
			120	0.97	3.87	1.00

C.3 *Two-step procedure*

In this appendix, we report simulation results regarding the finite-sample performance of our two-step procedure.

We simulate a design with independent time periods. Like our main design in Section 6 in the main text, the design studied here builds on the model estimated in Section 5 in the main text. A unit i corresponds to a US state and the "time periods" are given by observations of different counties in different quarters. The panel model is specified as in equation (12) in the main text, with coefficients equal to the estimated coefficients in Table 1 in the main text. The joint distribution of the regressors $\widetilde{\texttt{lmw}}_{it}$, $\widetilde{\texttt{lpop}}_{it}$, and $\widetilde{\texttt{lemp}}_{it}^{\text{TOT}}$ is defined from the data used in our empirical application. In particular, $\widetilde{\texttt{lmw}}_{it}$, $\widetilde{\texttt{lpop}}_{it}$, and $\widetilde{\texttt{lemp}}_{it}^{\text{TOT}}$ are sampled from the pooled empirical distribution of the respective fixed-effect transformations of $\log(\texttt{mw}_{ict})$, $\log(\texttt{pop}_{ict})$, and $\log(\texttt{emp}_{ict}^{\text{TOT}})$. The error component v_{it} is a standard normal noise term.

We set the distribution of (σ_i, T_i, g_i^0), that is, the distribution of heteroscedasticity and group membership, such that the simulation results reveal different aspects of the performance of the two-step procedure. We note that the two-step procedure is sensitive to this distribution. We determine it from the data by mapping each simulated unit i to

TABLE C.5. Simulation results for the two-step procedures (unit selection).

		success		failure		card with sel		card without sel			
m_σ	α/β	insignif	signif	insignif	signif	insignif	signif	insignif	signif	$\hat{N}$	coverage
						MVT					
0.25	10	0.55	0	0.00	0.00	1.48	1.00	2.01	1	10.06	1.00
	5	0.48	0	0.00	0.00	1.55	1.00	2.01	1	9.29	1.00
1.00	10	0.13	0	0.00	0.00	2.15	1.00	2.21	1	33.73	0.99
	5	0.00	0	0.01	0.03	2.22	1.00	2.21	1	32.35	1.00
4.00	10	0.00	0	0.50	0.45	2.75	1.02	2.72	1	50.95	0.96
	5	0.00	0	0.80	0.77	2.79	1.05	2.72	1	50.90	0.97
						SNS					
0.25	10	0.53	0	0.00	0.00	1.51	1.00	2.00	1	10.04	1.00
	5	0.46	0	0.00	0.00	1.58	1.00	2.00	1	9.32	1.00
1.00	10	0.13	0	0.00	0.00	2.21	1.00	2.27	1	33.73	1.00
	5	0.00	0	0.01	0.04	2.27	1.00	2.27	1	32.42	1.00
4.00	10	0.00	0	0.52	0.44	2.78	1.02	2.75	1	50.95	0.98
	5	0.00	0	0.81	0.73	2.82	1.05	2.75	1	50.92	0.98

one of the $N = 51$ units from our empirical application. We set σ_i equal to m_σ times the standard deviation of the empirical residuals for unit i, g_i^0 equal to the estimated group membership of i, and T_i equal to the number of observed "time periods" for unit i (i.e., number of counties times number of quarters). The parameter $m_\sigma = 1/4, 1, 4$ shifts the global level of uncertainty.

The other parameters for the simulations are set as follows. The nominal level of the simulated joint confidence set is $1 - \alpha = 0.95$. We simulate different values of the first-step parameter $\beta = \alpha/5, \alpha/10 = 0.01, 0.005$. The regularization sequence is specified as $\epsilon_N = 0.01$. We simulate the confidence set using our benchmark critical values defined in Section 2.4 of the main text (labeled MVT = multivariate t-distribution), as well as the SNS critical values defined in Section 4.1 of the main text (labeled SNS).

The simulation results are based on 1000 replications and reported in Table C.5. The columns labeled "insignif" give averages over units that are insignificant under no unit selection. Columns labeled "signif" give averages over units that are significant under no unit-selection. A unit is labeled as a "success" ("failure") if its marginal confidence set is strictly smaller (strictly larger) under unit-selection than under no unit selection. The columns labeled "card with sel" ("card without sel") give the cardinality of unitwise marginal confidence sets if unit selection is turned on (turned off). The column labeled $\hat{N}$ gives the simulated expected number of units that survive unit selection ($N = 51$). "coverage" gives the simulated joint coverage probability of the two-step joint confidence set (nominal level $1 - \alpha = 0.95$).

In all designs, unit selection produces a valid joint confidence set that covers the true group structure at the prescribed nominal level.

Unit selection aims to tighten the marginal confidence sets for units for which estimated group memberships are insignificant under a one-step procedure. Among such units, the expected proportion of units for which a two-step procedure tightens the marginal confidence set varies across the different designs. In the design with low error variances ($m_\sigma = 0.25$), this proportion ranges between 46% and 55%. This means that the two-step procedure improves the marginal confidence sets for roughly half of the units for which they can be improved. In the design with medium error variances ($m_\sigma = 1$), this proportion lies between 0% and 13%. In the design with high error variances ($m_\sigma = 4$), there are no improvements. This illustrates that the two-step procedures can only be successful if the overall uncertainty is low but unequally distributed across units. If overall uncertainty is high, then the first step cannot deselect units, and hence the second-step confidence sets cannot be tightened.

The two-step procedures can cause the confidence set to become wider if insufficiently many units are eliminated in the first step. This happens in the designs with high error variance ($m_\sigma = 4$): hardly any units are eliminated in the first step and the size of the marginal confidence sets increases both for units with significant and units with insignificant group membership estimates under the one-step procedure.

Using MVT instead of SNS critical values increases the power of our two-step procedure. In our designs, both choices of critical values select a similar number of units for the second step. Therefore, the power gain from using MVT critical values is almost entirely due to more efficient testing in the second step.

Appendix D: Weak group separation

D.1 *Introduction*

In this appendix, we consider grouped panel models in which groups are only weakly separated. By weak separation, we mean that groups are distinct but very similar to each other. We formalize this notion using an approach inspired by the local alternatives in asymptotic testing theory. In particular, we let the distance between groups shrink to zero at a fixed rate.

We offer new theoretical results on the rate of consistency of the *k-means* estimator under weak separation. In particular, we give conditions under which the estimated group-specific coefficients converge at the parametric $\sqrt{NT}$-rate if the distance between groups shrinks at a rate slower than $T^{-1/2}$. We then use this result to derive conditions under which our confidence set is valid under weak group separation.

In addition to the theoretical analysis, we provide simulation studies to investigate the finite sample behavior of the *k-means* estimator and our joint confidence set under weak separation and to verify our theoretical predictions.

This appendix is structured as follows. In Section D.2, we discuss existing results on *k-means* estimation in a setting where groups are not separated at all. We then turn to our analysis of the *k-means* estimation under weak separation. In Section D.3, we present asymptotic results. In Section D.4, we present simulation evidence. Proofs are given in Section D.5.

D.2 *No group separation*

We first discuss *k-means* estimation under no group separation. By "no group separation," we mean that there are at least two groups with identical coefficients. This corresponds to overspecification of the number of groups. Bonhomme and Manresa (2015) study this setting in their Supplemental Appendix. In this setting, the estimators of the group-specific coefficient converge at most at the rate of $T^{-1/2}$. As discussed in the main text of this paper, this rate is too slow to satisfy our conditions for the validity of the joint confidence set.

We consider a simple mean shift model where we observe y_{it}, for $i = 1, \dots, N$, and $t = 1, \dots T$. The parameter of interest is the mean of y_{it}. We assume that there is a latent group structure with G groups and that the mean of y_{it} may depend on unit i's group membership. Suppose that there is only one distinct group, that is, all units have the same mean, but we incorrectly set the number of groups to two. Specifically, the estimated model is

$$y_{it} = \alpha_{g_i} + v_{it},$$

where $g_i = 1, 2$ and v_{it} is assumed to be i.i.d. $N(0, \sigma^2)$. Let $\hat{\alpha}_1$ and $\hat{\alpha}_2$ be the estimators of α_1 and α_2, respectively, by the *k-means* method. By relabeling if necessary, we impose $\hat{\alpha}_1 \geq \hat{\alpha}_2$. The true model is homogeneous such that $\alpha = \alpha_1 = \alpha_2$.

Proposition S.2 of Bonhomme and Manresa (2015) (Supplemental Appendix) states that, as $N \to \infty$ with T fixed, it holds that in probability,

$$\hat{\alpha}_1 \to \alpha + \sqrt{\frac{2}{\pi T}}, \qquad \hat{\alpha}_2 \to \alpha - \sqrt{\frac{2}{\pi T}}.$$

We note that the model considered in Proposition S.2 of Bonhomme and Manresa (2015) includes regressors with common coefficients, but their presence does not affect the probability limits of $\hat{\alpha}_1$ and $\hat{\alpha}_2$.

The above result indicates that, even if we take $T \to \infty$ in addition to $N \to \infty$, the convergence rates of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are at most of order $T^{-1/2}$. In particular, the probability that $P(|\hat{\alpha}_g - \alpha| > CT^{-1/2})$ for fixed C does not converge to zero.

D.3 *Asymptotic analysis*

We now turn to the setting of weak group separation, where groups are distinct but very similar. The discussion given here is a simplified version of Lumsdaine, Okui, and Wang (2023, Supplemental Appendix C).

We observe (y_{it}, x_{it}) for $i = 1, \dots, N$ and $t = 1, \dots, T$. Units are divided into G groups, and all members of a group share the same value of the regression coefficient. The model is

$$y_{it} = x_{it}' \theta^0_{g_i^0} + u_{it},$$

where θ^0_g, $g = 1, \dots, G$, are group-specific coefficients, $g_i^0 \in \{1, \dots, G\}$ is unit i's true group membership, and u_{it} is an error term.

The parameters are estimated by the *k-means* method (Bonhomme and Manresa (2015)). Let $\mathbb{G} = \{1, \ldots, G\}$ be the set of groups. Then $\mathbb{G}^N$ is the parameter space for the group membership structure. A typical element of $\mathbb{G}$ is $\gamma = (g_1, \ldots, g_N)$. The true group membership structure is $\gamma^0 = (g_1^0, \ldots, g_N^0) \in \mathbb{G}$. The parameter space for the coefficients is denoted as $\mathcal{B} \subset \mathbb{R}^{Gp}$. The estimator is

$$(\hat{\gamma}, \hat{\theta}) = \arg \min_{\gamma \in \mathbb{G}^N, \theta \in \mathcal{B}} = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(y_{it} - x_{it}' \theta_{g_i}\right)^2.$$

We prove a rate of consistency of the *k-means* estimator under the following assumptions that are weaker than the standard set of assumptions imposed in the literature (see, e.g., Bonhomme and Manresa (2015)). Most importantly, the group separation condition is relaxed, allowing the difference between the slope coefficients associated with two groups to vanish asymptotically. In addition, we relax the conditions on the existence of moments and the mixing properties of the data.

Assumption D.1. 1. *Let z_{it} be $x_{it}' x_{it}$, or $\|u_{it} x_{it}\|$. Assume the following holds for any choice of z_{it}: z_{it} is a strictly stationary and strong mixing sequence over t whose mixing coefficients $a_i[t]$ are bounded by $a[t]$ such that $\max_{1 \le i \le N} a_i[t] \le a[t]$ and $\sum_{t=0}^{\infty} (t+1)^{r/2-1} a[t]^{b/r+b} < \infty$ for some $b > 0$, and $\max_{1 \le i \le N} E(|z_{it}|^{r+b}) < \infty$ for some $b > 0$.*

2. *$\mathcal{B}$ is compact.*

3. *Let $\rho_N(\gamma, g, \tilde{g})$ be the minimum eigenvalue of*

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}\{g_i^0 = g\} \mathbf{1}\{g_i = \tilde{g}\} x_{it} x_{it}' / (NT),$$

where $\gamma = (g_1, \ldots, g_N)$. For any $g \in \mathbb{G}$,

$$\min_{\gamma \in (\mathbb{G})^N} \max_{\tilde{g} \in \mathbb{G}} \rho_N(\gamma, g, \tilde{g}) > \hat{\rho},$$

where $\hat{\rho} \to_p \rho$ as $N, T \to \infty$ and $\rho > 0$.

4. *There exists $\hat{\rho}^*$ such that for any i,*

$$\lambda_{\min}\left(\frac{1}{T} \sum_{t=1}^{T} x_{it} x_{it}'\right) \ge \hat{\rho}^*$$

and $\hat{\rho}^* \to_p \rho^* > 0$ as $N, T \to \infty$, where $\lambda_{\min}$ gives the minimum eigenvalue of its argument.

5. *There exists a nonrandom sequence $c_T > T^{-1/2+e}$ for some $e > 0$ such that for any $g \neq \tilde{h}$ where $g, h \in \mathbb{G}$, it holds that $\|\theta_g^0 - \theta_h^0\| > c_T$.*

Assumption D.1.5 is the key assumption that replaces the standard group separation assumption by weak group separation. Similar to the standard assumption, any pair of groups must have distinct coefficients. In particular, the distance between their coefficients has to be bounded away from zero in any finite sample. We generalize the standard assumption and allow the distance to vanish asymptotically. In the limit, groups are not separated. We assume that the rate at which group differences vanish is slower than $T^{-1/2}$.

The mixing and moment conditions in Assumption D.1.1 are weaker than the standard assumptions imposed in the literature (see, e.g., Bonhomme and Manresa (2015)). However, we impose the additional assumption of strict stationarity. Under this assumption, we can relate the degree of weak group separation to a condition on the relative magnitudes of N and T.

All other assumptions are standard in the literature.

The following theorem derives an asymptotic equivalence between the *k-means* estimator $\hat{\theta}$ and the oracle estimator $\tilde{\theta}$ under a known group membership structure. The oracle estimator is trivially $\sqrt{NT}$-consistent.

Unlike most existing results on the consistency of the *k-means* estimator in grouped panels, the theorem holds under weak group separation. However, the degree of group separation affects the required condition on the relative magnitudes of N and T. The faster group separation converges to the limit of no group separation, the stronger the conditions on N and T. In particular, when group separation is weak, T must be large relative to N.

THEOREM D.1. *Suppose that Assumptions D.1.1–D.1.5 hold. As $N, T \to \infty$ with $NT^{-er} \to 0$, where e and r are defined in Assumptions D.1.1 and D.1.5, respectively, it holds that*

$$\hat{\theta} = \tilde{\theta} + o_p(1/\sqrt{NT}).$$

Since the oracle estimator $\tilde{\theta}$ is $\sqrt{NT}$-consistent, Theorem D.1 implies that $\hat{\theta}$ is $\sqrt{NT}$-consistent.

Under the conditions of Theorem D.1, group consistency still holds (see Lemma D.4). This is restrictive since the relevance of our testing problem relies on uncertainty about the group memberships even in the asymptotic limit. We leave a formal analysis that extends our results to settings with asymptotic misclassification for future research.

Our results indicate that such an extension is feasible. In our previous work in Dzemski and Okui (2021), we have shown that consistent estimation of group memberships is not a necessary condition for $\sqrt{NT}$-estimation of the group-specific coefficients under weak separation. We proved this result for the mean-shift model estimated from i.i.d. data. Theorem D.1 extends some of our previous analysis to a grouped panel regression model with weakly dependent time series. To simplify the derivations and make our main point (robustness to weak separation) in a clear and transparent manner, we impose a uniform bound on the variance of the error term (see our Assumption D.1.1). This bound implies group consistency. We conjecture that the uniform variance bound can be replaced by a set of more convoluted conditions (see condition (4) in Dzemski and Okui (2021)) that do not imply group consistency. We leave the details of this argument to future research.

Putting $\iota_N = c_T$, Theorem 2 in the main text yields

$$T^{1-e} r_{\theta,N} \to 0 \tag{D.1}$$

as $N, T \to \infty$ as a necessary condition (ignoring a log term) for the validity of our confidence set. Here, $r_{\theta,N}$ is the rate of convergence of $(\hat{\theta}_1, \ldots, \hat{\theta}_G)$. It indicates that the validity of our confidence set holds even when groups are only weakly separated as long as the cross-sectional sample is sufficiently large so that the group-specific coefficients converge sufficiently fast.

D.4 *Simulations*

We now report simulation evidence to study the finite sample effect of weak group separation on the rate of convergence of the *k-means* estimator and the validity of our joint confidence set for group membership.

The simulation design is a simplified version of the design in Section 6 in the main text. The data generating process is given by

$$\widetilde{\texttt{lemp}}_{it} = \theta_{g_i^0,1} \widetilde{\texttt{lmw}}_{it} + \theta_{g_i^0,2} \widetilde{\texttt{lpop}}_{it} + \theta_{g_i^0,3} \widetilde{\texttt{lemp}}_{it}^{\mathrm{TOT}} + \sigma_i v_{it}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, where g_i^0 is the group membership of unit i and takes either the value one or two with equal probability. For $g = 1, 2$, the group specific-coefficient is equal to

$$\theta_g = \left(1 - 2T^{-1/2+e}\right) \frac{\bar{\theta}_1 + \bar{\theta}_2}{2} + 2T^{-1/2+e} \bar{\theta}_g,$$

where $(\bar{\theta}_1, \bar{\theta}_2)$ are estimated by fitting the model to the data from the empirical application using the *k-means* algorithm and setting the number of groups to $G = 2$. The data generating process for the covariates and error is the same as in Section 6 in the main text, setting $\rho = 0$ and $\sigma = 0.2$. We simulate designs with $N = 50, 100, 200$, $T = 30, 60, 120$, and $e = -0.25, 0, 0.25$. The parameter e controls the rate at which the two group-specific coefficients converge to a common value as $T \to \infty$. Under $e = -0.25$, the two groups converge to a common group the fastest, and under $e = 0.25$, they converge the slowest. The case $e = 0.25$ is covered by the theoretical result in Theorem D.1. The case $e = 0$ is the infimum of the e considered in Theorem D.1. Under $e = -0.25$, group separation vanishes at a rate too fast to be covered by Theorem D.1. Table D.1 reports group separation between the two groups for the different choices of e.

To measure the distance between two sets of group-specific slope coefficients $\theta = (\theta_1, \theta_2)$ and $\theta' = (\theta'_1, \theta'_2)$, we define

$$\|\theta - \theta'\| = \sqrt{\sum_{g=1}^{2} \mathbb{E}\|\theta_g - \theta'_g\|_2^2}$$

TABLE D.1. Group separation and e.

T	$\|\theta_1^0 - \theta_2^0\|$			$\|\theta^0\|$		
	e = 0.25	e = 0.00	e = −0.25	e = 0.25	e = 0.00	e = −0.25
30	0.50	0.22	0.09	1.01	0.99	0.98
60	0.43	0.16	0.06	1.00	0.98	0.98
120	0.36	0.11	0.03	1.00	0.98	0.98

and

$$\|\theta\| = \sqrt{\sum_{g=1}^{2} \mathbb{E}\|\theta_g\|_2^2}$$

and $\|\cdot\|_2$ is the L_2-norm. Table D.1 shows that, for $\theta^0 = (\theta_1^0, \theta_2^0)$, $\|\theta^0\|$ is almost independent of e.

We simulate the joint confidence for group membership using the variance estimator for the case of no serial correlation (i.e., setting the bandwidth equal to zero). For all simulations, the nominal level for the joint confidence set is set to $1 - \alpha = 0.95$, and the number of replications is 500. The simulation results are summarized in Table D.2.

The column "coverage" gives the simulated coverage probability of the joint confidence set for group membership. The coverage is always conservative for the designs with slowly vanishing group separation ($e = 0.25$). For the designs with group separation that vanishes at a moderate or fast rate ($e = 0$ and $e = -0.25$, respectively), the confidence set has appropriate or conservative coverage provided N and T are large enough.

The columns labeled "$\hat{\theta} - \theta$" simulate the expected total error of *k-means* estimation $\|\hat{\theta} - \theta^0\|/\|\theta^0\|$, where $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ and the norm $\|\cdot\|$ is defined above. This error is relevant for assessing the finite sample validity of the assumptions we impose on coefficient estimation in Theorem 2 in the main text. When scaled by $T^{1/2}$, the error is approximately constant when increasing T and leaving N constant. This indicates that this is the rate at which time-series variation reveals information about the panel model. A smaller error can be achieved by using both time-series and cross-sectional variation, that is, by increasing both T and N. As discussed in Section D.3, a necessary condition for the asymptotic validity of our confidence set is

$$T^{1-e}\left\|\hat{\theta} - \theta^0\right\| \to 0$$

as $N, T \to \infty$. Clearly, this condition cannot be met by increasing T alone while keeping N constant. However, the estimation error scaled by T^{1-e} vanishes if N increases, suggesting that estimation error from *k-means* estimation is negligible in panels with a large cross-sectional dimension.

The columns labeled "$\hat{\theta} - \tilde{\theta}$" simulate $\|\tilde{\theta} - \hat{\theta}\|/\|\theta^0\|$. They are the normalized expected differences between the *k-means* estimator $\hat{\theta}$ and the oracle estimator $\tilde{\theta}$. Based on these columns, we assess the finite-sample relevance of our asymptotic result on *k-means* estimation under weak separation. Only the designs with $e = 0.25$ are covered by

TABLE D.2. Estimation error and confidence set coverage under shrinking group separation.

				Error			Error scaled by $T^{1/2}$			Error scaled by T^{1-e}		
e	N	T	coverage	$\hat{\theta}-\theta$	$\hat{\theta}-\tilde{\theta}$	$\tilde{\theta}-\theta$	$\hat{\theta}-\theta$	$\hat{\theta}-\tilde{\theta}$	$\tilde{\theta}-\theta$	$\hat{\theta}-\theta$	$\hat{\theta}-\tilde{\theta}$	$\tilde{\theta}-\theta$
0.25	50	30	0.97	0.28	0.12	0.25	1.52	0.66	1.37	3.55	1.54	3.22
		60	0.99	0.20	0.05	0.19	1.52	0.35	1.48	4.23	0.99	4.11
		120	1.00	0.14	0.00	0.14	1.56	0.04	1.56	5.16	0.15	5.16
	100	30	0.97	0.20	0.10	0.17	1.09	0.53	0.95	2.54	1.23	2.23
		60	0.99	0.14	0.04	0.14	1.10	0.30	1.06	3.07	0.85	2.96
		120	1.00	0.10	0.01	0.10	1.09	0.06	1.09	3.61	0.21	3.61
	200	30	0.98	0.15	0.08	0.12	0.80	0.43	0.68	1.87	1.00	1.59
		60	0.99	0.10	0.03	0.10	0.76	0.23	0.76	2.12	0.64	2.11
		120	1.00	0.07	0.00	0.07	0.79	0.05	0.78	2.61	0.16	2.58
0.00	50	30	0.84	0.39	0.24	0.26	2.16	1.33	1.41	2.16	1.33	1.41
		60	0.94	0.23	0.12	0.19	1.81	0.90	1.51	1.81	0.90	1.51
		120	0.99	0.15	0.03	0.14	1.61	0.33	1.58	1.61	0.33	1.58
	100	30	0.90	0.28	0.19	0.18	1.55	1.04	0.98	1.55	1.04	0.98
		60	0.96	0.17	0.09	0.14	1.35	0.71	1.09	1.35	0.71	1.09
		120	0.98	0.10	0.03	0.10	1.15	0.29	1.11	1.15	0.29	1.11
	200	30	0.91	0.23	0.17	0.13	1.26	0.94	0.70	1.26	0.94	0.70
		60	0.97	0.12	0.07	0.10	0.96	0.55	0.77	0.96	0.55	0.77
		120	0.99	0.07	0.02	0.07	0.82	0.21	0.79	0.82	0.21	0.79
−0.25	50	30	0.42	0.76	0.60	0.26	4.17	3.28	1.42	1.78	1.40	0.61
		60	0.58	0.42	0.30	0.19	3.23	2.31	1.51	1.16	0.83	0.54
		120	0.89	0.18	0.09	0.14	1.95	0.97	1.58	0.59	0.29	0.48
	100	30	0.46	0.55	0.45	0.18	3.02	2.44	0.98	1.29	1.04	0.42
		60	0.65	0.31	0.23	0.14	2.40	1.79	1.09	0.86	0.64	0.39
		120	0.93	0.13	0.07	0.10	1.40	0.77	1.11	0.42	0.23	0.33
	200	30	0.52	0.48	0.41	0.13	2.60	2.23	0.70	1.11	0.95	0.30
		60	0.68	0.24	0.20	0.10	1.88	1.52	0.77	0.67	0.55	0.28
		120	0.95	0.09	0.05	0.07	1.04	0.58	0.79	0.31	0.18	0.24

Theorem D.1. As predicted by the theorem, the difference between the *k-means* estimator and the oracle estimation vanishes at a faster rate than $T^{-1/2}$. We find the same result for $e=0, -0.25$, two cases not covered by our asymptotic results.

The columns labeled "$\tilde{\theta}-\theta$" report the simulated value of $\|\hat{\theta}-\theta^0\|/\|\theta^0\|$, that is, the expected error of the oracle estimator.

The columns labeled "$\hat{\theta}-\theta$" are equivalent to differently scaled versions of the convergence rate of $\hat{\theta}$, that is, of $r_{\theta,N}$ defined in Assumption 1.3. The column scaled by T^{1-e} checks the validity of condition (D.1) in finite samples. This condition is necessary for the validity of our joint confidence set under weak separation. For $e=0.25$, Theorem D.1 predicts that $T^{1-e}r_{\theta,N}\to\infty$ (i.e., condition (D.1) above) if N is sufficiently large compared to T. The simulation evidence confirms this prediction. The settings with $e=0, -0.25$ are not covered by the asymptotic analysis in Theorem D.1. Our simulation evidence suggests that the conditions for $T^{1-e} r_{\theta,N} \to 0$ are possibly even weaker under these settings.

D.5 *Proof of Theorem* D.1

Let

$$Q(\gamma, \theta) = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(y_{it} - x'_{it}\theta_{g_i}\right)^2.$$

Note that $Q(\gamma, \theta)$ is the objective function for the estimation. We also define

$$\tilde{Q}(\gamma, \theta) = \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(x'_{it}\left(\theta^0_{g_i^0} - \theta_{g_i}\right)\right)^2 + \frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} u_{it}^2.$$

The theorem follows from the following sequence of lemmas.

LEMMA D.1. *Suppose that Assumptions* D.1.1 *and* D.1.2 *hold. Then*

$$\sup_{\gamma \in \mathbb{G}^N, \theta \in \mathcal{B}} \left|\tilde{Q}(\gamma, \theta) - Q(\gamma, \theta)\right| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

PROOF. The proof is almost identical to the proof of Lemma S.3 of Bonhomme and Manresa (2015). We have

$$\tilde{Q}(\gamma, \theta) - Q(\gamma, \theta) = -2\frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} x'_{it}\left(\theta^0_{g_i^0} - \theta_{g_i}\right)u_{it}.$$

We rewrite a part of the right-hand side as

$$\frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} x'_{it}\theta^0_{g_i^0}u_{it} = \frac{1}{NT} \sum_{g \in \mathbb{G}^B} \sum_{t=1}^{T} \sum_{i=1}^{N} \mathbf{1}\left(g_i(B) = g\right)x'_{it}\theta^0_{g_i^0}u_{it}.$$

For each $g \in \mathbb{G}^B$, it holds that

$$\mathbb{E}\left(\frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} \mathbf{1}\left(g_i(B) = g\right)x'_{it}\theta^0_{g_i^0}u_{it}\right)^2 \leq C\mathbb{E}\left\|\frac{1}{NT} \sum_{t=1}^{T} \sum_{g_i(B)=g} x_{it}u_{it}\right\|^2 = O\left(\frac{1}{NT}\right),$$

where the inequality is the Cauchy–Schwarz inequality with C satisfying $\|\theta_{g_i}\|^2 < C$ for any $\theta \in \mathcal{B}$ (by Assumption D.1.2), and the equality follows since Theorem 1 of Yokoyama (1980) implies that under Assumption D.1.1, for any $L \subseteq \{1, \ldots, N\}$, there exists M, which does not depend on L such that

$$\mathbb{E}\left(\left\|\frac{1}{NT} \sum_{t=1}^{T} \sum_{i \in L} x_{it}u_{it}\right\|^2\right) \leq M\frac{|L|}{N^2 T}, \tag{D.2}$$

where $|L|$ denotes the cardinality of L. We then examine the other part of $\tilde{Q}(\gamma, \theta) - Q(\gamma, \theta)$. It follows that

$$\begin{aligned}\left(\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}x_{it}'\theta_{g_i}u_{it}\right)^2 &\leq \left(\frac{1}{NT}\sum_{i=1}^{N}\theta_{g_i}\sum_{t=1}^{T}x_{it}u_{it}\right)^2 \\ &\leq \left(\frac{1}{N}\sum_{i=1}^{N}\|\theta_{g_i}\|^2\right)\left(\frac{1}{NT^2}\sum_{i=1}^{N}\left\|\sum_{t=1}^{T}x_{it}u_{it}\right\|^2\right) \\ &= O_p\left(\frac{1}{T}\right),\end{aligned}$$

where the first inequality follows by the Cauchy–Schwarz inequality and the second inequality follows by Assumption D.1.2 and the Markov inequality by (D.2). We thus have

$$\tilde{Q}(\gamma, \theta) - Q(\gamma, \theta) = O\left(\frac{1}{\sqrt{NT}}\right) + O\left(\frac{1}{\sqrt{T}}\right)$$

uniformly over θ and γ, and consequently,

$$\sup_{\gamma\in\mathbb{G}, \theta\in\mathcal{B}}\left|\tilde{Q}(\gamma, \theta) - Q(\gamma, \theta)\right| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

□

LEMMA D.2. *Suppose that Assumptions* 1–3 *hold. Then*

$$\max_{g\in\mathbb{G}}\min_{\tilde{g}\in\mathbb{G}}\left\|\theta_g^0 - \hat{\theta}_{\tilde{g}}\right\|^2 = O_p(1/\sqrt{T}).$$

PROOF. The proof is almost identical to the proofs of Lemmas A.2 and B.3 of Bonhomme and Manresa (2015). Lemma D.1 implies

$$\begin{aligned}\tilde{Q}(\hat{\gamma}, \hat{\theta}) &= Q(\hat{\gamma}, \hat{\theta}) + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &\leq Q\left(\gamma^0, \theta^0\right) + O_p\left(\frac{1}{\sqrt{T}}\right) = \tilde{Q}\left(\gamma^0, \theta^0\right) + O_p\left(\frac{1}{\sqrt{T}}\right).\end{aligned}$$

The fact that $\tilde{Q}(\gamma, \theta)$ is minimized at (γ^0, θ^0) implies

$$\tilde{Q}(\hat{\gamma}, \hat{\theta}) - \tilde{Q}\left(\gamma^0, \theta^0\right) = O_p\left(\frac{1}{\sqrt{T}}\right).$$

We now establish a lower bound of $\tilde{Q}(\gamma, \theta) - \tilde{Q}(\gamma^0, \theta^0)$ such that

$$\begin{aligned}&\tilde{Q}(\gamma, \theta) - \tilde{Q}\left(\gamma^0, \theta^0\right) \\ &\quad = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(x_{it}'\left(\theta_{g_i^0}^0 - \theta_{g_i}\right)\right)^2\end{aligned}$$

$$= \frac{1}{NT}\sum_{t=1}^{T}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\sum_{i=1}^{N}\mathbf{1}\{g_i^0 = g\}\{g_i = \tilde{g}\}\big(x_{it}'(\theta_g^0 - \theta_{\tilde{g}})\big)^2$$

$$\geq \frac{1}{T}\sum_{t=1}^{T}\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\rho_N(\gamma, g, \tilde{g})\left\|\theta_g^0 - \theta_{\tilde{g}}\right\|^2$$

$$\geq \hat{\rho}G^2 \max_{g\in\mathbb{G}}\min_{\tilde{g}\in\mathbb{G}^B}\left\|\theta_g^0 - \theta_{\tilde{g}}\right\|^2,$$

where the first inequality follows by the definition of $\rho_N(\gamma, g, \tilde{g})$ and the second inequality is from the definition of $\hat{\rho}$. Now, Assumption D.1.4 implies

$$\max_{g\in\mathbb{G}}\min_{\tilde{g}\in\mathbb{G}}\left\|\theta_g^0 - \hat{\theta}_{\tilde{g}}\right\|^2 = O_p\left(\frac{1}{\sqrt{T}}\right). \qquad \square$$

Lemma D.3. *Suppose that Assumptions D.1.1–D.1.3, and D.1.5 are satisfied. Then there exist a permutation* $\sigma : \mathbb{G} \mapsto \mathbb{G}$ *such that* $\|\theta_g^0 - \hat{\theta}_{\sigma(g)}\|^2 = O_p(1/\sqrt{T})$ *for any* $g \in \mathbb{G}$.

Proof. We construct a permutation with the property stated in the lemma. Indeed, we show that

$$\sigma(g) = \arg\min_{\tilde{g}\in\mathbb{G}}\left\|\theta_g^0 - \hat{\theta}_{\tilde{g}}\right\|^2$$

is such a permutation. We first show that σ satisfies $\|\theta_g^0 - \hat{\theta}_{\sigma(g)}\|^2 = O_p(1/\sqrt{T})$ for any $g \in \mathbb{G}$, and that it is a permutation.

Lemma D.2 states that

$$\max_{g\in\mathbb{G}}\min_{\tilde{g}\in\mathbb{G}}\left\|\theta_g^0 - \hat{\theta}_{\tilde{g}}\right\|^2 = O_p(1/\sqrt{T}).$$

The map σ, by construction, satisfies $\|\theta_g^0 - \hat{\theta}_{\sigma(g)}\|^2 = O_p(1/\sqrt{T})$ for any $g \in \mathbb{G}$.

It remains to establish that σ is a permutation. For $g \neq \tilde{g}$, the triangular inequality gives

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

In the above, we have seen that $\|\theta_g^0 - \hat{\theta}_{\sigma(g)}\| = O_p(1/\sqrt{T})$ and $\|\theta_{\tilde{g}}^0 - \hat{\theta}_{\sigma(\tilde{g})}\| = O_p(1/\sqrt{T})$. Assumption 5 states that $\|\theta_g^0 - \theta_{\tilde{g}}^0\| > c_T$. The condition that $c_T > T^{-1/2+e}$ implies that $\|\theta_g^0 - \theta_{\tilde{g}}^0\| - \|\theta_g^0 - \hat{\theta}_{\sigma(g)}\| - \|\theta_{\tilde{g}}^0 - \hat{\theta}_{\sigma(\tilde{g})}\| > 0$ with probability approaching one. This means that $\sigma(g) \neq \sigma(\tilde{g})$ for $g \neq \tilde{g}$ with probability approaching one. Thus, σ possesses a well-defined inverse and is bijective; in other words, σ is a permutation. $\square$

From Lemmas D.2 and D.3, we observe that the Hausdorff distance between θ^0 and $\hat{\theta}$ converges to 0 at the rate of $\sqrt{T}$. By using the labeling such that $\sigma(g) = g$, we can write $\|\theta_g^0 - \hat{\theta}_g\|^2 = O_p(1/\sqrt{T})$ for any $g \in \mathbb{G}$.

We then establish that the group membership structure is correct asymptotically as long as the coefficients are in a neighborhood of the true value. Let $\mathcal{N} = \{\theta : \|\theta_g^0 - \theta_g\| < \eta = T^{-1/2+f}, \forall g \in \mathbb{G}\}$ for $0 < f < e$, where e is defined in Assumption D.1.5, for any $g \in \mathbb{G}$.

LEMMA D.4. *Suppose that Assumptions* D.1.1, D.1.2, D.1.4, *and* D.1.5 *hold. As* $N, T \to \infty$ *with* $NT^{-er} \to 0$, *where* e *and* r *are defined in Assumptions* D.1.1 *and* D.1.5, *respectively, it holds that*

$$P\{\hat{\gamma}(\theta) \neq \gamma^0 \text{ for some } \theta \in \mathcal{N}\} \to 0.$$

PROOF. We establish an equivalent statement:

$$\max_{1 \le i \le N} \sup_{\theta \in \mathcal{N}} \mathbf{1}\{\hat{g}_i(\theta) \neq g_i^0\} = o_p(1).$$

Note that

$$\mathbf{1}\{\hat{g}_i(\theta) \neq g_i^0\} = \max_{g \in \mathbb{G} \setminus \{g_i^0\}} \mathbf{1}\left(\sum_{t=1}^T (y_{it} - x_{it}'\theta_g)^2 < \sum_{t=1}^T (y_{it} - x_{it}'\theta_{g_i^0})^2\right).$$

We have

$$\sum_{t=1}^T \left((y_{it} - x_{it}'\theta_g)^2 - (y_{it} - x_{it}'\theta_{g_i^0})^2\right) = \sum_{t=1}^T 2u_{it}x_{it}(\theta_{g_i^0}^0 - \theta_g^0) + \sum_{t=1}^T \left(x_{it}'(\theta_{g_i^0}^0 - \theta_g^0)\right)^2 + \Psi,$$

where

$$\Psi = \sum_{t=1}^T 2u_{it}x_{it}(\theta_{g_i^0} - \theta_g - \theta_{g_i^0}^0 + \theta_g^0) + \sum_{t=1}^T (\theta_{g_i^0} - \theta_g - \theta_{g_i^0}^0 + \theta_g^0)' x_{it}x_{it}'(2\theta_{g_i^0}^0 - \theta_{g_i^0} - \theta_g)$$
$$+ \sum_{t=1}^T (\theta_{g_i^0}^0 - \theta_g^0)' x_{it}x_{it}'(\theta_{g_i^0}^0 - \theta_{g_i^0} - \theta_g + \theta_g^0).$$

Applying the Cauchy–Schwarz inequality and then Assumption D.1.2 and the definition of $\mathcal{N}$ gives

$$|\Psi| \le \eta C_1 \left\| \sum_{t=1}^T u_{it}x_{it} \right\| + \eta C_2 \left\| \sum_{t=1}^T x_{it}x_{it}' \right\|,$$

where C_1 and C_2 are constants independent of η and T. We thus have the following inequality:

$$\mathbf{1}\left(\sum_{t=1}^T (y_{it} - x_{it}'\theta_g)^2 < \sum_{t=1}^{k-1} (y_{it} - x_{it}'\theta_{g_i^0})^2\right)$$
$$\le \mathbf{1}\Bigg(\sum_{t=1}^{k^0-1} 2u_{it}x_{it}'(\theta_{g_i^0}^0 - \theta_g^0)$$
$$- \sum_{t=1}^T \left(x_{it}'(\theta_{g_i^0}^0 - \theta_{g,B}^0)\right)^2 + \eta C_1 \left\| \sum_{t=1}^T u_{it}x_{it} \right\| + \eta C_2 \left\| \sum_{t=1}^T x_{it}x_{it}' \right\|\Bigg).$$

Thus, we have

$$
\begin{aligned}
&P\Big(\sup_{\theta\in\mathcal{N}} \mathbf{1}\big(\hat{g}_i(\theta)\neq g_i^0\big)\neq 0\Big)\\
&\quad\le \sum_{g\in\mathbb{G}\setminus\{g_i^0\}}\left(P\left(\frac{1}{T}\sum_{t=1}^{T}\big(x_{it}'\big(\theta^0_{g_i^0}-\theta^0_g\big)\big)^2\le\frac{c_T''}{2}\right)+P\left(\left\|\frac{1}{T}\sum_{t=1}^{T}u_{it}x_{it}\right\|\ge M\right)\right.\\
&\qquad+P\left(\left\|\frac{1}{T}\sum_{t=1}^{T}x_{it}x_{it}'\right\|\ge M\right)\\
&\qquad\left.+P\left(\frac{1}{T}\sum_{t=1}^{T}2u_{it}x_{it}'\big(\theta^0_{g_i^0}-\theta^0_g\big)<-\frac{c_T''}{2}+\eta C_1M+\eta C_2M\right)\right),
\end{aligned}
$$

where we take $c_T''=c_T\times\rho^*$ for c in Assumption D.1.5 and ρ^* in Assumption D.1.4 and M is some large constant.

We now bound the second and third terms on the right-hand side of the inequality. First, we have

$$
P\left(\left\|\frac{1}{T}\sum_{t=1}^{T}x_{it}x_{it}'\right\|\ge M\right)\le P\left(\frac{1}{T}\sum_{t=1}^{T}\|x_{it}x_{it}'\|\ge M\right)=P\left(\frac{1}{T}\sum_{t=1}^{T}x_{it}'x_{it}\ge M\right).
$$

Assumption D.1.1 enables us to the Markov inequality and Theorem 1 of Yokoyama (1980) to $x_{it}'x_{it}-\mathbb{E}(x_{it}'x_{it})$, and we establish

$$
P\left(\left\|\frac{1}{T}\sum_{t=1}^{T}x_{it}x_{it}'\right\|\ge M\right)=O\big(T^{-r/2}\big),
$$

by taking M large enough such that $\sum_{t=1}^{T}\mathbb{E}(x_{it}'x_{it})/T<M$. A similar argument under Assumption D.1.1 implies that $P(\|(T)^{-1}\sum_{t=1}^{T}u_{it}x_{it}\|\ge M)=O(T^{-r/2})$.

Next, we consider the first term. We now use Assumptions D.1.4, D.1.5, and D.1.1. The Markov inequality combined with Theorem 1 of Yokoyama (1980) implies

$$
P\left(\left|\frac{1}{T}\sum_{t=1}^{T}\big(x_{it}'\big(\theta^0_{g_i^0}-\theta^0_g\big)\big)^2-\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\big(\big(x_{it}'\big(\theta^0_{g_i^0}-\theta^0_g\big)\big)^2\big)\right|\ge\frac{c_T''}{2}\right)=O\big(T^{-er}\big).
$$

We thus have uniformly over g:

$$
P\left(\frac{1}{T}\sum_{t=1}^{T}\big(x_{it}'\big(\theta^0_{g_i^0}-\theta^0_g\big)\big)^2\ge\frac{c_T''}{2}\right)=O\big(T^{-er}\big).
$$

Lastly, we consider the fourth term. It follows that

$$P\left(\frac{1}{T}\sum_{t=1}^{T}2u_{it}x_{it}'\left(\theta_{g_i^0}^0-\theta_g^0\right)<-\frac{c_T''}{2}+\eta C_1 M+\eta C_2 M\right)$$
$$\leq P\left(\frac{1}{T}\sum_{t=1}^{T}2u_{it}x_{it}'\left(\theta_{g_i^0}^0-\theta_g^0\right)<-\frac{c_T''}{4}\right)=O\left(T^{-er}\right)$$

uniformly over g under Assumption D.1.1. The inequality follows by $c_T''=O(c_T)=O(T^{-1/2+e})$ and $\eta=o(T^{-1/2+e})$. The equality holds by the Markov inequality and Theorem 1 of Yokoyama (1980).

To sum up, we have

$$P\Big(\max_{1\leq i\leq N}\sup_{\theta\in\mathcal{N}}\mathbf{1}\big(\hat{g}_i(\theta)\neq g_i^0\big)\neq 0\Big)\leq\sum_{i=1}^{N}P\Big(\sup_{\theta\in\mathcal{N}}\mathbf{1}\big(\hat{g}_i(\theta)\neq g_i^0\big)\neq 0\Big)$$
$$=O\big(N\big(T^{-er}+T^{-r/2}\big)\big)=O\big(NT^{-er}\big).\qquad\square$$

References

Bonhomme, Stéphane and Elena Manresa (2015), "Grouped patterns of heterogeneity in panel data." *Econometrica*, 83 (3), 1147–1184. [66, 67, 68, 72, 73]

Chang, Jinyuan, Xiaohui Chen, and Mingcong Wu (2024), "Central limit theorems for high dimensional dependent data." *Bernoulli*, 30 (1), 712–742. [42, 45]

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato (2017), "Central limit theorems and bootstrap in high dimensions." *The Annals of Probability*, 45 (4), 2309–2352. [5]

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato (2019), "Inference on causal and structural parameters using many moment inequalities." *Review of Economic Studies*, 86, 1867–1900. [43]

Davydov, Yu A. (1968), "Convergence of distributions generated by stationary stochastic processes." *Theory of Probability and Its Applications*, 13, 691–696. [44]

Davidson, James (1994), *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press, Oxford, UK. [43]

Duembgen, Lutz (2010), "Bounding standard Gaussian tail probabilities." Preprint. arXiv:1012.2063. [19, 43]

Dzemski, Andreas and Ryo Okui (2021), "Convergence rate of estimators of clustered panel models with misclassification." *Economics Letters*, 203, 109844. [68, 69]

Galvao, Antonio F. and Kengo Kato (2014), "Estimation and inference for linear panel data models under misspecification when both n and T are large." *Journal of Business & Economic Statistics*, 32 (2), 285–309. [44]

Kosorok, Michael R. (2008), *Introduction to Empirical Processes and Semiparametric Inference*. Springer. [42, 44, 45]

Li, Wenbo V. and Qi-Man Shao (2002), "A normal comparison inequality and its applications." *Probability Theory and Related Fields*, 122 (4), 494–508. [19]

Lumsdaine, Robin L., Ryo Okui, and Wendun Wang (2023), "Estimation of panel group structure models with structural breaks in group memberships and coefficients." *Journal of Econometrics*, 233 (1), 45–65. [66]

Merlevède, Florence, Magda Peligrad, and Emmanuel Rio (2011), "A Bernstein type inequality and moderate deviations for weakly dependent sequences." *Probability Theory and Related Fields*, 151, 435–474. [48]

Rio, Emmanuel (2017), *Asymptotic Theory of Weakly Dependent Random Processes*. Springer, Berlin, Heidelberg, Germany. [48]

Soms, Andrew P. (1976), "An asymptotic expansion for the tail area of the t-distribution." *Journal of the American Statistical Association*, 71 (355), 728–730. [52, 54]

Yokoyama, Ryozo (1980), "Moment bounds for stationary mixing sequences." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 52 (1), 45–57. [72, 76, 77]

Co-editor Stéphane Bonhomme handled this manuscript.

Manuscript received 31 August, 2021; final version accepted 9 January, 2024; available online 18 January, 2024.