Supplementary Material

Supplement to "Inference on heterogeneous treatment effects in high-dimensional dynamic panels under weak dependence"

(*Quantitative Economics*, Vol. 14, No. 2, May 2023, 471–510)

Vira Semenova
Department of Economics, University of California, Berkeley

Matt Goldman
Meta

Victor Chernozhukov
Department of Economics, MIT

Matt Taddy
Amazon

Appendix A presents and results on independence couplings. Appendix B develops concentration results for weakly-dependent panel data. Appendix C presents the results for high-dimensional CLT for weakly dependent data. Appendix D contains proofs for Section 4, and Appendix E for Section 5. Appendix F contains tail bounds for empirical rectangular matrices in operator norm. Appendix G contains the analysis of OLS used in stage 3 of our inference procedure.

Keywords. Orthogonal learning, residual learning, CATE, dynamic panel data, time series, mixing, cross-fitting, neighbors-left-out.

JEL classification. C14, C23, C33.

Notation We use the standard notation for numeric and stochastic dominance. For two numeric sequences, $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$, $a_n \lesssim b_n$ stands for $a_n = O(b_n)$. For two sequences of random variables $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$, $a_n \lesssim_P b_n$ stands for $a_n \lesssim_P (b_n)$. For a random vector V, let $V^0 := V - \mathrm{E}[V]$ be the demeaned vector. Let $[N] := \{1, 2, \dots, N\}$, $[T] := \{1, 2, \dots, T\}$, and $[j] := \{1, 2, \dots, d\}$. Let $a \wedge b = \min\{a, b\}$ and $a \vee b = \max\{a, b\}$.

Matrix and vector norms For a vector $v \in \mathrm{R}^d$, denote the ℓ_2-norm of v as $\|v\|_2 := \sqrt{\sum_{j=1}^d v_j^2}$, the ℓ_1-norm of v as $\|v\|_1 := \sum_{j=1}^d |v_j|$, the ℓ_∞-norm of v as $\|v\|_\infty := \max_{1\leq j\leq d} |v_j|$, and the ℓ_0-"norm" of v as $\|v\|_0 := \sum_{j=1}^d 1_{\{v_j\neq 0\}}$. Denote a unit sphere in

Vira Semenova: semenovavira@gmail.com
Matt Goldman: mattgoldman5850@gmail.com
Victor Chernozhukov: vchern@mit.edu
Matt Taddy: mataddy@gmail.com
We thank Michael Jansson, Pat Kline, Sylvia Klosin, Demian Pouzo and two anonymous referees whose comments helped improve the paper. We thank Microsoft Research Redmond and Simons Institute for the Theory of Computing for a productive research environment.

$\mathbb{R}^d$ as $\mathcal{S}^{d-1} = \{\alpha \in \mathbb{R}^d : \|\alpha\| = 1\}$. For a matrix $A = (a_{ij}) \in \mathbb{R}^{d\times d}$, let its operator norm be $\|A\|_2 = \sup_{\alpha\in\mathcal{S}^{d-1}} \|A\alpha\|_2$, the elementwise norm be $\|A\|_\infty = \max_{1\le i,j\le d} |a_{ij}|$, and the maximal ℓ_1-row-norm:

$$\|A\|_{1,\infty} = \max_{1\le j\le d} \sum_{i=1}^{d} |a_{ij}|.$$

Empirical process notation In what follows, we use the standard empirical process notation. For a generic measurable function $f : \mathcal{W} \to \mathbb{R}$ and a generic sample $\{\{W_{it}\}_{t=1}^T\}_{i=1}^N$, where W_{it}'s take values in $\mathcal{W}$, define the empirical expectation

$$\mathbb{E}_{NT} f(W_{it}) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} f(W_{it})$$

and the empirical process:

$$\mathbb{G}_{NT} f(W_{it}) = \sqrt{NT}\mathbb{E}_{NT}\big[f(W_{it}) - \mathrm{E}_{W_{it}} f(W_{it})\big].$$

Appendix A: Tools: Strassen and Berbee couplings. Implications for cross-fitting

A.1 *Strassen's coupling: Weak and strong form via Dudley–Philipp*

Let S be a Polish space and $P_{Z,W}$ be a law on $S \times S$, with marginal laws P_Z on S and P_W on S. Let $(\Omega, \mathcal{B}, \mathrm{P})$ be a probability space and Z be a random variable on Ω with values in S and law $\mathcal{L}(Z) = P_Z$. Assume that $(\Omega, \mathcal{B}, \mathrm{P})$ has been extended to carry a random variable U on Ω, independent of Z, with values in $[0, 1]$ and law $U(0, 1)$. The total variation norm of a signed measure ν on the Polish space T is defined as

$$\|\nu\|_{\mathrm{TV}} = \sup_{F \text{ closed}} \nu(F).$$

The total variation distance between laws P and Q defined on the Polish space T is defined by taking $\nu = P - Q$ in the definition above.

We also make use of the following Strassen's weak coupling result (e.g., Villani (2007, p. 7)):

$$\min_{Z^*,W^*} \big\{\mathrm{P}\big(Z^* \ne W^*\big) : \mathcal{L}\big(Z^*\big) = P_Z, \mathcal{L}\big(W^*\big) = P_W\big\} = \frac{1}{2}\|P_Z - P_W\|_{\mathrm{TV}}, \tag{A.1}$$

where minimization is done over space of random variables Z^* and W^* defined on the probability space $(\Omega, \mathcal{B}, \mathrm{P})$. Note that the problem above is the optimal transportation problem for 0-1 cost; see Villani (2007) for discussion. The above is a special case of Strassen's original result; Schwarz (1980) (Theorem 1) provides another proof of (A.1).

We now recall the following result.

Lemma A.1 (Strong Coupling; Lemma 2.11, Dudley and Philipp (1983)). *Let S and T be Polish spaces and Q a law on $S \times T$, with marginal P_Z on S. Let $(\Omega, \mathcal{B}, P)$ be a probability space and Z be a random variable on Ω with values in S and law $\mathcal{L}(Z) = P_Z$. Assume that there is a random variable U on Ω, independent of Z, with values in a separable metric space R and law $\mathcal{L}(U)$ on R having no atoms. Then there exists a random variable W on Ω with values in T and law $\mathcal{L}((Z, W)) = Q$.*

This result is quoted with minor adaptation of notation. This lemma implies the strong form of Strassen's weak coupling (A.1) as stated in the following lemma.

Lemma A.2 (Strong Form of Strassen's Coupling). *Given the setup above with a given random variable Z, there exists a random variable W taking values in S, defined on the same probability space, and having law $\mathcal{L}(W) = P_W$ such that*

$$\mathrm{P}(Z \neq W) = \frac{1}{2}\|P_Z - P_W\|_{\mathrm{TV}}. \tag{A.2}$$

Proof. Strassen's weak coupling implies that there is a pair of random variables (Z^*, W^*) with law Q and marginals P_Z and P_W such that

$$\mathrm{P}\big(Z^* \neq W^*\big) = \frac{1}{2}\|P_Z - P_W\|_{\mathrm{TV}}.$$

Application of the Dudley–Philipp lemma with $S = T$ and U taken to be uniform random variable implies that for the given Z there is a pairing random variable W, such that law of (Z, W) is Q. Therefore,

$$\mathrm{P}(Z \neq W) = \mathrm{P}\big(Z^* \neq W^*\big) = \frac{1}{2}\|P_Z - P_W\|_{\mathrm{TV}}. \qquad \square$$

A.2 *Independence coupling*

Consider a special case of the setup above with $S = S_1 \times S_2$ and $T = S$, where S_1 and S_2 are Polish spaces, and where $Z = (X, Y)$ is a pair of random variables such that $\mathcal{L}(X) = P_X$ on S_1 and $\mathcal{L}(Y) = P_Y$ on S_2, and $\mathcal{L}(X, Y) = P_{X,Y}$.

Lemma A.3 (Strong Coupling With Independence via Strassen–Dudley–Philip). *Consider the setup above. We can construct $\widetilde{Y}$ and $\widetilde{X}$ that are independent of each other with laws $\mathcal{L}(X) = P_X$ and $\mathcal{L}(Y) = P_Y$ such that*

$$\mathrm{P}\big\{(X, Y) \neq (\tilde{X}, \tilde{Y})\big\} = \frac{1}{2}\|P_{X,Y} - P_X \times P_Y\|_{\mathrm{TV}}.$$

Proof. In the previous lemma, take $Z = (X, Y)$ and $W = (\tilde{X}, \tilde{Y})$, and note that $P_W = P_X \times P_Y$. $\square$

A.3 *Berbee coupling extended*

Let (X, Y) be a pair of random variables taking values in the Polish space $S_1 \times S_2$ as in the setup above. Define their coefficient of dependence as

$$\gamma(X, Y) = \frac{1}{2}\|P_{X,Y} - P_X \times P_Y\|_{\mathrm{TV}}.$$

This coefficient vanishes if and only if X and Y are independent.

The following lemma is a minor extension of Lemma 2.1 of Berbee from real-valued to Polish-space valued random variables.

Lemma A.4 (Berbee Coupling on Polish Spaces). *Let $X = (X_i)_{i=1}^n$ be a collection of random variables taking values on the Polish space $S = (S_1 \times \cdots \times S_n)$, and defined on the same probability space $(\Omega, \mathcal{B}, \mathrm{P})$. Define for $1 \leq i < n$,*

$$\gamma^{(i)} = \gamma\big(X_i, (X_{i+1}, \ldots, X_n)\big).$$

The probability space can be extended so that there exist a collection of random variables $\tilde{X} = (\tilde{X}_i)_{i=1}^n$ that are mutually independent, such that each $\tilde{X}_i$ has the same law as X_i and

$$\mathrm{P}(X \neq \tilde{X}) \leq \gamma^{(1)} + \cdots + \gamma^{(n-1)}.$$

Proof. Assume that $(\Omega, \mathcal{B}, \mathrm{P})$ has been extended to carry a random variable U on Ω, independent of X, with values in $[0, 1]$ and law $U(0, 1)$.

1. Application of strong form of Strassen's coupling in Lemma A.2 implies that one can construct $\tilde{X}$ as in the statement of the lemma such that

$$\mathrm{P}(X \neq \tilde{X}) = \frac{1}{2}\|P_X - P_{\tilde{X}}\|.$$

2. (Identical to Berbee). To prove the claim of the lemma, we have to estimate the right hand side. If X, Y, and $\tilde{Y}$ are random variables, with Y and $\tilde{Y}$ having values in the same space, then

$$\begin{aligned}\|P_{X,Y} - P_{X,\tilde{Y}}\| &\leq \|P_{X,Y} - P_X \times P_Y\|_{\mathrm{TV}} + \|P_X \times P_Y - P_X \times P_{\tilde{Y}}\|_{\mathrm{TV}} \\ &= 2\gamma(X, Y) + \|P_Y - P_{\tilde{Y}}\|_{\mathrm{TV}}.\end{aligned}$$

Applying this rule successively, one obtains

$$\begin{aligned}&\frac{1}{2}\|P_{X_1,\ldots,X_n} - P_{X_1} \times \cdots \times P_{X_n}\| \\ &\quad\leq \gamma^{(1)} + \|P_{X_2,\ldots,X_n} - P_{X_2} \times \cdots \times P_{X_n}\| \\ &\quad\leq \cdots \leq \gamma^{(1)} + \cdots + \gamma^{(n-1)}.\end{aligned}$$

□

Corollary A.1 (Berbee's Coupling for Panel Data). *Let $\{X_{i1}, X_{i2}, \ldots, X_{iL}\}_{i=1}^N$ be real random matrices of possibly distinct dimensions. Suppose the sequences $(X_{i1}, X_{i2}, \ldots,$*

X_{iL}) *are independent over* i. *For each* i, $(X_{i1}, X_{i2}, \ldots, X_{iL})$ *is* β*-mixing whose coefficients are bounded as*

$$\sup_{1\le i\le N}\sup_{1\le l\le L-1}\gamma\big((W_{i,1},\ldots,W_{i,l-1}),(W_{i,l},\ldots,W_{i,L})\big)\le\epsilon. \tag{A.3}$$

The probability space can be extended with random variables X_{il}^* *distributed as* X_{il} *such that* X_{il}^* *are independent over* i, l, *and*

$$\mathrm{P}\big(X_{il}\neq X_{il}^*\ \textit{for some}\ i,l\big)\le N(L-1)\epsilon. \tag{A.4}$$

This follows immediately from the union bound and Lemma A.4.

A.4 *Applications to cross-fitting*

Here, we recall the setup induced by the NLO construction given in the main text. Let $\mathcal{M}_k$ and $\mathcal{M}_k^{\mathrm{qc}}$ be two NLO subsets of time indices $\{1,2,\ldots,T\}$, for $k=1,\ldots,K$. Define the data blocks

$$\begin{aligned} B_k &= \bigcup_{i=1}^N B_{ik}, \qquad B_{ik}=\{W_{it}\}_{t\in\mathcal{M}_k};\\ B_k^{\mathrm{qc}} &= \bigcup_{i=1}^N B_{ik}^{\mathrm{qc}}, \qquad B_{ik}^{\mathrm{qc}}=\{W_{it}\}_{t\in\mathcal{M}_k^{\mathrm{qc}}}. \end{aligned} \tag{A.5}$$

By construction, the time periods in $\mathcal{M}_k$ and $\mathcal{M}_k^{\mathrm{qc}}$ are separated by at least $T_k\ge T_{\mathrm{block}}:=\lfloor T/(K-1)\rfloor$ time periods.

Lemma A.5 (Approximate Independence of Separated Chunks). *Suppose Assumption* 4.1 *holds with*

$$\gamma(q):=\sup_{\bar t\le T,i\le N}\gamma\big(\{W_{it}\}_{t\le\bar t},\{W_{it}\}_{t\ge\bar t+q}\big)\le C_\kappa\exp(-\kappa q) \tag{A.6}$$

and $\log N/T_{\mathrm{block}}=o(1)$. *Then there exist random elements* B_k^* *and* B_k^{qc*} *such that* (1) B_k^* *and* B_k *are equal in law,* B_k^{qc*} *and* B_k^{qc} *are equal in law,* (2) B_k^* *and* B_k^{qc*} *are independent, and* (3) *the event*

$$\mathcal{E}_{\mathrm{berbee}}:=\big\{(B_k,B_k^{\mathrm{qc}})=B_k^*,B_k^{\mathrm{qc}*}),\ \textit{for all}\ k=1,\ldots,K\big\} \tag{A.7}$$

holds with probability $1-o(1)$, $NT\to\infty$.

Proof. Invoking Lemma A.3 shows that the required variables exist and obey

$$\mathrm{P}\big((B_k,B_k^{\mathrm{qc}})\neq(B_k^*,B_k^{\mathrm{qc}*})\big)\le\gamma\big(B_k,B_k^{\mathrm{qc}}\big)\le NC_\kappa\exp(-\kappa T_{\mathrm{block}}).$$

Invoking union bound over the partitions $k=1,2,\ldots,K$ gives

$$\mathrm{P}\big((B_k,B_k^{\mathrm{qc}})\neq(B_k^*,B_k^{\mathrm{qc}*}),\ \text{for some}\ k=1,\ldots,K\big)\le KNC_\kappa\exp(-\kappa T_{\mathrm{block}}),$$

since K is finite and $\log N/T_{\mathrm{block}}=o(1)$ gives $KNC_\kappa\exp(-\kappa T_{\mathrm{block}})=o(1)$. □

COROLLARY A.2 (Convenient Rate Implications). *Consider the setup above. Suppose there exists a sequence V_{NT} such that for some $\psi(B_k^*, B_k^{\text{qc}*})$ is $O_P(V_{NT})$ for some measurable function ψ. Then $\psi(B_k, B_k^{\text{qc}})$ is $O_P(V_{NT})$.*

PROOF OF COROLLARY A.2. Consider any sequence of constants such that $\ell_{NT} \to \infty$. Then

$$\begin{aligned} \mathrm{P}\big(\psi\big(B_k, B_k^{\text{qc}}\big) > \ell_{NT} V_{NT}\big) &\leq^{\text{i}} \mathrm{P}\big(\psi\big(B_k, B_k^{\text{qc}}\big) > \ell_{NT} V_{NT} \cap \mathcal{E}_{\text{berbee}}\big) + \mathrm{P}\big(\mathcal{E}_{\text{berbee}}^c\big) \\ &\leq^{\text{ii}} \mathrm{P}\big(\psi\big(B_k^*, B_k^{\text{qc}*}\big) > \ell_{NT} V_{NT}\big) + \mathrm{P}\big(\mathcal{E}_{\text{berbee}}^c\big) \\ &=^{\text{iii}} o(1), \end{aligned}$$

where (i) follows from union bound, (ii) holds since $\psi(B_k^*, B_k^{\text{qc}*}) = \psi(B_k, B_k^{\text{qc}})$ on $\mathcal{E}_{\text{berbee}}$, and (iii) is assumed in the statement of lemma. □

Consider the following setup. We assume all spaces to be separable and complete. Consider the parameter space $\mathcal{T}$ with elements η, typically a space of functions. Consider also a measurable function (the estimation map) $b^{qc} \mapsto \bar{\eta}(b^{qc})$ that maps $\mathcal{W}^{T-q+1}$ to $\mathcal{T}$. Here, $\mathcal{W}$ is the metric space containing realizations of W_{it} for all i and t. Let $\widehat{\eta}_k = \bar{\eta}(B_k^{\text{qc}})$ denote an estimator constructed on the data B_k^{qc}. Let $b \mapsto \phi(b; \eta)$ be another measurable mapping, indexed by η that maps $\mathcal{W}^q$ to R^{d_ϕ}. We assume that the composition map $(b, b^{qc}) \mapsto \phi(b; \bar{\eta}(b^{qc}))$ is measurable.[1]

COROLLARY A.3. *Suppose there exists a sequence of sets $\{\bar{\mathcal{T}}_{N,T}\} \subseteq \bar{\mathcal{T}}$ obeying the conditions as $NT \to \infty$: (A) $\mathrm{P}(\widehat{\eta}_k \in \bar{\mathcal{T}}_{N,T}) = 1 - o(1)$ and (B) for any sequence $\{\eta_{NT}\} \in \bar{\mathcal{T}}_{N,T}$, $\phi(B_k, \eta_{NT}) = O_P(V_{NT})$. Then $\phi(B_k, \widehat{\eta}_k) = O_P(V_{NT})$.*

PROOF OF COROLLARY A.3. Invoke Lemma A.2 with

$$\psi\big(b, b^{qc}\big) := \phi\big(b, \bar{\eta}\big(b^{qc}\big)\big) 1_{\{\bar{\eta}(b^{qc}) \in \bar{\mathcal{T}}_{N,T}\}}.$$

Union bound implies

$$\begin{aligned} \mathrm{P}\big(\phi(B_k, \widehat{\eta}_k) \geq \ell_{NT} V_{NT}\big) &\leq \mathrm{P}\big(\phi(B_k, \widehat{\eta}_k) \geq \ell_{NT} V_{NT} \cap \widehat{\eta}_k \in \bar{\mathcal{T}}_{N,T}\big) + \mathrm{P}(\widehat{\eta}_k \notin \bar{\mathcal{T}}_{N,T}) \\ &= \mathrm{P}\big(\psi\big(B_k, B_k^{\text{qc}}\big) \geq \ell_{NT} V_{NT} \cap \widehat{\eta}_k \in \bar{\mathcal{T}}_{N,T}\big) + \mathrm{P}(\widehat{\eta}_k \notin \bar{\mathcal{T}}_{N,T}) \\ &\leq \mathrm{P}\big(\psi\big(B_k, B_k^{\text{qc}}\big) \geq \ell_{NT} V_{NT}\big) + o(1), \end{aligned}$$

where the last inequality holds by Condition A. We have that

$$\mathrm{P}\big(\psi\big(B_k, B_k^{\text{qc}}\big) \geq \ell_{NT} V_{NT}\big) \leq \mathrm{P}\big(\psi\big(B_k^*, B_k^{\text{qc}*}\big) \geq \ell_{NT} V_{NT}\big) + o(1),$$

from the previous proof. By Condition B,

$$\begin{aligned} &\mathrm{P}\big(\psi\big(B_k^*, B_k^{qc*}\big) > \ell_{NT} V_{NT} \mid B_k^{\text{qc}*}\big) \\ &\quad = \mathrm{P}\big(\phi\big(B_k^*, \bar{\eta}\big(B_k^{qc*}\big)\big) 1\big\{\bar{\eta}\big(B_k^{qc*}\big) \in \bar{\mathcal{T}}_{N,T}\big\} > \ell_{NT} V_{NT} \mid B_k^{\text{qc}*}\big) = o_P(1). \end{aligned}$$

[1] Otherwise, can use outer probability measures to work with the bounds below.

Therefore, using LIE

$$\mathrm{P}\big(\psi\big(B_k^*, B_k^{\mathrm{qc}*}\big) \geq \ell_{NT} V_{NT}\big) = \mathrm{E}\big[\mathrm{P}\big(\psi\big(B_k^*, B_k^{\mathrm{qc}*}\big) \geq \ell_{NT} V_{NT} | B_k^{\mathrm{qc}*}\big)\big] = o(1),$$

where the final conclusion holds by the boundedness (and, therefore, uniform integrability) of the integrand. □

LEMMA A.6 (Bounds on Cross-Fit Sample Averages). *Let $w \mapsto A(w, \eta)$ be a generic (measurable) matrix-valued function defined on $\mathcal{W}$, indexed by the parameter $\eta \in \bar{\mathcal{T}}$. Define*

$$B_{Ak}(\eta) := (NT_k)^{-1} \sum_{i=1}^{N} \sum_{t \in \mathcal{M}_k} \mathrm{E}_{W_{it}} A(W_{it}, \eta), \tag{A.8}$$

$$V_{Ak}(\eta) := (NT_k)^{-1} \sum_{i=1}^{N} \sum_{t \in \mathcal{M}_k} \big[A(W_{it}, \eta) - \mathrm{E}_{W_{it}} A(W_{it}, \eta)\big]. \tag{A.9}$$

Suppose there exist sequences of constants ζ_{NT}^B and ζ_{NT}^V so that as $NT \to \infty$ for each $k = 1, \ldots, K$:

(1) $\mathrm{P}(\widehat{\eta}_k \in \bar{\mathcal{T}}_{N,T}) = 1 - o(1)$.

(2) *For any sequence $\{\eta_{NT}\} \in \bar{\mathcal{T}}_{N,T}$ and any norm $|||\cdot|||$,*

$$\big|\big|\big|B_{Ak}(\eta_{NT})\big|\big|\big| = O\big(\zeta_{NT}^B\big), \qquad \big|\big|\big|V_{Ak}(\eta_{NT})\big|\big|\big| \lesssim_P \big(\zeta_{NT}^V\big).$$

Then

$$\Bigg|\Bigg|\Bigg|(NT)^{-1} \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{t \in \mathcal{M}_k} A(W_{it}, \hat{\eta}_k)\Bigg|\Bigg|\Bigg| \lesssim_P \big(\zeta_{NT}^V + \zeta_{NT}^B\big).$$

In our case, we will either use $|||\cdot||| = \|\cdot\|_\infty$ (sup-norm) or $|||\cdot||| = \|\cdot\|_2$ (operator norm).

PROOF OF LEMMA A.6. We invoke Corollary A.3 with $\phi(B_k, \eta) := B_{Ak}(\eta) + V_{Ak}(\eta)$. The conditions (A) and (B) are directly assumed in Lemma A.6 as conditions (1) and (2), respectively. Therefore, for each $k \leq K$,

$$\big|\big|\big|B_{Ak}(\widehat{\eta}_k) + V_{Ak}(\widehat{\eta}_k)\big|\big|\big| \lesssim_P \big(\zeta_{NT}^V + \zeta_{NT}^B\big).$$

We next note that with probability converging to one,

$$(NT)^{-1} \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{t \in \mathcal{M}_k} A(W_{it}, \hat{\eta}) = \frac{T_k}{T} \sum_{k=1}^{K} \big[B_{Ak}(\widehat{\eta}_k) + V_{Ak}(\widehat{\eta}_k)\big].$$

Since $T_k \asymp T$, the claim holds by the triangle inequality and the union bound. □

APPENDIX B: TOOLS: TAIL BOUNDS FOR MAXIMA OF SUMS FOR WEAKLY DEPENDENT PANELS

Here, we collect and develop some useful lemmas, some of which can be of interest.

B.1 *Properties of products of sub-Gaussians*

A random variable ξ is (σ^2, α)-sub-exponential if

$$\mathrm{E}e^{\lambda\xi} \leq e^{\lambda^2\sigma^2/2} \quad \text{a.s. } \forall\lambda : |\lambda| \leq \alpha^{-1}. \tag{B.1}$$

A $(\sigma^2, 0)$-sub-exponential is σ^2-sub-Gaussian. Lemma B.2 states concentration inequality for a sub-exponential martingale difference sequence (m.d.s.).

Lemma B.1 (Properties of Sub-Gaussian Random Variables). (1) *Let* $\sigma_X, \sigma_Y > 0$. *If* X *is* σ_X^2*-sub-Gaussian and* Y *is* σ_Y^2*-sub-Gaussian, then* $X + Y$ *is* $(\sigma_X + \sigma_Y)^2$*-sub-Gaussian.* (2) *Let* $\{X_m\}_{m=1}^M$ *be a sequence of* σ^2*-sub-Gaussian random variables. Then* (2*a*) $\sum_{m=1}^M X_m$ *is* $(M^2\sigma^2)$*-sub-Gaussian and* (2*b*) $\max_{1\leq m\leq M} X_m \lesssim_P (\sigma\sqrt{\log d})$. (3) *Furthermore,* $\sum_{m=1}^M X_m$ *is* $(M\sigma^2)$*-sub-Gaussian if* $\{X_m\}_{m=1}^M$ *are independent.* (4) *If* $Y \in [-B, B]$ *a.s.,* Y *is* B^2*-sub-Gaussian.* (5) *If* X *is* σ_X^2*-sub-Gaussian conditional on* Y*, and* $Y \in [-B, B]$ *a.s., then* $X \cdot Y$ *is* $\sigma_X^2 B^2$*-sub-Gaussian.* (6) *If* X_{mn} *are* $\bar{\sigma}^2$*-sub-Gaussian for* $n = 1, 2, \ldots, \bar{N}$ ($\bar{N}$ *finite*) *and* $m = 1, 2, \ldots, M$*, then* $\max_{1\leq m\leq M}\prod_{n=1}^{\bar{N}} |X_{mn}| \lesssim_P ((2\bar{\sigma})^{\bar{N}} \times \log^{\bar{N}/2}(M\bar{N}))$.

Proof of Lemma B.1. We prove (1). By Holder inequality, for any p, q in $[1, \infty)$ such that $1/p + 1/q = 1$,

$$\mathrm{E}e^{\lambda(X+Y)} \leq \left(\mathrm{E}e^{\lambda pX}\right)^{1/p}\left(\mathrm{E}e^{\lambda qY}\right)^{1/q} \leq e^{\lambda^2/2(p\sigma_X^2 + q\sigma_Y^2)}. \tag{B.2}$$

Plugging $p = (\sigma_Y + \sigma_X)/\sigma_X$ and $q = (\sigma_Y + \sigma_X)/\sigma_Y$ into (B.2) gives (4.3) with $\sigma^2 = (\sigma_X + \sigma_Y)^2$. We prove (2a) by induction over M. The statement holds for $M = 1$. The inductive step follows from (1) with $\sigma_X = (M-1)\sigma$ and $\sigma_Y = \sigma$. (2b) is Theorem 1.14 in Rigollet and Hutter (2017). The statements (3) and (4) are Theorem 1.6 and Lemma 1.8 in Rigollet and Hutter (2017), respectively. To see (5), observe that $\mathrm{E}[X \mid Y] = 0$ a.s. by assumption. LIE gives

$$\mathrm{E}_{X,Y}[X \cdot Y] = \mathrm{E}_Y \mathrm{E}[X \mid Y]Y = 0.$$

Furthermore,

$$\mathrm{E}_Y \mathrm{E}\left[e^{\lambda XY} \mid Y\right] \leq \mathrm{E}_Y e^{\lambda^2\sigma^2 Y^2/2} \leq e^{\lambda^2\sigma^2 B^2/2},$$

which gives the result. (6) Invoking union bound for any $t > 0$,

$$\begin{aligned}\mathrm{P}\left(\max_{1\leq m\leq M}\prod_{n=1}^{\bar{N}} |X_{mn}| > t\right) &\leq \sum_{m=1}^M \mathrm{P}\left(\left|\prod_{n=1}^{\bar{N}} X_{mn}\right| > t\right) \\ &\leq \sum_{m=1}^M \sum_{n=1}^{\bar{N}} \mathrm{P}\left(|X_{mn}| > t^{1/\bar{N}}\right) \leq 2M\bar{N}e^{-t^{2/\bar{N}}/2\bar{\sigma}^2}.\end{aligned}$$

Taking $t := \bar{C}(2\bar{\sigma})^{\bar{N}} \log^{\bar{N}/2}(M\bar{N})$ and setting $\bar{C} \to \infty$ makes the R.H.S. above $o(1)$. Conclude that

$$\max_{1\le m\le M} \prod_{n=1}^{\bar{N}} |X_{mn}| \lesssim_P \big((2\bar{\sigma})^{\bar{N}} \log^{\bar{N}/2}(M\bar{N})\big). \qquad \square$$

B.2 *Tails bounds for maxima of sums of martingale differences*

Lemma B.2 (Martingale Difference Sequences; Theorem 2.19 in Wainwright (2019)). (1) *Let* $\{(\xi_m, \Phi_m)\}_{m=1}^M$ *be an m.d.s. obeying*

$$\mathrm{E}\big[e^{\lambda \xi_m} \mid \Phi_{m-1}\big] \le e^{\lambda^2\sigma^2/2} \quad a.s.$$

for any λ *such that* $|\lambda| \le \alpha^{-1}$. *Then the sum* $\sum_{m=1}^M \xi_m$ *is* $(\sigma^2 M, \alpha)$*-sub-exponential and satisfies concentration inequality*

$$\mathrm{P}\left(\left|\sum_{m=1}^M \xi_m\right| \ge t\right) \le \begin{cases} 2e^{-t^2/(2M\sigma^2)}, & 0 \le t \le M\sigma^2/\alpha, \\ 2e^{-t/(2\alpha)}, & t > M\sigma^2/\alpha. \end{cases}$$

(2) *For each* $j : 1 \le j \le d$, *let* $\{(\xi_{mj}, \Phi_m)\}_{m=1}^M$ *be an m.d.s. obeying the conditions above with the same parameters* (σ^2, α). *Then*

$$\mathrm{P}\left(\left\|M^{-1}\sum_{m=1}^M \xi_m\right\|_\infty > t\right) \le \begin{cases} 2e^{\log d - t^2 M/(2\sigma^2)}, & 0 \le t \le \sigma^2/\alpha, \\ 2e^{\log d - tM/(2\alpha)}, & t > \sigma^2/\alpha. \end{cases} \tag{B.3}$$

Proof of Lemma B.2. Lemma B.2 is essentially Theorem 2.19 in Wainwright (2019). Replacing ξ by $c \cdot \xi$ in (B.1) shows that $c \cdot \xi$ is $(c^2\sigma^2, c\alpha)$-sub-exponential. $\square$

Lemma B.3. *Let* $1 \le i \le N$ *and* $1 \le t \le T$ *be the unit and the time indices. Denote the index* m *as*

$$m = m(i,t) = T(i-1) + t. \tag{B.4}$$

Consider a sequence

$$\xi_m = V_{it}U_{it}, \quad m = 1, 2, \dots, M = NT. \tag{B.5}$$

Under Assumption 4.3,

A $\{\xi_m\}_{m=1}^M$ *is a martingale difference sequence with respect to natural filtration*:

$$\mathrm{E}[\xi_m \mid \Phi_{m-1}] := \mathrm{E}[\xi_m \mid \xi_1, \dots, \xi_{m-1}] = 0, \quad \forall m = 1, 2, \dots M.$$

B *Given a large enough constant* C_{VU} *large enough, there exists* NT *large enough such that the maximal norm of the empirical moment vector obeys*

$$\mathrm{P}\big(\|\mathbb{E}_{NT} V_{it}U_{it}\|_\infty > C_{VU}\sqrt{\log d/NT}\big) \le 2/d = o(1). \tag{B.6}$$

PROOF OF LEMMA B.3. By conditional sequential exogeneity (2.4) and independence over i,

$$\mathrm{E}\left[U_{it} \,\Big|\, \bigcup_{t<t'} V_{it'}U_{it'}, \bigcup_{j\neq i}\{(V_{jt}, U_{jt})_{t=1}^{T}\}\right] = 0 \quad \forall i, t.$$

Therefore, the martingale difference property A holds. Union bound and Assumption 4.3 imply

$$\mathrm{P}(|V_{it,j}U_{it}| > t) \le \mathrm{P}(|V_{it,j}| > \sqrt{t}) + \mathrm{P}(|U_{it}| > \sqrt{t}) \le 2e^{-t/2\bar{\sigma}^2}.$$

By Theorem 2.13 in Wainwright (2019), $V_{it,j}U_{it}$ is (σ^2, α)-sub-exponential for some $\sigma, \alpha > 0$ that do not depend on j, N or T. Since the cut-off point σ^2/α in (B.3) does not depend on N, T, for C_{VU} large enough and sample size NT,

$$t := C_{VU}\sqrt{\log d/NT} \le \sigma^2/\alpha.$$

The bound follows:

$$\mathrm{P}(\|\mathbb{E}_{NT}V_{it}U_{it}\|_\infty > C_{VU}\sqrt{\log d/NT}) \le 2/d = o(1). \qquad \square$$

B.3 *Tail bounds for maxima of sums of sub-Gaussian products*

LEMMA B.4 (Tail Bounds for Weakly Dependent Matrices, ℓ_∞-Norm). *Suppose Assumption* 4.1(1) *holds. For each* $j = 1, 2, \dots, d$, *let* $\phi_j(W_{it})$ *be centered* σ^2*-sub-Gaussian random variable for all* i, t *where* $\sigma = \sigma(N, T)$ *and* $\phi_j(W_{it})$ *can depend on* N, T. *Then*

$$\|S\|_\infty := \max_{1\le j\le d}\left|(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi_j(W_{it})\right| \lesssim_P \sigma\sqrt{\log(NT)\log d/NT}. \tag{B.7}$$

REMARK B.1 (Triangular Arrays). Note that all variables and σ can be indexed by (N, T), but we omit the indexing to keep the notation light. Thus, this lemma and all other lemmas stated below apply to triangular arrays.

PROOF OF LEMMA B.4. Let q be the block size such that $1 \le q \le T/2$ and let $L = [T/2q]$. Define the odd blocks

$$B_{i(2l-1)} := [W_{i,(2l-2)q+1}, W_{i,(2l-2)q+2}, \dots, W_{i,(2l-2)q+q}], \quad l = 1, 2, \dots, L, \tag{B.8}$$

the even blocks

$$B_{i(2l)} := [W_{i,(2l-1)q+1}, W_{i,(2l-1)q+2}, \dots, W_{i,(2l-1)q+q}], \quad l = 1, 2, \dots, L, \tag{B.9}$$

and the remainder block, which can be empty, as

$$B_{ir} := [W_{i,2Lq+1}, W_{i,2Lq+2}, \dots, W_{i,T}]. \tag{B.10}$$

Note that $\{B_{i(2l-1)}\}_{l=1}^{L}$ obeys (A.3) and $\{B_{i(2l)}\}_{l=1}^{L}$ obeys (A.3) with $\epsilon = \gamma(q)$. Let $B^*_{i(2l-1)}$ be the Berbee copy of $B_{i(2l-1)}$. Define the Berbee event

$$\mathcal{I}_1 := \left\{B^*_{i(2l-1)} = B_{i(2l-1)} \text{ for all } i, l\right\}.$$

Likewise, let $B^*_{i(2l)}$ and $\mathcal{I}_2$ be the analogs of $B^*_{i(2l-1)}$ and $\mathcal{I}_1$ for even indices. Define the blockwise sum

$$\phi\big(B^*_{i(2l-1)}\big) := \sum_{t=(2l-2)q+1}^{t=(2l-2)q+q} \phi(W^*_{it}), \tag{B.11}$$

$$S^*_{\text{odd}}(q) := (NT)^{-1} \sum_{i=1}^{N}\sum_{l=1}^{L} \phi\big(B^*_{i(2l-1)}\big). \tag{B.12}$$

Let $S^*_{\text{even}}(q)$ be the analog of $S^*_{\text{odd}}(q)$ for even indices. If $T \neq 2Lq$, the remainder block is nonempty, in which case define

$$\phi(B_{ir}) := \sum_{t=2Lq+1}^{T} \phi(W_{it}), \tag{B.13}$$

$$S_{\text{rem}}(q) := (NT)^{-1} \sum_{i=1}^{N} \phi(B_{ir}). \tag{B.14}$$

On the event $\mathcal{I}_1 \cap \mathcal{I}_2$, the union bound gives

$$\|S\|_\infty \le \big\|S^*_{\text{odd}}(q)\big\|_\infty + \big\|S^*_{\text{even}}(q)\big\|_\infty + \big\|S_{\text{rem}}(q)\big\|_\infty. \tag{B.15}$$

Thus,

$$\begin{aligned}
\mathrm{P}\big(\|S\|_\infty \ge 3t\big) \le{}& \mathrm{P}\left(\left\|\sum_{i=1}^{N}\sum_{l=1}^{L}\phi\big(B^*_{i(2l-1)}\big)\right\|_\infty \ge t\right) \\
&+ \mathrm{P}\left(\left\|\sum_{i=1}^{N}\sum_{l=1}^{L}\phi\big(B^*_{i(2l)}\big)\right\|_\infty \ge t\right) + \mathrm{P}\left(\left\|\sum_{i=1}^{N}\phi(B_{ir})\right\|_\infty \ge t\right) + 2NL\gamma(q).
\end{aligned}$$

For each j, $S^*_{\text{odd}_j}(q)$ is $(NT)^{-2}(NL)q^2\sigma^2 \le (q/NT)\sigma^2$-sub-Gaussian by Lemma B.1; similarly, for each j, $S^*_{\text{even}_j}(q)$ is $(q/NT)\sigma^2$-sub-Gaussian. Note that here the dependency on L is linear and not square, because the Berbee blocks are independent. For the remainder block, for each j, $S_{\text{rem}_j}(q)$ is $(NT)^{-2}(N)q^2\sigma^2 \le (q/NT)\sigma^2$-sub-Gaussian since $q \le T$ by Lemma B.1, where we use only independence across i. Since $S^*_{\text{odd}_j}(q)$ is $(q/NT)\sigma^2$-sub-Gaussian for each j, $\|S^*_{\text{odd}}(q)\|_\infty \lesssim_P \sigma\sqrt{q\log d/NT}$ by Lemma B.1(2b). Likewise, $\|S_{\text{rem}}(q)\|_\infty \lesssim_P \sigma\sqrt{q\log d/NT}$ by Lemma B.1(2b). Given the parameter κ in mixing coefficient (4.1), we set block size q to be

$$q = \big\lfloor (2/\kappa)\log(NT)\big\rfloor. \tag{B.16}$$

Invoking the bound (4.1) in Assumption 4.1 gives

$$\begin{aligned} \mathrm{P}(\mathcal{I}_1^c) + \mathrm{P}(\mathcal{I}_2^c) &\le 2N(L-1)\gamma(q) \le (2NT/q)\gamma(q) \\ &\le 2(NT/q)(NT)^{-2} = o\big((NT)^{-1}\big) = o(1), \quad NT \to \infty, \end{aligned}$$

which implies (B.7). □

The following is an extension/clarification of a useful lemma due to Kock and Tang (2019).

Lemma B.5 (Concentration of Products Of Sub-Gaussian Random Variables With Independent Blocks). *Suppose the random variables $Z_{n,m,v,j}$ are uniformly $\bar{\sigma}_n^2$-sub-Gaussian as in (4.3) for $n = 1, 2, \ldots, \bar{N}$, ($\bar{N} \ge 2$ is fixed and finite), $m = 1, 2, \ldots, M$, $v = 1, 2, \ldots, V$, and $j = 1, 2, \ldots, d$. Suppose Z_{n_1,m_1,v_1,j_1} and Z_{n_2,m_2,v_2,j_2} are independent as long as $m_1 \neq m_2$ regardless of the values of other subscripts. Then*

$$\max_{j,v,m} \mathrm{E}\left|\prod_{n=1}^{\bar{N}} Z_{n,m,v,j}\right| \le C_A \prod_{n=1}^{\bar{N}} \bar{\sigma}_n,$$

for some positive constant C_A that depends on $\bar{N}$ and with probability approaching 1,

$$\max_{1\le j\le d}\left|(MV)^{-1}\sum_{m=1}^{M}\sum_{v=1}^{V}\left(\prod_{n=1}^{\bar{N}} Z_{n,m,v,j} - \mathrm{E}\prod_{n=1}^{\bar{N}} Z_{n,m,v,j}\right)\right| \le C_V\sqrt{\log^{\bar{N}+1}(dV)/M}\prod_{n=1}^{\bar{N}}\bar{\sigma}_n,$$

for some positive constant C_V that depends only on $\bar{N}$.

Lemma B.6 (Concentration of Products of Sub-Gaussian Random Variables Under Weak Dependence). *Suppose Assumption* 4.1(1) *holds, and let $\varphi_{nj}(\cdot): \mathcal{W} \to \mathrm{R}$ be a deterministic function. Suppose that $\varphi_{nj}(W_{it})$ are uniformly $\bar{\sigma}_n$-sub-Gaussian as in* (4.3) *for $n = 1, 2, \ldots, \bar{N}$ ($\bar{N} \ge 2$ is fixed and finite) and $j = 1, 2, \ldots, d$ and any i, t. Then*

$$\max_{j,i,t}\left|\mathrm{E}\left[\prod_{n=1}^{\bar{N}}\varphi_{nj}(W_{it})\right]\right| \le C_A\prod_{n=1}^{\bar{N}}\bar{\sigma}_n, \tag{B.17}$$

for some positive constant C_A that depends on $\bar{N}$ and with probability approaching 1,

$$\begin{aligned} \|S\|_\infty &:= \max_{1\le j\le d}\left|(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\prod_{n=1}^{\bar{N}}\varphi_{nj}(W_{it}) - \mathrm{E}\left[\prod_{n=1}^{\bar{N}}\varphi_{nj}(W_{it})\right]\right]\right| \\ &\le \bar{C}_V\sqrt{\log^{\bar{N}+1}\big(d\log(NT)\big)\log(NT)/NT}\prod_{n=1}^{\bar{N}}\bar{\sigma}_n. \end{aligned} \tag{B.18}$$

for some positive constant $\bar{C}_V$ that depends only on $\bar{N}$.

PROOF OF LEMMA B.6. Define

$$\phi(W_{it}) := \{\phi_j(W_{it})\}_{j=1}^d, \qquad \phi_j(W_{it}) := \prod_{n=1}^{\bar{N}} \varphi_{nj}(W_{it}).$$

Let the block size q, the odd blocks, even blocks, and remainder blocks, and events $\mathcal{I}_1$ and $\mathcal{I}_2$ be as defined in the proof of Lemma B.4. Likewise, let $S^*_{\text{odd}}(q)$ be as in (B.11), that is,

$$\phi\big(B^*_{i(2l-1)}\big) := \sum_{t=(2l-2)q+1}^{t=(2l-2)q+q} \phi\big(W^*_{it}\big), \qquad S^*_{\text{odd}}(q) := (NT)^{-1}\sum_{i=1}^{N}\sum_{l=1}^{L}\phi\big(B^*_{i(2l-1)}\big),$$

$S^*_{\text{even}}(q)$ be its analog for the even-numbered blocks, and $S^*_{\text{rem}}(q)$ be as in (B.13). The first claim (B.17) is immediate from the previous lemma. Lemma B.5 with $L \geq 2$ and $M = NL$ and $V = q$ implies that w.p. $1 - o(1)$,

$$\begin{aligned}\big\|S^*_{\text{odd}}(q)\big\|_\infty &:= (Lq/T)\left\|(NLq)^{-1}\sum_{i=1}^{N}\sum_{l=1}^{L}\{\varphi\big(B^*_{i(2l-1)}\big) - \mathrm{E}\varphi\big(B^*_{i(2l-1)}\big)\}\right\|_\infty \\ &\leq (Lq/T)C_V\Big(\sqrt{\log^{\bar{N}+1}(dq)/NL}\Big) \leq^{\mathrm{i}} C_V\Big(\sqrt{\log^{\bar{N}+1}(dq)q/NT}\Big),\end{aligned}$$

where (i) follows from $L = \lfloor T/2q \rfloor \leq T/2q$ and $L \geq \lfloor T/2q \rfloor \geq T/2q - 1 \geq T/4q$. A similar bounds holds for $S^*_{\text{even}}(q)$. If $T_{\text{rem}} \neq 0$, Lemma B.5 with $M = N$ and $V = T_{\text{rem}} \leq q$ implies that w.p. $1 - o(1)$:

$$\begin{aligned}\big\|S_{\text{rem}}(q)\big\|_\infty &:= T_{\text{rem}}/T\left\|(NT_{\text{rem}})^{-1}\sum_{i=1}^{N}\big(\varphi(B_{ir}) - \mathrm{E}\varphi(B_{ir})\big)\right\|_\infty \\ &\leq T_{\text{rem}}/TC_V\Big(\sqrt{\log^{\bar{N}+1}(dT_{\text{rem}})/N}\Big) \\ &\leq q/TC_V\Big(\sqrt{\log^{\bar{N}+1}(dq)/N}\Big).\end{aligned}$$

Plugging $q^2/NT^2 \leq q/NT$ into the R.H.S. above gives the bound $C_V(\sqrt{\log^{\bar{N}+1}(dq)q/NT})$. Let NT be large enough so that $L = \lfloor T/2q \rfloor \geq 2$ and $(2/\kappa) \leq \log(NT)$ so that $q \leq \log^2(NT)$ and $dq \leq (d\log(NT))^2$. Collecting the bounds gives (B.18). Adding up the bounds and plugging choice of $q = (2/\kappa)\log(NT)$ as in (B.16) and noting that $L = \lfloor T/2q \rfloor \geq 2$ for T large enough gives (B.18). □

COROLLARY B.1. *Suppose Assumption* 4.1(1) *holds. Suppose $Z_{1,nit}$ and $Z_{2,nit}$ are d-vectors obtained as (measurable) transformations of W_{it}, whose entries are uniformly $\bar{\sigma}_1^2$ and $\bar{\sigma}_2^2$-sub-Gaussian for $n = 1, 2, \ldots, \bar{N}$. Let U_{it} be uniformly $\bar{\sigma}^2$-sub-Gaussian and $g \geq 0$ be a finite power. Then*

$$\max_{1\leq k,j\leq d}\max_{i,t}\left|\mathrm{E}\left[\prod_{n=1}^{\bar{N}} Z_{1,nitk}Z_{2,nitj}U_{it}^{2g}\right]\right| \leq C_A\big(\bar{\sigma}_1\bar{\sigma}_2\bar{\sigma}^{2g}\big)^{\bar{N}}, \tag{B.19}$$

$$\max_{1\le k,j\le d}\left|(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\prod_{n=1}^{\bar N} Z_{1,nitk}Z_{2,nitj}U_{it}^{2g} - \mathrm{E}\left[\prod_{n=1}^{\bar N} Z_{1,nitk}Z_{2,nitj}U_{it}^{2g}\right]\right]\right|$$
$$\le \bar C_V\big(\bar\sigma_1\bar\sigma_2\bar\sigma^{2g}\big)^{\bar N}\Big(\sqrt{\log^{2\bar N+2g+1}\big(d^2\log(NT)\big)\log(NT)/NT}\Big). \tag{B.20}$$

Remark B.2. Suppose Assumptions 4.1 and 4.3 hold. Invoking (B.19) with $\bar N = 1$ and $Z_{1,it} = Z_{2,it} = V_{it}$ implies for some finite $\sigma_V < \infty$,

$$\max_{it}\big\|\mathrm{E}V_{it}V_{it}'\big\|_\infty \le \max_{itj}\mathrm{E}V_{itj}^2 \le \sigma_V^2.$$

Likewise, Assumption 4.3 implies for some finite $\sigma_{VU} < \infty$,

$$\sup_{it}\mathrm{E}\big[U_{it}^2 \mid V_{it}\big] \le \sigma_{VU}^2 \quad \text{a.s.}$$

B.4 *Some technical lemmas*

Here, we provide technical extensions of the results in Kock and Tang (2019), keeping the notation as in the original Kock and Tang (2019) and references therein.

Lemma B.7 (Theorem 2.1 in Fan, Grama, and Liu (2012), Proposition F.1 in Kock and Tang (2019)). *Let $\alpha \in (0,1)$. Assume that $(X_i,\mathcal{F}_i)_{i=1}^n$ is a sequence of supermartingale differences satisfying* $\sup_i \mathrm{E}[e^{|X_i|^{\frac{2\alpha}{1-\alpha}}}] \le C_1$ *for some constant $C_1 \in (0,\infty)$. Define $S_k := \sum_{i=1}^k X_i$. Then, for all $\epsilon > 0$,*

$$\mathrm{P}\Big(\max_{1\le k\le n} S_k \ge n\epsilon\Big) \le C(\alpha,n,\epsilon)e^{-(\epsilon/4)^{2\alpha}n^\alpha},$$

where

$$C(\alpha,n,\epsilon) := 2 + 35C_1\left[\frac{1}{16^{1-\alpha}(n\epsilon^2)^\alpha} + \frac{1}{n\epsilon^2}\left(\frac{3(1-\alpha)}{2\alpha}\right)^{\frac{1-\alpha}{\alpha}}\right].$$

Lemma B.8 (Proposition F.2 in Kock and Tang (2019)). *Let $\alpha \in (0,1)$. Assume that $(X_i,\mathcal{F}_i)_{i=1}^n$ is a sequence of martingale differences satisfying* $\sup_i \mathrm{E}[e^{D|X_i|^{\frac{2\alpha}{1-\alpha}}}] \le C_1$ *for some positive constant D, where $C_1 \ge 1$ can change with the sample size n. Then, for all $\epsilon \ge 1/\sqrt{n}$,*

$$\mathrm{P}\left(\left|\sum_{i=1}^n X_i\right| \ge n\epsilon\right) \le C_1 A(\alpha)e^{-K(\epsilon^2 n)^\alpha}, \quad K = \big(D^{\frac{1-\alpha}{2\alpha}}/4\big)^{2\alpha},$$

where

$$A = A(\alpha) = 2 + 35\left[\frac{1}{16^{1-\alpha}} + \left(\frac{3(1-\alpha)}{2\alpha}\right)^{\frac{1-\alpha}{\alpha}}\right].$$

Lemma B.8 restates Proposition F.2 in Kock and Tang (2019) with explicit constants in tail bounds.

PROOF. Note that for some positive constant D,

$$\mathrm{P}\left(\sum_{i=1}^{n} X_i \geq n\epsilon\right) = \mathrm{P}\left(\sum_{i=1}^{n} D^{\frac{1-\alpha}{2\alpha}} X_i \geq nD^{\frac{1-\alpha}{2\alpha}}\epsilon\right) = \mathrm{P}\left(\sum_{i=1}^{n} Y_i \geq n\delta\right),$$

where $Y_i := D^{\frac{1-\alpha}{2\alpha}} X_i$ and $\delta := D^{\frac{1-\alpha}{2\alpha}}\epsilon$. Now $(Y_i)_{i=1}^n$ is a sequence of martingale differences satisfying $\sup_i \mathrm{E}[e^{|Y_i|^{\frac{2\alpha}{1-\alpha}}}] \leq C_1$. Invoking the preceding theorem, we have

$$\mathrm{P}\left(\sum_{i=1}^{n} Y_i \geq n\delta\right) \leq C(\alpha, n, \delta) e^{-(\delta/4)^{2\alpha} n^{\alpha}}.$$

$(-Y_i)_{i=1}^n$ is also a sequence of martingale differences satisfying the same exponential moment condition. Thus,

$$\begin{aligned}\mathrm{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq n\epsilon\right) &= \mathrm{P}\left(\left|\sum_{i=1}^{n} Y_i\right| \geq n\delta\right) \leq 2C(\alpha, n, \delta) e^{-(\delta/4)^{2\alpha} n^{\alpha}} \\ &= 2C\big(\alpha, n, D^{\frac{1-\alpha}{2\alpha}}\epsilon\big) e^{-(D^{\frac{1-\alpha}{2\alpha}}\epsilon/4)^{2\alpha} n^{\alpha}} \leq C_1 A(\alpha) e^{-K\epsilon^{2\alpha} n^{\alpha}},\end{aligned}$$

where we can select

$$A = A(\alpha) = 2 + 35\left[\frac{1}{16^{1-\alpha}} + \left(\frac{3(1-\alpha)}{2\alpha}\right)^{\frac{1-\alpha}{\alpha}}\right]$$

and K as defined above. □

The following lemma is inspired by Proposition F.3 of Kock and Tang (2019). The difference is that the constants are made explicit to make the results applicable to arrays; and part of the proof was replaced by another argument (as we were not able to follow one step in their proof).[2]

LEMMA B.9. *Suppose we have random variables $Z_{l,i,t,j}$ uniformly $(K, \sigma_l^2) > 0$ sub-Gaussian for $l = 1, \ldots, L$ ($L \geq 2$ fixed), $i = 1, \ldots, N$, $t = 1, \ldots, T$ and $j = 1, \ldots, p$, that is,*

$$\mathrm{P}\big(\big|\sigma_l^{-1} Z_{l,t,i,j}\big| \geq \varepsilon\big) \leq K \exp\big(-\varepsilon^2\big),$$

and Z_{l_2,i_2,t_2,j_2} are independent as long as $i_1 \neq i_2$ regardless of the values of other subscripts. Then we have that (1)

$$\max_{j,t,i} \mathrm{E}\left|\prod_{l=1}^{L} Z_{l,i,t,j}\right| \leq \big(L!(\log 2)^{-1/2}(1+K)^{1/2}\big) \prod_{l=1}^{L} \sigma_l,$$

[2]KT's proof uses the inequality $(x-(y\wedge x))^{2/L} \leq x^{2/L} - (y\wedge x)^{2/L}$, for $x > 0$ and $y > 0$. This inequality is not true (e.g., with $x = 10$, $y = 1$, $L = 4$, the inequality implies $3 < 2.163$), so we changed the middle part of the proof; the end result is preserved; none of conclusions in KT are affected.

and (2) *with probability* $1 - A'(pT)^{-1/2}$,

$$\max_{1 \le j \le d} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\prod_{l=1}^{L} Z_{l,i,t,j} - \mathrm{E}\left[\prod_{l=1}^{L} Z_{l,i,t,j} \right] \right) \right| \le M \left(\sqrt{\frac{\left(\log(pT)\right)^{L+1}}{N}} \right) \prod_{l=1}^{L} \sigma_l,$$

for $M > M'$, *and some positive constants* A' *and* M' *that only depend on* L *and* K.

Proof. Hölder's inequality gives

$$\max_{j,t,i} \mathrm{E} \left| \prod_{l=1}^{L} \sigma_l^{-1} Z_{l,i,t,j} \right| \le \max_{j,t,i} \prod_{l=1}^{L} \left(\mathrm{E} \left| \sigma_l^{-1} Z_{l,i,t,j} \right|^L \right)^{\frac{1}{L}},$$

where

$$\begin{aligned} \left(\mathrm{E} \left| \sigma_l^{-1} Z_{l,i,t,j} \right|^L \right)^{\frac{1}{L}} &\le L! \left\| \sigma_l^{-1} Z_{l,i,t,j} \right\|_{\psi_1} \\ &\le L! (\log 2)^{-1/2} \left\| \sigma_l^{-1} Z_{l,i,t,j} \right\|_{\psi_2} \le L! (\log 2)^{-1/2} (1+K)^{1/2} =: A, \end{aligned}$$

where the first two inequalities are from van der Vaart and Wellner (1996, page 95) and the third inequality from Lemma 2.2.1 in van der Vaart and Wellner (1996). Thus,

$$\max_{j,t,i} \mathrm{E} \left| \prod_{l=1}^{L} \sigma_l^{-1} Z_{l,i,t,j} \right| \le \left(L! (\log 2)^{-1/2} (1+K)^{1/2} \right) =: A.$$

This implies the first claim, after multiplying both sides by $\prod_{l=1}^{L} \sigma_l$. Let

$$X_{i,t,j} = \prod_{l=1}^{L} \sigma_l^{-1} Z_{l,i,t,j} - \mathrm{E}\left[\prod_{l=1}^{L} \sigma_l^{-1} Z_{l,i,t,j} \right].$$

For every $\epsilon \ge 0$,

$$\begin{aligned} \mathrm{P}\left(|X_{i,t,j}| \ge 2\epsilon \right) &\le \mathrm{P}\left(\left| \prod_{l=1}^{L} \sigma_l^{-1} Z_{l,i,t,j} \right| \ge \epsilon \right) + \mathrm{P}\left(\left| \mathrm{E} \prod_{l=1}^{L} \sigma_l^{-1} Z_{l,i,t,j} \right| \ge \epsilon \right) \\ &\le \sum_{l=1}^{L} \mathrm{P}\left(\sigma_l^{-1} |Z_{l,i,t,j}| \ge \varepsilon^{1/L} \right) + 1(\epsilon \le A) \\ &\le LKe^{-\epsilon^{2/L}} + 1\left(\epsilon^{2/L} \le A^{2/L} \right) \\ &\le LKe^{-\epsilon^{2/L}} + e^{A^{2/L}} e^{-\epsilon^{2/L}} = K' e^{-\epsilon^{2/L}}, \quad \left(K' := \left(LK + e^{A^{2/L}} \right) \right). \end{aligned}$$

Let

$$X_{i,j} := \frac{1}{T} \sum_{t=1}^{T} \left(\prod_{l=1}^{L} \sigma_l^{-1} Z_{l,i,t,j} - \mathrm{E}\left[\prod_{l=1}^{L} \sigma_l^{-1} Z_{l,i,t,j} \right] \right).$$

For every $\epsilon \geq 0$,

$$\mathrm{P}\big(|X_{i,j}| \geq 2\epsilon\big) \leq \mathrm{P}\Big(\max_{1\leq t\leq T} |X_{i,t,j}| \geq 2\epsilon\Big) \leq TK' e^{-\epsilon^{2/L}}.$$

Consider some positive constant $D < 1$, then as van der Vaart and Wellner (1996, page 96), using Fubini and change of order of integration:

$$\mathrm{E}\big[e^{D|X_{i,j}/2|^{2/L}}\big] = \int_{x\in\mathbb{R}} \int_0^{|x/2|^{2/L}} De^{Ds}\, ds P(dx) + 1 = \int_0^\infty De^{Ds}\mathrm{P}\big(|X_{i,j}| > 2s^{L/2}\big)\, ds + 1.$$

This is further bounded by

$$\int_0^\infty TK'De^{(D-1)s}\, ds + 1 = \frac{TK'D}{1-D} + 1 \leq BT; \quad \Big(B := \frac{K'D}{1-D} + 1\Big).$$

Then we can use independence across i to invoke the previous Lemma B.8 with $\alpha = \frac{1}{L+1}$ and $C_1 = BT$, for $\epsilon \geq \frac{1}{\sqrt{N}}$,

$$\mathrm{P}\left(\left|\sum_{i=1}^N \frac{1}{T}\sum_{t=1}^T X_{i,t,j}\right| \geq 2N\epsilon\right) \leq A'Te^{-K''(\epsilon^2 N)^{\frac{1}{L+1}}}$$

for positive constants A' and K'' that depend only on K, L, and D.

Setting

$$\epsilon = \sqrt{\frac{M\big(\log(pT)\big)^{L+1}}{N}}$$

for some $M \geq 1$, we have

$$\mathrm{P}\left(\max_{1\leq j\leq p}\left|\sum_{i=1}^N \frac{1}{T}\sum_{t=1}^T X_{i,t,j}\right| \geq 2N\epsilon\right) \leq pA'Te^{-K''(\epsilon^2 N)^{\frac{1}{L+1}}} = A'(pT)^{1-K''M\frac{1}{L+1}}.$$

Therefore, with probability $1 - A'(pT)^{1-K''M\frac{1}{L+1}}$,

$$\max_j \left|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T (X_{i,t,j})\right| \leq 2M\left(\sqrt{\frac{\big(\log(pT)\big)^{L+1}}{N}}\right),$$

for any $M \geq 1$. Setting M large enough such that

$$1 - K''M\frac{1}{L+1} < -\frac{1}{2},$$

guarantees that the bound holds with probability at most $A'(pT)^{-1/2}$. The bounds can be then be restated as in the statement of the theorem. □

Appendix C: Tools: High-dimensional central limit theorems for weakly dependent data

Let $\{X_m\}_{m=1}^M$ be a weakly dependent martingale difference sequence (m.d.s.) with respect to natural filtration. Define its β-mixing coefficient,

$$\gamma_X(q) = \sup_{m \leq M} \gamma\big((X_1, \ldots, X_{m-1}, X_m), (X_{m+q}, X_{m+q+1}, \ldots)\big).$$

The scaled sum

$$S_X = M^{-1/2} \sum_{m=1}^{M} X_m$$

has the variance

$$\Sigma_G := M^{-1} \sum_{m=1}^{M} \mathrm{E} X_m X_m'. \tag{C.1}$$

The distribution of the scaled sum over the cubes can be approximated by the Gaussian distribution $N(0, \Sigma_G)$ over the cubes, as shown in the lemma below.

We will introduce the following notation. For some numbers $\bar{r} = \bar{r}_{NT}$, $\bar{q} = \bar{q}_{NT}$, and $L = \lfloor M/(\bar{q} + \bar{r}) \rfloor$, define Bernstein's "large" and "small" blocks of size $\bar{q}$ and $\bar{r}$:

$$P_l = \big\{(l-1)(\bar{q}+\bar{r})+1, \ldots, (l-1)(\bar{q}+\bar{r})+\bar{q}\big\}, \quad l = 1, 2, \ldots, L,$$
$$Q_l = \big\{(l-1)(\bar{q}+\bar{r})+1+\bar{q}, \ldots, l(\bar{q}+\bar{r})\big\}$$

and let

$$S_l := \sum_{m \in P_l} X_m, \qquad U_l := \sum_{m \in Q_l} X_m, \qquad U_{L+1} := \sum_{m=L(\bar{q}+\bar{r})+1}^{M} X_m.$$

Denote

$$\Sigma_P := (L\bar{q})^{-1} \sum_{l=1}^{L} \mathrm{E} S_l S_l' = (L\bar{q})^{-1} \sum_{l=1}^{L} \sum_{m \in P_l} \mathrm{E} X_m X_m' \tag{C.2}$$

and observe that

$$\Sigma_G = (L\bar{q}/M)\Sigma_P + M^{-1} \sum_{l=1}^{L+1} \mathrm{E} U_l U_l'.$$

The following result is useful both in the proof below and also for performing Gaussian inference, where we replaced unknown variance-covariance matrix by an estimated one.

LEMMA C.1 (Comparison of Distributions). *Let* $X \sim N(0, \Sigma_X)$ *and* $Y \sim N(0, \Sigma_Y)$ *be centered normal* d*-vectors, and let* $\Delta_{XY} := \|\Sigma_X - \Sigma_Y\|_\infty$. *Suppose* $\min_{1\le j\le d}(\Sigma_Y)_{jj} > 0$. *Then*

$$\sup_{t\ge 0}\big|P\big(\|X\|_\infty \le t\big) - P\big(\|Y\|_\infty \le t\big)\big| \le C'\big(\Delta_{XY}\log^2(2d)\big)^{1/2}, \tag{C.3}$$

where $C' > 0$ *depends only on* $\min_{1\le j\le d}(\Sigma_Y)_{jj}$ *and* $\max_{1\le j\le d}(\Sigma_Y)_{jj}$.

Lemma C.1 follows from Proposition 2.1 in Chernozhukov, Chetverikov, Kato, and Koike (2019) for vectors $\overline{X} = (X, -X)$ and $\overline{Y} = (Y, -Y)$ and

$$\Sigma_{\overline{X}} = \begin{pmatrix} \Sigma_X & -\Sigma_X \\ -\Sigma_X & \Sigma_X \end{pmatrix}, \qquad \Sigma_{\overline{Y}} = \begin{pmatrix} \Sigma_Y & -\Sigma_Y \\ -\Sigma_Y & \Sigma_Y \end{pmatrix}, \qquad \|\Sigma_{\overline{X}} - \Sigma_{\overline{Y}}\|_\infty = \Delta_{XY}.$$

Another result is the following anticoncentration property. This result is useful for showing that linearization errors do not impact the behavior of the key statistics. The statistics are approximate means, namely averages of some centered influence functions plus linearization errors.

LEMMA C.2 (Anticoncentration). *Let* $X = (X_1, X_2, \dots, X_d)' \sim N(0, \Sigma_X)$ *be a centered Gaussian random vector in* R^d. *Assume* $\min_{1\le j\le d}(\Sigma_X)_{jj} > 0$. *Then*

$$\sup_{t\in \mathrm{R}} P\big(\big|\|X\|_\infty - t\big| \le \epsilon\big) \le C\epsilon\sqrt{1 \vee \log(2d/\epsilon)}, \tag{C.4}$$

where $C > 0$ *depends on* $\min_{1\le j\le d}(\Sigma_X)_{jj}$ *and* $\max_{1\le j\le d}(\Sigma_X)_{jj}$.

Lemma C.2 follows from Corollary 1 in Chernozhukov, Chetverikov, and Kato (2015) with $\overline{X} = (X, -X)$.

The following result is a consequence of Theorem E.1 in Chernozhukov, Chetverikov, and Kato (2019) for martingale difference sequence.

LEMMA C.3 (High-Dimensional CLT for Martingale Difference Sequence Under Weak Dependence). *Let* $\{X_m\}_{m=1}^M$ *be a weakly dependent m.d.s. of* d*-vectors obeying for* $D_M \ge 1$:

$$\sup_{m\le M}\|X_m\|_\infty \le D_M \quad a.s.$$

Suppose there exist constants $0 < a_1 \le A_1$ *and* $0 < c_2 < 1/4$ *such that*

$$a_1 \le \min_{1\le j\le d}\min_{1\le m\le M} \mathrm{Var}\, X_{mj} \le \max_{1\le j\le d}\sup_{1\le m\le M} \mathrm{Var}\, X_{mj} \le A_1, \tag{C.5}$$

and let $\bar{r}$ *and* $\bar{q}$ *be such that* $\bar{r}/\bar{q} \le A_1 M^{-c_2}\log^{-2} d$ *and*

$$\max\big\{\bar{r}D_M\log^{3/2} d, \bar{q}D_M\log^{1/2} d, \sqrt{\bar{q}}D_M\log^{7/2}(dM)\big\} \le A_1 M^{1/2-c_2}. \tag{C.6}$$

Then there exist constants c_X, $C_X > 0$ depending only on a_1, A_1, c_2 such that

$$\sup_{t\geq 0}\left|\mathrm{P}\big(\|S_X\|_\infty \leq t\big) - \mathrm{P}\big(\|G_P\|_\infty < t\big)\right| \leq 2\frac{M}{\bar{q}+\bar{r}}\gamma_X(\bar{r}) + C_X M^{-c_X}, \tag{C.7}$$

where $G_P \sim N(0, \Sigma_P)$ is a centered normal d-vector.

Note that this result uses Σ_P as the variance in the Gaussian approximation. In our application, we will be using Σ_G in place of Σ_P (i.e., Lemma C.5) so as not to worry about omitting small blocks. Therefore below, we will provide a sequence of the results that allow this replacement.

Proof of Lemma C.3. Let

$$\overline{X}_m := (X_m, -X_m), \quad m = 1, 2, \ldots, M$$

be a sequence of $2d$-vectors. Observe that $\{\overline{X}_m\}_{m=1}^M$ is an m.d.s. It obeys

$$\sup_{m\leq M}\|\overline{X}_m\|_\infty \leq D_M, \quad \text{a.s.}, \qquad \gamma_{\bar{X}}(q) = \gamma_X(q) \quad \forall q.$$

By construction, for any integer r,

$$\bar{\sigma}^2(r) = \max_{1\leq j\leq d}\max_I \mathrm{Var}\left(r^{-1/2}\sum_{m\in I} X_{mj}\right) = \max_{1\leq j\leq 2d}\max_I \mathrm{Var}\left(r^{-1/2}\sum_{m\in I}\overline{X}_{mj}\right),$$

$$\underline{\sigma}^2(r) = \min_{1\leq j\leq d}\min_I \mathrm{Var}\left(r^{-1/2}\sum_{m\in I} X_{mj}\right) = \min_{1\leq j\leq 2d}\min_I \mathrm{Var}\left(r^{-1/2}\sum_{m\in I}\overline{X}_{mj}\right),$$

where $\max_I$ and $\min_I$ are taken over the sets $I = \{i+1, i+2, \ldots, i+r\}$ of size r. Theorem E.1 in Chernozhukov, Chetverikov, and Kato (2019) requires

$$a_1 \leq \underline{\sigma}^2(\bar{q}) \leq \bar{\sigma}^2(\bar{q}) \vee \bar{\sigma}^2(\bar{r}) \leq A_1. \tag{C.8}$$

Because $\{X_m\}_{m=1}^M$ is an m.d.s.,

$$\mathrm{Cov}(X_{m_1}, X_{m_2}) = 0 \in \mathrm{R}^{d\times d} \quad \text{for } m_1 \neq m_2.$$

Therefore, for any r and any $I = \{i+1, i+2, \ldots, i+r\}$,

$$a_1 \leq \mathrm{Var}\left(r^{-1/2}\sum_{m\in I} X_{mj}\right) = r^{-1}\sum_{m\in I}\mathrm{Var}(X_{mj}) \leq A_1, \quad 1 \leq j \leq d,$$

which implies (C.8). All other conditions of Theorem E.1 in Chernozhukov, Chetverikov, and Kato (2019) are satisfied. Invoking Theorem E.1 in Chernozhukov, Chetverikov, and Kato (2019) with

$$T := \max_{1\leq j\leq 2d} M^{-1/2}\sum_{m=1}^M \overline{X}_{mj} = \|S_X\|_\infty$$

and

$$\overline{G}_P \sim N(0, \Sigma_{G_P})$$

being a centered normal $(2d)$-vector with

$$\Sigma_{G_P} = \begin{pmatrix} \Sigma_P & -\Sigma_P \\ -\Sigma_P & \Sigma_P \end{pmatrix},$$

gives (C.7). □

LEMMA C.4 (Comparison of Distributions, cont.). *Consider the setup above with* $\Sigma_X = \Sigma_G$ *and* $\Sigma_Y = \Sigma_P$*, where* Σ_G *and* Σ_P *are as in* (C.2) *and* (C.1) *where*

$$\sup_{1\le m\le M} \|\mathrm{E} X_m X_m'\|_\infty \le \sup_{1\le m\le M} \sup_{1\le j\le d} \mathrm{Var}(X_{mj}) \le A_1.$$

For some $c_2 \in (0, 1/4)$*, assume that the growth condition holds*:

$$D_M \log d \log M \log^{7/2}(dM) \lesssim M^{1/2-2c_2}$$

and $\log^4 d \log^2 M = o(\sqrt{M})$*. Then the max distance* $\Delta_{GP} := \|\Sigma_G - \Sigma_P\|_\infty$ *obeys*

$$\left(\Delta_{GP} \log^2 d\right)^{1/2} \lesssim M^{-c_2/2}.$$

PROOF OF LEMMA C.4. Observe that

$$\Sigma_G - \Sigma_P = (L\bar{q}/M - 1)\Sigma_P + M^{-1}\sum_{l=1}^{L+1} \mathrm{E} U_l U_l'.$$

Since $L = \lfloor M/(\bar{q}+\bar{r})\rfloor$, $L \ge M/(\bar{q}+\bar{r}) - 1$. Therefore,

$$1 - L\bar{q}/M \le 1 - \bar{q}/(\bar{q}+\bar{r}) + \bar{q}/M = \bar{r}/(\bar{q}+\bar{r}) + \bar{q}/M \le \bar{r}/\bar{q} + \bar{q}/M.$$

Furthermore, $(L+1)/M \le 2L/M \le 2/\bar{q}$. The following bound holds:

$$\Delta_{GP} \le \left((1 - L\bar{q}/M) + (L+1)/M\right) \sup_{1\le m\le M} \|\mathrm{E} X_m X_m'\|_\infty = O(\bar{r}/\bar{q} \vee \bar{q}/M \vee 1/\bar{q}).$$

Taking $\bar{q} = M^{c_2}\log^2 d \log^2 M$ and $\bar{r} = (2/\kappa)\log M$ give

$$\begin{aligned} \bar{r}/\bar{q} &= (2/\kappa)M^{-c_2}\log^{-2} d \log^{-1} M = o\left(M^{-c_2}\log^{-2} d\right), \\ \bar{q}/M &= M^{c_2-1}\log^2 d \log^2 M \overset{\mathrm{i}}{=} o\left(M^{-c_2}\log^{-2} d\right), \\ 1/\bar{q} &= M^{-c_2}\log^{-2} d \log^{-2} M = o\left(M^{-c_2}\log^{-2} d\right), \end{aligned}$$

where (i) follows from $c_2 < 1/4$ and

$$\log^4 d \log^2 M = o\left(M^{1-2\cdot 1/4}\right) = o\left(M^{1-2c_2}\right).$$

Plugging $\Delta_{GP} = o(M^{-c_2}\log^{-2} d)$ into $(\Delta_{GP}\log^2(2d))^{1/2}$ gives

$$\left(\Delta_{GP}\log^2(2d)\right)^{1/2} = o\left(M^{-c_2/2}\right).$$

□

REMARK C.1 (Sufficient Growth Condition). If the growth condition holds,

$$D_M \log d \log M \log^{7/2}(dM) \lesssim M^{1/2-2c_2}, \tag{C.9}$$

then

$$\bar{r} = \log M, \qquad \bar{q} = M^{c_2} \log^2 d \log^2 M, \tag{C.10}$$

obeys (C.6) and $\bar{r}/\bar{q} \leq A_1 M^{-c_2} \log^{-2} d$ for M large enough.

PROOF OF REMARK C.1. Let M be large enough such that $M^{-c_2/2} \leq A_1$ and $(2/\kappa) \times \log^{-1} M \leq A_1$. Then the growth condition

$$D_M \log d \log M \log^{7/2}(dM) \lesssim M^{1/2-2c_2} \leq A_1 M^{1/2-3/2c_2}$$

implies the third inequality in (C.6),

$$\sqrt{\bar{q}} D_M \log^{7/2}(dM) \leq A_1 M^{1/2-c_2}.$$

Next, for $d \geq e$ such that $\log d \geq 1$, and

$$\begin{aligned} M^{-c_2} D_M \bar{q} \log^{1/2} d &= D_M \log^{5/2} d \log^2 M \\ &\leq D_M \log^{5/2}(dM) \log M \log(dM) \log d \leq A_1 M^{1/2-3/2c_2}. \end{aligned}$$

Multiplying both sides by M^{c_2} gives

$$D_M \bar{q} \log^{1/2} d \leq A_1 M^{1/2-c_2},$$

which coincides with the second inequality in (C.6). Finally,

$$D_M \bar{r} \log^{3/2} d = (2/\kappa) D_M \log M \log^{3/2} d \leq D_M \bar{q} \log^{1/2} d,$$

as long as $(2/\kappa) \leq \log M$, which verifies (C.6). For M large enough, $\bar{r}/\bar{q} = 2/\kappa M^{-c_2} \times \log^{-2} d \log^{-1} M \leq A_1$. □

LEMMA C.5 (Summary). *Let $\{X_m\}_{m=1}^M$ be a weakly dependent m.d.s. of d-vectors obeying for $D_M \geq 1$:*

$$\sup_{m \leq M} \|X_m\|_\infty \leq D_M \quad a.s.$$

Suppose there exist constants $0 < a_1 \leq A_1$ such that

$$a_1 \leq \min_{1 \leq j \leq d} \min_{1 \leq m \leq M} \operatorname{Var} X_{mj} \leq \max_{1 \leq j \leq d} \sup_{1 \leq m \leq M} \operatorname{Var} X_{mj} \leq A_1.$$

For some constant $c_2 \in (0, 1/4)$, the growth condition (C.5) holds, namely

$$D_M \log d \log M \log^{7/2}(dM) \lesssim M^{1/2-2c_2},$$

and $\log^4 d \log^2 M = o(M^{1/2})$*. Then there exist constants* $c_X, C_X > 0$ *depending only on* a_1, A_1, c_2 *such that for* $\bar{r} = (2/\kappa \log M)$ *and* $\bar{q} = M^{c_2} \log^2 d \log^2 M$,

$$\sup_{t\geq 0} \left| \mathrm{P}\big(\|S_X\|_\infty \leq t\big) - \mathrm{P}\big(\|G_\Sigma\|_\infty < t\big) \right| \lesssim C_X M^{-c_X} + M^{-c_2/2}, \tag{C.11}$$

where $G_\Sigma \sim N(0, \Sigma_G)$ *is a centered normal* d*-vector.*

Triangular inequality gives

$$\sup_{t\geq 0} \left| \mathrm{P}\big(\|S_X\|_\infty \leq t\big) - \mathrm{P}\big(\|G_\Sigma\|_\infty < t\big) \right| \tag{C.12}$$

$$\leq \sup_{t\geq 0} \left| \mathrm{P}\big(\|S_X\|_\infty \leq t\big) - \mathrm{P}\big(\|G_P\|_\infty < t\big) \right| + \sup_{t\geq 0} \left| \mathrm{P}\big(\|G_P\|_\infty < t\big) - \mathrm{P}\big(\|G_\Sigma\|_\infty < t\big) \right|$$

$$\lesssim 2\frac{M}{\bar{q}+\bar{r}} \gamma(\bar{r}) + C_X M^{-c_X} + M^{-c_2/2} = o\big(M^{-c_2/2} + M^{-c_X}\big). \tag{C.13}$$

Appendix D: Proofs for Section 4

D.1 *Bounds on errors for estimating Q and gradient S*

Below, we define the following terms that appear in the analysis of $\widehat{Q}$ and the least squares gradient S. In what follows, we use the notation defined in the main text heavily, without further warning.

Define the first-stage approximation error as a function of $d(\cdot)$ and $l(\cdot)$:

$$R_{it}(\mathbf{d}, \mathbf{l}) := l_{i0}(X_{it}) - l_i(X_{it}) - \big(d_{i0}(X_{it}) - d_i(X_{it})\big)' \beta_0. \tag{D.1}$$

Define the first-order error terms

$$\bar{a} := \mathbb{E}_{NT} V_{it}\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big) = \mathbb{E}_{NT} V_{it}(\widehat{V}_{it} - V_{it}), \tag{D.2}$$

$$\bar{m} = \mathbb{E}_{NT} V_{it}\big(l_{i0}(X_{it}) - \widehat{l}_i(X_{it})\big) = \mathbb{E}_{NT} V_{it}(\widehat{\widetilde{Y}}_{it} - \widetilde{Y}_{it}), \tag{D.3}$$

$$\bar{f} = \mathbb{E}_{NT} U_{it}\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big) = \mathbb{E}_{NT} U_{it}(\widehat{V}_{it} - V_{it}), \tag{D.4}$$

$$\bar{e} = \mathbb{E}_{NT} V_{it} R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}}) = \bar{m} - \bar{a}' \beta_0, \tag{D.5}$$

the second-order error terms,

$$\bar{b} = \mathbb{E}_{NT}\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)', \tag{D.6}$$

$$\bar{z} = \mathbb{E}_{NT}\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)\big(l_{i0}(X_{it}) - \widehat{l}_i(X_{it})\big), \tag{D.7}$$

$$\bar{g} = \mathbb{E}_{NT}\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big) R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}}) = \bar{z} - \bar{b}' \beta_0. \tag{D.8}$$

Lemma D.1 (First-Order Terms). *Under Assumptions* 4.1–4.5, *we have that*

$$\|\bar{a}\|_\infty \lesssim_P \big(\mathbf{d}_{NT,\infty} \sqrt{\log(dNT)/NT}\big), \tag{D.9}$$

$$\|\bar{m}\|_\infty \lesssim_P \big(\mathbf{l}_{NT,\infty} \sqrt{\log(dNT)/NT}\big), \tag{D.10}$$

$$\|\bar{f}\|_\infty \lesssim_P \big(\mathbf{d}_{NT,\infty}\sqrt{\log(dNT)/NT}\big), \tag{D.11}$$

$$\|\bar{e}\|_\infty \lesssim_P \big(\sqrt{\log(dNT)/NT}\big(\mathbf{d}_{NT,\infty}\|\beta_0\|_1 + \mathbf{l}_{NT,\infty}\big)\big). \tag{D.12}$$

Proof of Lemma D.1. Define

$$\zeta^V_{NT} := \mathbf{d}_{NT,\infty}\sqrt{\log(dNT)/NT}, \qquad \zeta^B_{NT} = 0,$$

and the A-function as

$$A(W_{it}, \eta) = V_{it}\big(d_{i0}(X_{it}) - d_i(X_{it})\big).$$

Define $B_{Ak}(\eta)$ and $V_{Ak}(\eta)$ with $\eta = \mathbf{d}$ as in (A.8)–(A.9).

Consider any $\eta = \eta_{NT} \in D_{NT}$ in what follows. Since V_{it} obeys the martingale difference property by assumption, we have that

$$\mathrm{E}\Bigg[V_{it}\,\Bigg|\,\bigcup_{t' \le t, t' \in \mathcal{M}_k}(V_{it'}, X_{it'})\Bigg] = 0, \tag{D.13}$$

and it follows that $\|B_{Ak}(\eta_{NT})\|_\infty = 0$. By Assumption 4.3 and Lemma B.1, each entry of $V_{it}(d_{i0}(X_{it}) - d_i(X_{it}))$ is $\bar{\sigma}^2\mathbf{d}^2_{NT,\infty}$-sub-Gaussian. Invoking Lemma B.2 gives

$$\big\|V_{Ak}(\eta_{NT})\big\|_\infty \lesssim_P (\bar{\sigma}\mathbf{d}_{NT,\infty}\sqrt{\log d/NT_k}) = o_P\big(\zeta^V_{NT}\big)$$

since $T_k \asymp T$ (as we keep number of blocks K fixed). By Assumption 4.5, we have that $\mathrm{P}(\widehat{\mathbf{d}}_k \in D_{NT}, \forall k = 1, \ldots, K) \to 1$. We conclude by Lemma A.6 that (D.9) holds. Repeating the same argument for

$$A(W_{it}, \eta) = V_{it}\big(l_{i0}(X_{it}) - l_i(X_{it})\big) \quad \text{and} \quad A(W_{it}, \eta) = U_{it}\big(d_{i0}(X_{it}) - d_i(X_{it})\big)$$

establishes claims (D.10) and (D.11). Finally, (D.12) holds by definition of $\bar{e} = \bar{m} - \bar{a}'\beta_0$ and Holder inequalities. □

Lemma D.2 (Second-Order Term). *Under Assumptions* 4.1–4.5, *we have that*

$$\|\bar{z}\|_\infty \lesssim_P \big(\mathbf{d}_{NT}\mathbf{l}_{NT} + \mathbf{d}_{NT,\infty}\mathbf{l}_{NT,\infty}\sqrt{(NT)^{-1}\log(NT)\log d}\big), \tag{D.14}$$

$$\|\bar{b}\|_\infty \lesssim_P \big(\mathbf{d}^2_{NT} + \mathbf{d}^2_{NT,\infty}\sqrt{(NT)^{-1}\log(NT)\log d}\big), \tag{D.15}$$

$$\|\bar{g}\|_\infty \lesssim_P \big(\|\beta_0\|_1\mathbf{d}^2_{NT} + \mathbf{d}_{NT}\mathbf{l}_{NT} + \big(\|\beta_0\|_1\mathbf{d}^2_{NT,\infty} + \mathbf{d}_{NT,\infty}\mathbf{l}_{NT,\infty}\big)\sqrt{(NT)^{-1}\log(NT)\log d}\big). \tag{D.16}$$

Proof of Lemma D.2. Define the A-function as

$$A(W_{it}, \eta) = \big(d_{i0}(X_{it}) - d_i(X_{it})\big)\big(l_{i0}(X_{it}) - l_i(X_{it})\big), \quad \eta = (\mathbf{d}, \mathbf{l}).$$

Let $B_{Ak}(\eta)$ and $V_{Ak}(\eta)$ be defined according to (A.8)–(A.9). Let

$$\zeta^B_{NT} = \mathbf{l}_{NT}\mathbf{d}_{NT}, \qquad \zeta^V_{NT} = \sqrt{\mathbf{l}^2_{NT,\infty}\mathbf{d}^2_{NT,\infty}\log d\log NT/NT}.$$

For any i and t, m and j, the Cauchy–Schwarz inequality gives

$$\mathrm{E}\big[\big|\big(d_{i0}(X_{it})-d_i(X_{it})\big)_m\big(l_{i0}(X_{it})-l_i(X_{it})\big)\big|\big]$$
$$\leq \sqrt{\mathrm{E}\big(d_{i0}(X_{it})-d_i(X_{it})\big)_m^2\mathrm{E}\big(l_{i0}(X_{it})-l_i(X_{it})\big)^2} =: \sqrt{a_{it}^2 b_{it}^2} = |a_{it}||b_{it}|.$$

Another application of the Cauchy–Schwarz gives

$$(T_kN)^{-1}\sum_i\sum_{t\in\mathcal{M}_k}|a_{it}||b_{it}| \leq \sqrt{(T_kN)^{-1}\sum_{i=1}^N\sum_{t\in\mathcal{M}_k}a_{it}^2}\sqrt{(T_kN)^{-1}\sum_{i=1}^N\sum_{t\in\mathcal{M}_k}b_{it}^2}$$
$$\leq \sqrt{(T_kN)^{-1}\sum_{i=1}^N\sum_{t=1}^T a_{it}^2}\sqrt{(T_kN)^{-1}\sum_{i=1}^N\sum_{t=1}^T b_{it}^2} \leq \mathbf{d}_{NT}\mathbf{l}_{NT}T/T_k.$$

Therefore, $\|B_{Ak}(\eta_{NT})\|_\infty = O(\zeta_{NT}^B)$. Furthermore, each entry of $A(W_{it},\eta)$ is bounded by $\mathbf{d}_{NT,\infty}\mathbf{l}_{NT,\infty}$ and, therefore, is $\mathbf{d}_{NT,\infty}^2\mathbf{l}_{NT,\infty}^2$-sub-Gaussian. By Lemma B.4,

$$\big\|V_{Ak}(\eta_{NT})\big\|_\infty \lesssim_P (\zeta_{NT}^V),$$

since $T_k \asymp T$. Furthermore, by Assumption 4.5, $\mathrm{P}((\widehat{\mathbf{d}}_k,\widehat{\mathbf{l}}_k)\in D_{NT}\times L_{NT},\forall k = 1,\ldots,K)\to 1$. We conclude by Lemma A.6 that (D.14) holds. The bound (D.15) follows from the same argument. Finally, the bound (4.17) follows from the definition $\bar{g} = \bar{z} - \bar{b}'\beta_0$ and Holder inequality and union bounds. We obtain

$$\|\bar{g}\|_\infty \lesssim_P \Big(\|\beta_0\|_1\big(\mathbf{d}_{NT}^2 + \mathbf{d}_{NT,\infty}^2\sqrt{(NT)^{-1}\log(NT)\log d}\big)$$
$$+\mathbf{d}_{NT}\mathbf{l}_{NT} + \mathbf{d}_{NT,\infty}\mathbf{l}_{NT,\infty}\sqrt{(NT)^{-1}\log(NT)\log d}\Big).$$

Then we rewrite the bound as in (4.17). □

Define

$$\widehat{Q} = \mathbb{E}_{NT}\widehat{V}_{it}\widehat{V}_{it}', \qquad \widetilde{Q} = \mathbb{E}_{NT}V_{it}V_{it}', \qquad \widehat{S} := \mathbb{E}_{NT}\widehat{V}_{it}\big(\widehat{\widetilde{Y}}_{it} - \widehat{V}_{it}'\beta_0\big), \qquad S := \mathbb{E}_{NT}V_{it}U_{it}$$

and the following rates:

$$\kappa_{NT} := \sqrt{\log^3\big(d^2\log(NT)\big)\log(NT)/NT}, \tag{D.17}$$
$$q_{NT} := \mathbf{d}_{NT,\infty}\sqrt{\log(dNT)/NT} + \mathbf{d}_{NT}^2 + \mathbf{d}_{NT,\infty}^2\sqrt{\log(NT)\log(d)/NT}. \tag{D.18}$$

We will also use the following rates defined in the Section 4 of main text:

$$\rho_{NT} := \mathbf{d}_{NT,\infty}\sqrt{\log(dNT)/NT} + \sqrt{\log(dNT)/NT}\big(\mathbf{d}_{NT,\infty}\|\beta_0\|_1 + \mathbf{l}_{NT,\infty}\big) + r_{NT},$$
$$r_{NT} := \|\beta_0\|_1\mathbf{d}_{NT}^2 + \mathbf{d}_{NT}\mathbf{l}_{NT} + \big(\|\beta_0\|_1\mathbf{d}_{NT,\infty}^2 + \mathbf{l}_{NT,\infty}\big)\sqrt{(NT)^{-1}\log(NT)\log d}.$$

LEMMA D.3 (Summary of Gram Matrix and Gradient Error Bounds). *Suppose Assumptions* 4.1–4.5 *hold. Then the following bounds hold w.p.* $1-o(1)$:

$$\|\widetilde{Q}-Q\|_\infty \lesssim_P o\big(\kappa_{NT}\log\big(d^2NT\big)\big), \tag{D.19}$$

$$\|\widetilde{Q}-\widehat{Q}\|_\infty \lesssim_P (q_{NT}) = o_P\big((NT)^{-1/2}\big), \tag{D.20}$$

$$\|\widehat{Q}-Q\|_\infty \lesssim_P o\big(\kappa_{NT}\log\big(d^2NT\big)\big), \tag{D.21}$$

$$\|\widehat{S}-S\|_\infty \lesssim_P (\rho_{NT}) = o_P\big((NT)^{-1/2}\big). \tag{D.22}$$

PROOF OF LEMMA D.3. Decomposing matrix first-stage estimation error gives

$$\begin{aligned}
\widehat{Q} &= \mathbb{E}_{NT}\widehat{V}_{it}\widehat{V}_{it}' \\
&= \mathbb{E}_{NT}\big(V_{it}+\big(d_{i0}(X_{it})-\widehat{d}_i(X_{it})\big)\big)\big(V_{it}+\big(d_{i0}(X_{it})-\widehat{d}_i(X_{it})\big)\big)' \\
&= \mathbb{E}_{NT}V_{it}V_{it}' \\
&\quad + \mathbb{E}_{NT}V_{it}\big(d_{i0}(X_{it})-\widehat{d}_i(X_{it})\big)' + \big(\mathbb{E}_{NT}V_{it}\big(d_{i0}(X_{it})-\widehat{d}_i(X_{it})\big)'\big)' \\
&\quad + \mathbb{E}_{NT}\big(d_{i0}(X_{it})-\widehat{d}_i(X_{it})\big)\big(d_{i0}(X_{it})-\widehat{d}_i(X_{it})\big)' \\
&= \widetilde{Q}+\bar{a}+\bar{a}'+\bar{b}.
\end{aligned}$$

Then, an application of Lemma B.6 with $\bar{N}=2$ gives w.p. $1-o(1)$ $\|\widetilde{Q}-Q\|_\infty \le \bar{C}_\kappa\kappa_{NT}$ for large enough $\bar{C}_\kappa$. The bounds on $\|\bar{a}\|_\infty$ and $\|\bar{b}\|_\infty$ are given in (D.9) and (D.15), respectively. Collecting terms gives the bound (D.20). The (D.21) follows from the triangle inequality and $q_{NT}=o_P(\kappa_{NT})$. We can decompose the gradient error $\widehat{S}-S$ as follows. Note that

$$\begin{aligned}
\widehat{\widetilde{Y}}_{it}-\widetilde{Y}_{it} &= Y_{it}-\widehat{l}_i(X_{it})-\big(Y_{it}-l_{i0}(X_{it})\big) = l_{i0}(X_{it})-\widehat{l}_i(X_{it}), \\
\widehat{V}_{it}-V_{it} &= D_{it}-\widehat{d}_i(X_{it})-\big(D_{it}-d_{i0}(X_{it})\big) = d_{i0}(X_{it})-\widehat{d}_i(X_{it}).
\end{aligned}$$

The difference of the two equations is

$$\widehat{\widetilde{Y}}_{it}-\widetilde{Y}_{it}-(\widehat{V}_{it}-V_{it})'\beta_0 = R_{it}(\widehat{\mathbf{d}},\widehat{\mathbf{l}}).$$

Therefore,

$$\widehat{\widetilde{Y}}_{it}-\widehat{V}_{it}'\beta_0 = \big(\widetilde{Y}_{it}-V_{it}'\beta_0\big)+\big((\widehat{\widetilde{Y}}_{it}-\widetilde{Y}_{it})-(\widehat{V}_{it}-V_{it})'\beta_0\big) = U_{it}+R_{it}(\widehat{\mathbf{d}},\widehat{\mathbf{l}}). \tag{D.23}$$

Decompose the gradient:

$$\widehat{S} = \mathbb{E}_{NT}\widehat{V}_{it}\big(\widehat{\widetilde{Y}}_{it}-\widehat{V}_{it}'\beta_0\big) = \mathbb{E}_{NT}V_{it}\big(\widehat{\widetilde{Y}}_{it}-\widehat{V}_{it}'\beta_0\big)+\mathbb{E}_{NT}(\widehat{V}_{it}-V_{it})\big(\widehat{\widetilde{Y}}_{it}-\widehat{V}_{it}'\beta_0\big) = \widehat{S}_1+\widehat{S}_2,$$

where

$$\begin{aligned}
\widehat{S}_1 &= \mathbb{E}_{NT}V_{it}U_{it}+\mathbb{E}_{NT}V_{it}R_{it}(\widehat{\mathbf{d}},\widehat{\mathbf{l}}) = S+\bar{e}, \\
\widehat{S}_2 &= \mathbb{E}_{NT}(d_{i0}\big(X_{it}-\widehat{d}_i(X_{it})\big)U_{it}+\mathbb{E}_{NT}\big(d_{i0}(X_{it})-\widehat{d}_i(X_{it})\big)R_{it}(\widehat{\mathbf{d}},\widehat{\mathbf{l}}) = \bar{f}+\bar{g}.
\end{aligned}$$

Invoking bounds on $\bar{e}$, $\bar{f}$, and $\bar{g}$ in (D.12)–(D.16) gives the result. □

D.2 *Proof of orthogonal lasso rate: Theorem 4.1*

D.2.1 *Group sparsity notation* We use the same notation as Lounici, Pontil, van de Geer, and Tsybakov (2011). Consider a generic covariate vector of size $g \cdot d$, where d is the number of groups and g is the group size. Partition the set of indices $\{1, 2, \ldots, gd\}$ into d groups of size g:

$$J_j := \{j, d+j, \ldots, (g-1)d + j\}, \quad j = 1, 2, \ldots, d, \, |J_j| = g.$$

For a group index j and a subset of group indices $\mathcal{T}$, and vector $\Delta \in \mathrm{R}^{gd}$, denote

$$\Delta^j = (\Delta_m)_{m \in J_j} \in \mathrm{R}^g, \qquad \Delta^{\mathcal{T}} = (\Delta_m)_{\{m \in J_j, j \in \mathcal{T}\}} \in \mathrm{R}^{|\mathcal{T}| \cdot g}.$$

For any $\Delta \in \mathrm{R}^{gd}$, define the group-vector norms

$$\|\Delta\|_{2,\infty} = \max_{1 \le j \le d} \left\|\Delta^j\right\|_2, \qquad \|\Delta\|_{2,1} = \sum_{j=1}^{d} \left\|\Delta^j\right\|_2.$$

For a symmetric matrix M, define

$$\|M\|_{2,\infty} = \left\|M'\right\|_{2,\infty} = \max_{1 \le i \le dg} \max_{1 \le j \le d} \left(\sum_{k \in J_j} M_{i,k}^2\right)^{1/2}.$$

Define the group restricted cone as

$$\mathsf{REG}(\bar{c}) := \left\{\Delta \in \mathrm{R}^{gd} : \sum_{j \in \mathcal{T}^c} \left\|\Delta^j\right\|_2 \le \bar{c} \sum_{j \in \mathcal{T}} \left\|\Delta^j\right\|_2, \, \Delta \ne 0\right\}.$$

Given a matrix $M \in \mathrm{R}^{gd} \times \mathrm{R}^{gd}$, define the restricted group-sparse eigenvalue

$$\kappa_g(M, \mathcal{T}, \bar{c}) = \min_{\Delta \in \mathsf{REG}(\bar{c})} \frac{\sqrt{s}\left(\Delta' M \Delta\right)^{1/2}}{\left\|\Delta^{\mathcal{T}}\right\|_{2,1}}.$$

When the group size g is equal to 1, the objects above reduce to the following quantities:

$$\Delta^j = \Delta_j, \qquad \Delta^{\mathcal{T}} = \Delta_{\mathcal{T}} = (\Delta_m)_{\{m \in \mathcal{T}\}}, \qquad \|M\|_{2,\infty} = \|M\|_\infty,$$

the group restricted cone is regular restricted cone

$$\mathsf{REG}(\bar{c}) = \mathsf{RE}(\bar{c}) = \left\{\Delta \in \mathrm{R}^d : \|\Delta_{\mathcal{T}^c}\|_1 \le \bar{c}\|\Delta_{\mathcal{T}}\|_1, \, \Delta \ne 0\right\},$$

and the restricted group-sparse eigenvalue reduces to restricted eigenvalue

$$\kappa_1(M, \mathcal{T}, \bar{c}) = \kappa(M, \mathcal{T}, \bar{c}) = \min_{\Delta \in \mathsf{RE}(\bar{c})} \frac{\sqrt{s}\left(\Delta' M \Delta\right)^{1/2}}{\|\Delta_{\mathcal{T}}\|_1}.$$

Let $\bar{X}_{it} \in \mathrm{R}^{gd}$ be a generic covariate (dg)-vector and $\bar{Y}_{it}$ be a generic outcome. Given a parameter $\bar{\beta}_0$, decompose

$$\bar{Y}_{it} = \bar{X}_{it}'\bar{\beta}_0 + U_{it}.$$

The least squares loss function is

$$\mathcal{Q}(\bar{\beta}) := 1/2(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\big(\bar{Y}_{it} - \bar{X}_{it}'\bar{\beta}\big)^2.$$

The group lasso estimator is

$$\widehat{\bar{\beta}} := \arg\min_{\bar{\beta}} \mathcal{Q}(\bar{\beta}) + \lambda\|\bar{\beta}\|_{2,1}. \tag{D.24}$$

The least squares gradient is

$$\mathcal{S}(\bar{\beta}_0) := \nabla_{\bar{\beta}_0}\mathcal{Q}(\bar{\beta}_0) = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\big(\bar{Y}_{it} - \bar{X}_{it}'\bar{\beta}_0\big)\bar{X}_{it},$$

and the Hessian is

$$\mathcal{H}(\bar{\beta}_0) := (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{X}_{it}\bar{X}_{it}'.$$

Lemma D.4 (Grouped Norm Inequalities). *For any two vectors $a, b \in \mathrm{R}^{gd}$, and matrix $M \in \mathrm{R}^{gd} \times \mathrm{R}^{gd}$, the following inequalities hold*:

$$\big|a'b\big| \le \|a\|_{2,1}\|b\|_{2,\infty}, \tag{D.25}$$

$$\big|v'Mv\big| \le \sqrt{g}\|v\|_{2,1}^2 \cdot \|M\|_{2,\infty}, \tag{D.26}$$

$$\|M\|_{2,\infty} \le \|M\|_\infty\sqrt{g}. \tag{D.27}$$

Proof. For each group $j = 1, 2, \dots, d$, Cauchy inequality gives

$$\left|\sum_{k\in J_j} a_k b_k\right| \le \left(\sum_{k\in J_j} a_k^2\right)^{1/2}\left(\sum_{k\in J_j} b_k^2\right)^{1/2} \le \max_{1\le j\le d}\left(\sum_{k\in J_j} a_k^2\right)^{1/2}\|b^j\|_2 = \Big(\max_{1\le j\le d}\|a^j\|_2\Big)\|b^j\|_2,$$

which implies

$$\big|a'b\big| \le \sum_{j=1}^{d}\left|\sum_{k\in J_j} a_k b_k\right| \le \Big(\max_{1\le j\le d}\|a^j\|_2\Big)\sum_{j=1}^{d}\|b^j\|_2 = \|a\|_{2,\infty}\|b\|_{2,1}.$$

For each index i, $1 \le i \le kg$, the following bound holds:

$$\left|\sum_{k=1}^{gd} M_{i,k}v_k\right| \le \sum_{j=1}^{d}\left|\sum_{k\in J_j} M_{i,k}v_k\right| \le \sum_{j=1}^{d}\left(\sum_{k\in J_j} M_{i,k}^2\right)^{1/2}\left(\sum_{k\in J_j} v_k^2\right)^{1/2}$$

$$\leq \max_{1\leq j\leq d}\left(\sum_{k\in J_j} M_{i,k}^2\right)^{1/2}\sum_{j=1}^{d}\left(\sum_{k\in J_j} v_k^2\right)^{1/2} \leq \|M\|_{2,\infty}\|v\|_{2,1}.$$

Then

$$\|Mv\|_{2,\infty} = \max_{1\leq j\leq d}\left\|(Mv)^j\right\|_2 = \max_{1\leq j\leq d}\left(\sum_{i\in J_j}|Mv|_i^2\right)^{1/2}$$

$$\leq \max_{1\leq j\leq d}\left(\sum_{i\in J_j}\|v\|_{2,1}^2\|M\|_{2,\infty}^2\right)^{1/2} \leq \sqrt{g}\|v\|_{2,1}\|M\|_{2,\infty}.$$

Therefore, we obtain (D.26) by combining inequalities above:

$$\left|v'Mv\right| \leq \|v\|_{2,1}\|Mv\|_{2,\infty} \leq \sqrt{g}\|M\|_{2,\infty}\cdot\|v\|_{2,1}^2.$$

Finally, the bound (D.27) follows from

$$M_{2,\infty} = \max_{1\leq j\leq d}\left\|M^j\right\|_2 \leq \max_{1\leq j\leq d}\sqrt{g}\left\|M^j\right\|_\infty$$

using the fact that $\|v\|_2 \leq \sqrt{\dim(v)}\|v\|_\infty$. □

LEMMA D.5 (First-Stage Effect on the Curvature). *Let* $\mathbf{M}_1, \mathbf{M}_2 \in \mathrm{R}^{gd\times gd}$ *be two matrices. Let* $\lambda_{NT}^M := \|\mathbf{M}_1 - \mathbf{M}_2\|_\infty$. *On the event* $\kappa_g^2(\mathbf{M}_2, \mathcal{T}, \bar{c}) > 0$, *for any* $\Delta \in \mathsf{REG}(\bar{c})$,

$$\left|\kappa_g^2(\mathbf{M}_2, \mathcal{T}, \bar{c}) - \kappa_g^2(\mathbf{M}_1, \mathcal{T}, \bar{c})\right| \leq \lambda_{NT}^M(1+\bar{c})^2 sg. \tag{D.28}$$

PROOF OF LEMMA D.5. For any $\Delta \in \mathrm{R}^{gd}$, the difference can be bounded as

$$\left|\Delta'(\mathbf{M}_1 - \mathbf{M}_2)\Delta\right| \leq^{\mathrm{i}} \sqrt{g}\|\mathbf{M}_1 - \mathbf{M}_2\|_{2,\infty}\|\Delta\|_{2,1}^2 \leq^{\mathrm{ii}} g\lambda_{NT}^M\|\Delta\|_{2,1}^2, \tag{D.29}$$

where (i) follows from (D.26) and (ii) from (D.27). For any $\Delta \in \mathsf{REG}(\bar{c})$,

$$\|\Delta\|_{2,1}^2 \leq (1+\bar{c})^2\left\|\Delta^{\mathcal{T}}\right\|_{2,1}^2 \leq \frac{(1+\bar{c})^2 s}{\kappa_g^2(\mathbf{M}_2, \mathcal{T}, \bar{c})}\Delta'\mathbf{M}_2\Delta =: \gamma\cdot\Delta'\mathbf{M}_2\Delta. \tag{D.30}$$

Combining (D.29) and (D.30) give

$$\left|\Delta'(\mathbf{M}_1 - \mathbf{M}_2)\Delta\right| \leq \left(g\lambda_{NT}^M\gamma\right)\cdot\Delta'\mathbf{M}_2\Delta. \tag{D.31}$$

Noting that $x \leq |x|$ gives

$$\Delta'(\mathbf{M}_1 - \mathbf{M}_2)\Delta \leq \left|\Delta'(\mathbf{M}_1 - \mathbf{M}_2)\Delta\right| \leq \left(g\lambda_{NT}^M\gamma\right)\cdot\Delta'\mathbf{M}_2\Delta,$$

which implies

$$\Delta'\mathbf{M}_1\Delta \leq \Delta'\mathbf{M}_2\Delta\left(1 + g\lambda_{NT}^M\gamma\right). \tag{D.32}$$

Noting that $-x \leq |x|$ gives

$$\Delta'(\mathbf{M}_2 - \mathbf{M}_1)\Delta \leq \big(g\lambda^M_{NT}\gamma\big) \cdot \big(\Delta'\mathbf{M}_2\Delta\big) \tag{D.33}$$

which implies

$$\Delta'\mathbf{M}_1\Delta \geq \Delta'\mathbf{M}_2\Delta \cdot \big(1 - g\lambda^M_{NT}\gamma\big). \tag{D.34}$$

Rearranging (D.32) gives an upper bound on $\kappa_g(\mathbf{M}_1, \mathcal{T}, \bar{c})$:

$$\begin{aligned}
\kappa_g(\mathbf{M}_1, \mathcal{T}, \bar{c}) &:= \min_{\Delta \in \mathsf{REG}(\bar{c})} \frac{\sqrt{s}\big(\Delta'\mathbf{M}_1\Delta\big)^{1/2}}{\|\Delta^{\mathcal{T}}\|_{2,1}} \\
&\leq \min_{\Delta \in \mathsf{REG}(\bar{c})} \frac{\sqrt{s}\big(\Delta'\mathbf{M}_2\Delta\big)^{1/2}}{\|\Delta^{\mathcal{T}}\|_{2,1}} \sqrt{1 + g\lambda^M_{NT}\gamma} \\
&= \kappa_g(\mathbf{M}_2, \mathcal{T}, \bar{c})\sqrt{1 + g\lambda^M_{NT}\gamma}.
\end{aligned}$$

A lower bound on $\kappa_g(\mathbf{M}_1, \mathcal{T}, \bar{c})$ follows analogously, that is,

$$\begin{aligned}
\kappa_g(\mathbf{M}_1, \mathcal{T}, \bar{c}) &\geq \min_{\Delta \in \mathsf{REG}(\bar{c})} \frac{\sqrt{s}\big(\Delta'\mathbf{M}_2\Delta\big)^{1/2}}{\|\Delta^{\mathcal{T}}\|_{2,1}} \sqrt{1 - g\lambda^M_{NT}\gamma} \\
&= \kappa_g(\mathbf{M}_2, \mathcal{T}, \bar{c})\sqrt{1 - g\lambda^M_{NT}\gamma}.
\end{aligned}$$

Taking the squares of both sides of the inequality and rearranging gives (D.28). □

LEMMA D.6 (Oracle Inequality for Group Lasso). *On the event $\mathcal{G}_1 := \{\lambda \geq c\sqrt{g}\|\mathcal{S}(\bar{\beta}_0)\|_\infty\}$, the error vector $\Delta = \widehat{\bar{\beta}} - \bar{\beta}_0$ belongs to the restricted set:*

$$\Delta \in \mathsf{REG}(\bar{c})$$

and obeys the bound

$$\big(\Delta'\mathcal{H}(\bar{\beta}_0)\Delta\big) \leq 2\lambda\bar{c}\|\Delta^{\mathcal{T}}\|_{2,1}, \tag{D.35}$$

where $\bar{c} := (c+1)/(c-1)$.

PROOF OF LEMMA D.6. Assume the event $\mathcal{G}_1$ holds throughout, which implies $\lambda \geq c\|\mathcal{S}(\bar{\beta}_0)\|_{2,\infty}$. Negahban, Ravikumar, Wainwright, and Yu (2012) establishes

$$\|\bar{\beta}_0\|_{2,1} - \|\widehat{\bar{\beta}}\|_{2,1} \leq \|\Delta^{\mathcal{T}}\|_{2,1} - \|\Delta^{\mathcal{T}^c}\|_{2,1}, \tag{D.36}$$

and shows that $\Delta \in \mathsf{REG}(\bar{c})$, which implies

$$\|\Delta\|_{2,1} \leq (1+\bar{c})\|\Delta^{\mathcal{T}}\|_{2,1}. \tag{D.37}$$

Note that $\widehat{\bar{\beta}}$ solves group lasso minimization problem (D.24), so that

$$\mathcal{Q}(\widehat{\bar{\beta}}) + \lambda\|\widehat{\bar{\beta}}\|_{2,1} \leq \mathcal{Q}(\bar{\beta}_0) + \lambda\|\bar{\beta}_0\|_{2,1}.$$

Expanding the least squares criterion gives

$$\mathcal{Q}(\widehat{\bar{\beta}}) - \mathcal{Q}(\bar{\beta}_0) = \mathcal{S}(\bar{\beta}_0)'\Delta + 1/2\big(\Delta'\mathcal{H}(\bar{\beta}_0)\Delta\big) \leq \lambda\big(\|\bar{\beta}_0\|_{2,1} - \|\widehat{\bar{\beta}}\|_{2,1}\big).$$

Invoking inequality (D.25) for $\mathcal{S}(\bar{\beta}_0)'\Delta$ gives

$$1/2\big(\Delta'\mathcal{H}(\bar{\beta}_0)\Delta\big) \leq \lambda\big(\|\bar{\beta}_0\|_{2,1} - \|\widehat{\bar{\beta}}\|_{2,1}\big) + \big\|\mathcal{S}(\bar{\beta}_0)\big\|_{2,\infty}\|\Delta\|_{2,1}.$$

Then

$$\begin{aligned}
1/2\big(\Delta'\mathcal{H}(\bar{\beta}_0)\Delta\big) &\leq^{\text{i}} \lambda\big(\big\|\Delta^{\mathcal{T}}\big\|_{2,1} - \big\|\Delta^{\mathcal{T}^c}\big\|_{2,1}\big) + \lambda/c\|\Delta\|_{2,1} \\
&\leq \lambda\big\|\Delta^{\mathcal{T}}\big\|_{2,1} + 0 + \lambda/c\|\Delta\|_{2,1} \\
&\leq^{\text{ii}} \lambda\big\|\Delta^{\mathcal{T}}\big\|_{2,1} + (\lambda/c)(1+\bar{c})\big\|\Delta^{\mathcal{T}}\big\|_{2,1} \\
&=^{\text{iii}} \lambda\bar{c}\big\|\Delta^{\mathcal{T}}\big\|_{2,1},
\end{aligned} \tag{D.38}$$

where (i) follows from (D.36), (ii) from (D.37), and (iii) from

$$1 + c^{-1}(\bar{c}+1) = \big(c + (c+1)/(c-1)\big)/c = (c+1)/(c-1) = \bar{c}.$$

Since $\Delta \in \mathsf{REG}(\bar{c})$, (D.35) follows. □

Proof of Theorem 4.1. We invoke Lemma D.6 with the group size $g = 1$, $\bar{\beta}_0 = \beta_0$, and $\bar{U}_{it} = U_{it} + R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}})$. The gradient $\mathcal{S}(\beta_0) = \widehat{S}$, the Hessian is $\mathcal{H}(\beta_0) = \widehat{Q}$, and the penalty $\lambda = \lambda_\beta$. Note that $\delta \in \mathsf{RE}(\bar{c})$ has been established in the proof of Lemma D.6.

Step 1. Union bound implies

$$\begin{aligned}
\mathrm{P}\big(\lambda_\beta \leq c\sqrt{g}\|\widehat{S}\|_\infty\big) &\leq \mathrm{P}\big(\lambda_\beta/2 \leq c\sqrt{g}\|S\|_\infty\big) + \mathrm{P}\big(\lambda_\beta/2 \leq c\sqrt{g}\|\widehat{S} - S\|_\infty\big) \\
&= P_S + P_{\widehat{S}-S} \leq o(1) + o(1),
\end{aligned}$$

where $P_S \leq 2/d = o(1)$ is given in (B.6) and $P_{\widehat{S}-S} = o(1)$ since

$$\|\widehat{S} - S\|_\infty \lesssim_P (\rho_{NT}) = o_P(\sqrt{\log d/NT}).$$

Step 2. Let $\mathbf{M}_2 := Q = (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \mathrm{E}V_{it}V_{it}'$ and $\mathbf{M}_1 := \widetilde{Q} = \mathbb{E}_{NT}V_{it}V_{it}'$. Observe that

$$\kappa^2(Q, \mathcal{T}, \bar{c}) = \min_{\delta\in\mathsf{RE}(\bar{c})} \frac{s\delta' Q\delta}{\|\delta_{\mathcal{T}}\|_1^2} \geq \min_{\delta\in\mathsf{RE}(\bar{c})} \frac{s\min\operatorname{eig}(Q)\|\delta\|_2^2}{\|\delta_{\mathcal{T}}\|_1^2} \geq^{\text{i}} \min\operatorname{eig}(Q), \tag{D.39}$$

where (i) follows from

$$s\|\delta\|_2^2 \geq s\|\delta_{\mathcal{T}}\|_2^2 \geq \|\delta_{\mathcal{T}}\|_1^2 \quad \forall\delta \in \mathrm{R}^d.$$

The bounds (D.28) and (D.19) imply

$$\left|\kappa^2(\widetilde{Q}, \mathcal{T}, \bar{c}) - \kappa^2(Q, \mathcal{T}, \bar{c})\right| \leq s\|\widetilde{Q} - Q\|_\infty (1+\bar{c})^2 \lesssim_P (s\kappa_{NT}).$$

Therefore, the event $\mathcal{G}_2 := \{\kappa^2(\widetilde{Q}, \mathcal{T}, \bar{c}) > C_{\min}/2\}$ holds w.p. $1 - o(1)$.

Step 3. Invoke Lemma D.5 on the event $\mathcal{G}_2$ with $\mathbf{M}_2 := \widetilde{Q}$ and $\mathbf{M}_1 := \widehat{Q}$. The bound (D.28) gives

$$\left|\kappa^2(\widehat{Q}, \mathcal{T}, \bar{c}) - \kappa^2(\widetilde{Q}, \mathcal{T}, \bar{c})\right| \leq s\|\widehat{Q} - \widehat{Q}\|_\infty (1+\bar{c})^2 \lesssim_P (sq_{NT}),$$

which implies

$$\left|\kappa^2(\widehat{Q}, \mathcal{T}, \bar{c}) - \kappa^2(Q, \mathcal{T}, \bar{c})\right| \lesssim_P \left(s(q_{NT} + \kappa_{NT})\right).$$

Therefore, the event $\{\kappa^2(\widehat{Q}, \mathcal{T}, \bar{c}) > C_{\min}/2\}$ holds w.p. $1 - o(1)$. Thus, the event

$$\mathcal{G}_3 := s\|\widehat{Q} - \widetilde{Q}\|_\infty (1+\bar{c})^2 / \kappa^2(\widehat{Q}, \mathcal{T}, \bar{c}) < 1/2$$

is well-defined and holds w.p. $1 - o(1)$.

Step 4. On the event $\mathcal{G}_1 \cap \mathcal{G}_2 \cap \mathcal{G}_3$, invoking (D.34) with $\mathbf{M}_2 = \widetilde{Q}$ and $\mathbf{M}_1 = \widehat{Q}$ gives

$$\delta' \widehat{Q} \delta \geq (1/2) \cdot \delta' \widetilde{Q} \delta.$$

Combining inequality above with (D.35) gives

$$\delta' \widetilde{Q} \delta \leq 2\delta' \widehat{Q} \delta \leq 4\lambda_\beta \bar{c} \|\delta_{\mathcal{T}}\|_1 \leq \sqrt{s}\lambda_\beta \frac{4\bar{c}\left(\delta' \widetilde{Q} \delta\right)^{1/2}}{\kappa(\widetilde{Q}, \mathcal{T}, \bar{c})}.$$

Dividing L.H.S. and R.H.S. by $(\delta' \widetilde{Q} \delta)^{1/2}$ give

$$\left(\delta' \widetilde{Q} \delta\right)^{1/2} \leq \sqrt{s}\lambda_\beta \frac{4\bar{c}}{\kappa(\widetilde{Q}, \mathcal{T}, \bar{c})} \lesssim_P (\sqrt{s}\lambda_\beta)$$

and

$$\|\delta\|_1 \leq (1+\bar{c})\|\delta_{\mathcal{T}}\|_1 \leq (1+\bar{c}) \frac{\sqrt{s}\left(\delta' \widetilde{Q} \delta\right)^{1/2}}{\kappa(\widetilde{Q}, \mathcal{T}, \bar{c})} \leq 4(1+\bar{c}) \frac{s\lambda_\beta \bar{c}}{\kappa^2(\widetilde{Q}, \mathcal{T}, \bar{c})}. \qquad \square$$

D.3 *Proof of Theorem 4.2*

In what follows, we use the notation $Q^{-1} = (\omega^0_{ij})$ and $Q^{-1}_{\cdot,j} := \omega^0_j$. Define the following quantities:

$$s_j(\lambda) := \left\|1\left\{\left|\omega^0_j\right| \geq \lambda\right\}\right\|_1, \qquad r_j(\lambda) := \left\|\left(\omega^0_j\right) 1\left\{\left|\omega^0_j\right| \leq \lambda\right\}\right\|_1.$$

Remark D.1. Assumption 4.6 implies the following bounds:

$$\left\|Q^{-1}\right\|_{1,\infty} = \max_{1 \leq j \leq d} \left\|\omega^0_j\right\|_1 \leq A_Q \sum_{j=1}^{p} j^{-a_Q} \leq A_Q \int_1^\infty j^{-a_Q}\, dj \leq A_Q/(a_Q - 1). \tag{D.40}$$

Furthermore, if $A_Q j^{-a_Q} \le \lambda$, then $j \ge j_Q^* := (A_Q/\lambda)^{1/a_Q}$. This implies

$$s_j(\lambda) := \big\| 1\{|\omega_j^0| \ge \lambda\} \big\|_1 \le \big\| 1\{A_Q j^{-a_Q} \ge \lambda\} \big\|_1 \le \sum_{j=1}^{j_Q^*} 1 = j_Q^* = (A_Q/\lambda)^{1/a_Q},$$

$$r_j(\lambda) \le \int_{j_Q^*}^{\infty} A_Q j^{-a_Q} \, dj = A_Q \frac{(j_Q^*)^{1-a_Q}}{a_Q - 1} = A_Q \frac{(A_Q/\lambda)^{(1-a_Q)/a_Q}}{a_Q - 1} = \frac{A_Q^{1/a_Q}}{(a_Q - 1)} \lambda^{1-1/a_Q}.$$

PROOF OF LEMMA 4.2. *Step 0.* Suppose Assumptions 4.1–4.6 hold. We claim that the event

$$\mathcal{G}_Q := \big\{ \|\widehat{Q} - Q\|_\infty \big\| Q^{-1} \big\|_{1,\infty} \le \lambda_Q \big\}, \tag{D.41}$$

holds w.p. $1 - o(1)$. On this event $\mathcal{G}_Q$, by definition of $\widehat{\Omega}$, we have

$$\|\widehat{\Omega}\|_{1,\infty} \le \big\| Q^{-1} \big\|_{1,\infty}, \tag{D.42}$$

and, therefore,

$$\big\| \widehat{\Omega}^{\text{CLIME}} \big\|_{1,\infty} \le \|\widehat{\Omega}\|_{1,\infty} \le \big\| Q^{-1} \big\|_{1,\infty}. \tag{D.43}$$

To show that $\mathrm{P}(\mathcal{G}_Q) = 1 - o(1)$, decompose

$$\widehat{Q} Q^{-1} - I_d = \widehat{Q} Q^{-1} - Q Q^{-1} = (\widehat{Q} - Q) Q^{-1}.$$

By Lemma D.3 for some $\bar{C}_\kappa > 0$, w.p. $1 - o(1)$,

$$\|\widehat{Q} - Q\|_\infty \le \bar{C}_\kappa \kappa_{NT}.$$

Therefore, w.p. $1 - o(1)$,

$$\big\| \widehat{Q} Q^{-1} - I_d \big\|_\infty \le \|\widehat{Q} - Q\|_\infty \big\| Q^{-1} \big\|_{1,\infty} \le \bar{C}_\kappa 2\kappa_{NT} \big\| Q^{-1} \big\|_{1,\infty} \le \lambda_Q \tag{D.44}$$

as long as $C_Q \ge 2\bar{C}_\kappa \|Q^{-1}\|_{1,\infty}$. Since $\|Q^{-1}\|_{1,\infty} \le A_Q/(a_Q - 1)$, $C_Q \ge 2\bar{C}_\kappa \|Q^{-1}\|_{1,\infty}$ holds by Assumption 4.6.

Step 1. We establish (4.10). Specifically, we show that, on the event $\mathcal{G}_Q$, we have

$$\big\| \widehat{\Omega}^{\text{CLIME}} - Q^{-1} \big\|_\infty \le \big\| \widehat{\Omega} - Q^{-1} \big\|_\infty \le \frac{4A_Q}{(a_Q - 1)} \lambda_Q.$$

The argument repeats the proof of equation (13) in Cai, Liu, and Luo (2011, Theorem 6). On the event $\mathcal{G}_Q$, the bound holds

$$\begin{aligned}
\|Q\widehat{\Omega} - I_d\|_\infty &= \big\| Q(\widehat{\Omega} - Q^{-1}) \big\|_\infty \\
&\le \big\| (Q - \widehat{Q})(\widehat{\Omega} - Q^{-1}) \big\|_\infty + \big\| \widehat{Q}(\widehat{\Omega} - Q^{-1}) \big\|_\infty \\
&\le \|Q - \widehat{Q}\|_\infty \big\| \widehat{\Omega} - Q^{-1} \big\|_{1,\infty} + \|\widehat{Q}\widehat{\Omega} - I_d\|_\infty + \big\| I_d - \widehat{Q} Q^{-1} \big\|_\infty \\
&\le \|Q - \widehat{Q}\|_\infty \big(\big\| Q^{-1} \big\|_{1,\infty} + \|\widehat{\Omega}\|_{1,\infty} \big) + \lambda_Q + \big\| \widehat{Q} Q^{-1} - I_d \big\|_\infty.
\end{aligned}$$

Invoking (D.42) and (D.44) give

$$\|Q\widehat{\Omega} - I_d\|_\infty \le 2\|Q - \widehat{Q}\|_\infty \|Q^{-1}\|_{1,\infty} + \lambda_Q + \lambda_Q \le 2\lambda_Q + 2\lambda_Q = 4\lambda_Q. \tag{D.45}$$

Premultiplying $Q\widehat{\Omega} - I_d$ by Q^{-1} and invoking (D.40) give

$$\|\widehat{\Omega} - Q^{-1}\|_\infty = \|Q^{-1}(Q\widehat{\Omega} - I_d)\|_\infty \le \|Q^{-1}\|_{\infty,1}\|Q\widehat{\Omega} - I_d\|_\infty \le 4\frac{A_Q}{a_Q - 1}\lambda_Q.$$

Since Q is a symmetric matrix, so is Q^{-1}, and

$$|\widehat{\Omega}^{\text{CLIME}}_{mj} - Q^{-1}_{mj}| \le \max(|\widehat{\Omega}_{mj} - Q^{-1}_{mj}|, |\widehat{\Omega}_{jm} - Q^{-1}_{jm}|) \le \|\widehat{\Omega} - Q^{-1}\|_\infty,$$

which implies (4.10).

Step 2. We show that (4.11) holds. Specifically, we show that on the event $\mathcal{G}_Q$ we have that

$$\|\widehat{\Omega}^{\text{CLIME}} - Q^{-1}\|_{1,\infty} \le \bar{C}_Q \lambda_Q{}^{1-1/a_Q},$$

for some constant $\bar{C}_Q$ that depends on Q. We closely follow the proof of (14), page 605 in Cai, Liu, and Luo (2011). Using their notation, let

$$t_n := \|\widehat{\Omega}^{\text{CLIME}} - Q^{-1}\|_\infty, \qquad \omega^0_j := Q^{-1}_{\cdot,j},$$
$$h_j := \widehat{\Omega}^{\text{CLIME}}_{\cdot,j} - \omega^0_j, \qquad h^1_j := \big(\widehat{\omega}_{ij} 1\{|\widehat{\omega}_{ij}| \ge 2t_n\}\big)^p_{i=1} - \omega^0_j, \qquad h^2_j := h_j - h^1_j.$$

By definition of CLIME, on the event $\mathcal{G}_G$,

$$\begin{aligned}\|\omega^0_j\|_1 - \|h^1_j\|_1 + \|h^2_j\|_1 &\le \|h^1_j + \omega^0_j\|_1 + \|h^2_j\|_1 \\ &=^{\text{i}} \|h^2_j + h^1_j + \omega^0_j\|_1 = \|\widehat{\Omega}^{\text{CLIME}}_{\cdot,j}\|_1 \le \|\widehat{\Omega}_{\cdot,j}\|_1 \le \|\omega^0_j\|_1,\end{aligned}$$

where (i) follows from $h^1_j + \omega^0_j$ and h^2_j having nonoverlapping support. This implies

$$\|h_j - h^1_j\|_1 := \|h^2_j\|_1 \le \|h^1_j\|_1, \quad \|h_j\|_1 \le 2\|h^1_j\|_1.$$

Then the following bound holds:

$$\begin{aligned}\|h^1_j\|_1 &= \sum_{i=1}^d |\widehat{\omega}_{ij} 1\{|\widehat{\omega}_{ij}| \ge 2t_n\} - \omega^0_{ij}| \\ &\le \sum_{i=1}^d |\omega^0_{ij}| 1\{\omega^0_{ij} \le 2t_n\} + \sum_{i=1}^d |\widehat{\omega}_{ij} 1\{|\widehat{\omega}_{ij}| \ge 2t_n\} - \omega^0_{ij} 1\{|\omega^0_{ij}| \ge 2t_n\}| \\ &\le r_j(2t_n) + t_n \sum_{i=1}^d 1\{|\widehat{\omega}_{ij}| \ge 2t_n\} + \sum_{i=1}^d |\omega^0_{ij}| |(1\{|\widehat{\omega}_{ij}| \ge 2t_n\} - 1\{|\omega^0_{ij}| \ge 2t_n\})| \\ &\le r_j(2t_n) + t_n \sum_{i=1}^d 1\{|\omega^0_{ij}| \ge t_n\} + \sum_{i=1}^d |\omega^0_{ij}| I\{||\omega^0_{ij}| - 2t_n| \le |\hat{\omega}_{ij} - \omega^0_{ij}|\}\end{aligned}$$

$$
\begin{aligned}
&\leq r_j(2t_n) + t_n s_j(t_n) + \sum_{i=1}^{d} |\omega_{ij}^0| 1\{\omega_{ij}^0 \leq 3t_n\} \\
&\leq r_j(2t_n) + t_n s_j(t_n) + r_j(3t_n) \\
&\leq C_Q' t_n^{1-1/a_Q}, \quad \left(C_Q' := \frac{A_Q^{1/a_Q}}{(a_Q - 1)} (2^{1-1/a_Q} + (a_Q - 1) + 3^{1-1/a_Q}) \right).
\end{aligned}
$$

Since $t_n \leq \|\widehat{\Omega} - Q^{-1}\|_\infty$ from Step 1, we have

$$
\begin{aligned}
\left\| (\widehat{\Omega}^{\text{CLIME}} - Q^{-1})' \right\|_{1,\infty} = \left\| \widehat{\Omega}^{\text{CLIME}} - Q^{-1} \right\|_{1,\infty} &:= \max_{1 \leq j \leq d} \|h_j\|_1 \leq C_Q' \left(\left\| \widehat{\Omega} - Q^{-1} \right\|_\infty \right)^{1-1/a_Q} \\
&\leq \bar{C}_Q {\lambda_Q}^{1-1/a_Q},
\end{aligned}
$$

where $\bar{C}_Q = C_Q'(4A_Q/(a_Q - 1))^{1-1/a_Q}$ is a constant that depends on Q. Thus, (4.11) follows.

Step 3. We show (4.12). Specifically, we show that on the event $\mathcal{G}_Q$ and $\|\widehat{Q} - Q\| \leq 1$ and once $\lambda_Q \leq 1$, we have that

$$
\left\| I_d - \widehat{\Omega}^{\text{CLIME}} \widehat{Q} \right\|_\infty = \left\| I_d - \widehat{Q} \widehat{\Omega}^{\text{CLIME}} \right\|_\infty \leq C_Q' {\lambda_Q}^{1-1/a_Q},
$$

for some constant C_Q' that depends only on Q. Indeed,

$$
\begin{aligned}
\left\| I_d - \widehat{Q} \widehat{\Omega}^{\text{CLIME}} \right\|_\infty &\leq \left\| I_d - \widehat{Q} Q^{-1} \right\|_\infty + \left\| \widehat{Q} (Q^{-1} - \widehat{\Omega}^{\text{CLIME}}) \right\|_\infty \\
&\leq \left\| I_d - \widehat{Q} Q^{-1} \right\|_\infty + \left(\|Q\|_\infty + 1 \right) \left\| \widehat{\Omega}^{\text{CLIME}} - Q^{-1} \right\|_{1,\infty} \\
&\leq \lambda_Q + \left(\|Q\|_\infty + 1 \right) \bar{C}_Q \lambda_Q^{1-1/a_Q} \leq C_Q' \lambda_Q^{1-1/a_Q} \qquad \text{(D.46)}
\end{aligned}
$$

for example, taking C_Q' to bound:

$$
\left(\lambda_Q^{1-1/a_Q} + \left(\|Q\|_\infty + 1 \right) C_Q \right) \leq \left(1 + \left(\|Q\|_\infty + 1 \right) C_Q \right) =: C_Q'. \qquad \square
$$

LEMMA D.7 (Linearization in Sup-Norm). *Suppose Assumptions* 4.1–4.6 *hold. Then the debiased estimator* $\widehat{\beta}_{\text{DL}}$ *is asymptotically linear*:

$$
\sqrt{NT}(\widehat{\beta}_{\text{DL}} - \beta_0) = Q^{-1} \mathbb{G}_{NT} V_{it} U_{it} + R_{NT}, \qquad \text{(D.47)}
$$

$$
\|R_{NT}\|_\infty \lesssim_P {\lambda_Q}^{1-1/a_Q} \sqrt{s^2 \log d} + \sqrt{NT} \rho_{NT} = o_P(1). \qquad \text{(D.48)}
$$

PROOF OF LEMMA D.7. *Step 1.* Recall that

$$
R_{it}(\mathbf{d}, \mathbf{l}) := l_{i0}(X_{it}) - l_i(X_{it}) - \left(d_{i0}(X_{it}) - d_i(X_{it}) \right)' \beta_0.
$$

and invoking (D.23), which states that

$$
\widehat{\widetilde{Y}}_{it} - \widehat{V}_{it}' \beta_0 = \left(\widetilde{Y}_{it} - V_{it}' \beta_0 \right) + \left((\widehat{\widetilde{Y}}_{it} - \widetilde{Y}_{it}) - (\widehat{V}_{it} - V_{it})' \beta_0 \right) = U_{it} + R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}}),
$$

we can see that

$$\widehat{\widetilde{Y}}_{it} - \widehat{V}_{it}'\widehat{\beta}_L = \widehat{\widetilde{Y}}_{it} - \widehat{V}_{it}'\beta_0 + \widehat{V}_{it}'(\beta_0 - \widehat{\beta}_L),$$

$$\mathbb{E}_{NT}\widehat{V}_{it}\big(\widehat{\widetilde{Y}}_{it} - \widehat{V}_{it}'\widehat{\beta}_L\big) = \mathbb{E}_{NT}\widehat{V}_{it}\big(U_{it} + R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}})\big) + \widehat{Q}(\beta_0 - \widehat{\beta}_L).$$

Since

$$\widehat{\beta}_{\text{DL}} - \beta_0 = \widehat{\beta}_L - \beta_0 + \widehat{\Omega}^{\text{CLIME}}\big(\mathbb{E}_{NT}\widehat{V}_{it}\big(\widehat{\widetilde{Y}}_{it} - \widehat{V}_{it}'\widehat{\beta}_L\big)\big)$$

we have that

$$\begin{aligned}
\widehat{\beta}_{\text{DL}} - \beta_0 &= \widehat{\Omega}^{\text{CLIME}}(\mathbb{E}_{NT}\widehat{V}_{it}\big(U_{it} + R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}})\big) + \widehat{\Omega}^{\text{CLIME}}\widehat{Q}(\beta_0 - \widehat{\beta}_L) + \widehat{\beta}_L - \beta_0 \\
&= \widehat{\Omega}^{\text{CLIME}}(\mathbb{E}_{NT}\widehat{V}_{it}\big(U_{it} + R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}})\big) + \underbrace{\big(I_d - \widehat{\Omega}^{\text{CLIME}}\widehat{Q}\big)(\widehat{\beta}_L - \beta_0)}_{L_3} \\
&= Q^{-1}\mathbb{E}_{NT}V_{it}U_{it} + \big(\widehat{\Omega}^{\text{CLIME}} - Q^{-1}\big)\mathbb{E}_{NT}V_{it}U_{it} \\
&\quad + \widehat{\Omega}^{\text{CLIME}}(\mathbb{E}_{NT}\big[\widehat{V}_{it}\big(U_{it} + R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}}) - V_{it}U_{it}\big]\big) + L_3 \\
&= Q^{-1}\mathbb{E}_{NT}V_{it}U_{it} + L_1 + L_2 + L_3,
\end{aligned}$$

where

$$\begin{aligned}
L_1 &= \big(\widehat{\Omega}^{\text{CLIME}} - Q^{-1}\big)\mathbb{E}_{NT}V_{it}U_{it}, \\
L_2 &= \widehat{\Omega}^{\text{CLIME}}\mathbb{E}_{NT}\big[V_{it}R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}}) + (\widehat{V}_{it} - V_{it})\big(U_{it} + R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}})\big)\big], \\
L_3 &= \big(I_d - \widehat{\Omega}^{\text{CLIME}}\widehat{Q}\big)(\widehat{\beta}_L - \beta_0).
\end{aligned}$$

Term L_1. The bounds (4.11) and (B.6) imply

$$\begin{aligned}
\|L_1\|_\infty &\leq \big\|\widehat{\Omega}^{\text{CLIME}} - Q^{-1}\big\|_{1,\infty}\sqrt{NT}\|\mathbb{E}_{NT}V_{it}U_{it}\|_\infty \\
&\lesssim_P {\lambda_Q}^{1-1/a_Q}\sqrt{NT}\sqrt{\log d/NT} = o_P(1),
\end{aligned} \tag{D.49}$$

because ${\lambda_Q}^{1-1/a_Q} = o(s^{-1}\log^{-1/2} d) = o(\log^{-1/2} d)$ as assumed in (4.11).

Term L_2. The bounds (D.43) and the gradient error bound (D.22) imply

$$\begin{aligned}
\|L_2\|_\infty &\leq \big\|\widehat{\Omega}^{\text{CLIME}}\big\|_{\infty,1}\big\|\sqrt{NT}\mathbb{E}_{NT}\big[V_{it}R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}}) + (\widehat{V}_{it} - V_{it})\big(U_{it} + R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}})\big)\big]\big\|_\infty \\
&\lesssim_P 1\sqrt{NT}\rho_{NT} = o(1),
\end{aligned}$$

because $\sqrt{NT}\rho_{NT} = o(1)$ is implied by our assumption Assumptions 4.1–4.5.

Term L_3. The conditions (4.12) and (4.8) imply

$$\begin{aligned}
\sqrt{NT}\|L_3\|_\infty &= \sqrt{NT}\big\|\big(I_d - \widehat{\Omega}^{\text{CLIME}}\widehat{Q}\big)(\widehat{\beta}_L - \beta_0)\big\|_\infty \\
&\leq \sqrt{NT}\big\|I_d - \widehat{\Omega}^{\text{CLIME}}\widehat{Q}\big\|_\infty\|\widehat{\beta}_L - \beta_0\|_1 \\
&\lesssim_P \big({\lambda_Q}^{1-1/a_Q}\sqrt{NT}\sqrt{s^2\log d/NT}\big) = o(1),
\end{aligned}$$

where $\|I_d - \widehat{\Omega}^{\text{CLIME}}\widehat{Q}\|_\infty \lesssim_P {\lambda_Q}^{1-1/a_Q} = o(s^{-1}\log^{-1/2} d)$ as assumed in (4.11). □

PROOF OF THEOREM 4.2. *Step 1.* Let $\alpha \in \mathrm{R}^d$ be such that $\|\alpha\|_1 = K_\alpha = O(1)$ and $\|\alpha\|_2 = 1$. Lemma D.7 implies

$$\alpha'\big(\alpha'\Sigma\alpha\big)^{-1/2}\big(\sqrt{NT}(\widehat{\beta}_{\mathrm{DL}} - \beta_0)\big) = \alpha'\big(\alpha'\Sigma\alpha\big)^{-1/2}Q^{-1}\sqrt{NT}\mathbb{E}_{NT}V_{it}U_{it} + o_P(1),$$

where $(\alpha'\Sigma\alpha)^{-1/2} = O(1)$ because

$$\alpha'\Sigma\alpha \geq \underline{\sigma}^2\alpha'Q^{-1}\alpha \geq \underline{\sigma}^2 C_{\max}^{-1} > 0 \tag{D.50}$$

by the assumptions of the theorem, so that

$$\big|\alpha'\big(\alpha'\Sigma\alpha\big)^{-1/2}R_{NT}\big| \leq O(1)K_\alpha\|R_{NT}\|_\infty = o_P(1). \tag{D.51}$$

Consider a sequence

$$\xi_m(\alpha) := \alpha'Q^{-1}\big(\alpha'\Sigma\alpha\big)^{-1/2}V_mU_m, \quad m = 1, 2, \ldots, M$$

with

$$m = m(i,t) = T(i-1) + t, \quad 1 \leq t \leq T, 1 \leq i \leq N.$$

As shown in Corollary B.3, $\{\xi_m(\alpha)\}_{m=1}^M$ is a martingale difference sequence w.r.t. natural filtration with $M = NT$. By the law of large numbers in Hansen (2019) and the assumed Lindeberg condition

$$\frac{1}{NT}\sum_{m=1}^{NT}\xi_m^2(\alpha) \to_p \frac{\alpha'Q^{-1}\Gamma Q^{-1}\alpha}{\alpha'\Sigma\alpha} = 1.$$

As discussed in McLeish (1974), the Lindeberg condition assumed in the Theorem 4.2 implies conditions (i) and (ii) in Theorem 2.3 of McLeish (1974), which implies the first part of the theorem:

$$\mathrm{P}\big(\alpha'\big(\alpha'\Sigma\alpha\big)^{-1/2}\sqrt{NT}(\widehat{\beta}_{\mathrm{DL}} - \beta_0) \leq t\big) \to \Phi(t).$$

By Polya's theorem, the convergence is uniform in $t \in \mathrm{R}$. Since the result holds for any sequence $\{\alpha\}$ (indexed by N, T obeying conditions above, the convergence is uniform over such sequences).

Step 2. Let K_α be a finite constant in the statement of the theorem. Thus,

$$\sup_{\alpha:\|\alpha\|_2\leq 1, \|\alpha\|_1 \leq K_\alpha}\big|\alpha'(\widehat{\Sigma} - \Sigma)\alpha\big| \leq K_\alpha^2\|\widehat{\Sigma} - \Sigma\|_\infty = o_P(1)$$

by assumption. Since $\min_{\|\alpha\|_2=1}\alpha'\Sigma\alpha \geq \underline{\sigma}^2 C_{\max}^{-1}$, by assumption, we conclude that, for N and T large enough, the event

$$\mathcal{G}_K := \Big\{\inf_{\|\alpha\|_2=1, \|\alpha\|_1\leq K_\alpha}\alpha'\widehat{\Sigma}\alpha > \underline{\sigma}^2 C_{\max}^{-1}/2\Big\}$$

occurs w.p. $1 - o(1)$. Hence, w.p. $1 - o(1)$.

$$\alpha_{NT} - 1 := \frac{\big(\alpha'\Sigma\alpha\big)^{1/2}}{\big(\alpha'\widehat{\Sigma}\alpha\big)^{1/2}} - 1 \tag{D.52}$$

obeys

$$|\alpha_{NT} - 1| \le \left(\underline{\sigma}^2 C_{\max}^{-1}/2\right)^{-1} K_\alpha^2 \|\widehat{\Sigma} - \Sigma\|_\infty = o_P(1),$$

which follows from the inequality:

$$\left|1 - \frac{\sqrt{x}}{\sqrt{y}}\right| = \frac{|\sqrt{x} - \sqrt{y}|}{\sqrt{y}} = \frac{|x-y|}{\sqrt{y}(\sqrt{x}+\sqrt{y})}; \quad x > 0, y > 0.$$

Then

$$\begin{aligned} \left|(\alpha_{NT} - 1)\alpha'\left(\alpha'\Sigma\alpha\right)^{-1/2}\sqrt{NT}(\widehat{\beta}_L - \beta_0)\right| &\le |\alpha_{NT} - 1|\left|\alpha'\left(\alpha'\Sigma\alpha\right)^{-1/2}\sqrt{NT}(\widehat{\beta}_L - \beta_0)\right| \\ &= o_p(1)O_P(1). \end{aligned}$$

Therefore,

$$\alpha'\left(\alpha'\widehat{\Sigma}\alpha\right)^{-1/2}\sqrt{NT}(\widehat{\beta}_{\mathrm{DL}} - \beta_0) = \alpha'\left(\alpha'\Sigma\alpha\right)^{-1/2}\sqrt{NT}(\widehat{\beta}_L - \beta_0) + o_P(1).$$

Then convergence in distribution for the L.H.S. follows by Slutsky's lemma and Step 1. □

D.4 *Estimation of* Σ*: Proof of Lemma 4.3*

Define the following terms:

$$\bar{b}_1 = \mathbb{E}_{NT} V_{it} V_{it}' U_{it}^2 - \Gamma, \tag{D.53}$$

$$\bar{b}_2 = \mathbb{E}_{NT}\left(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\right) V_{it} U_{it}^2, \tag{D.54}$$

$$\bar{b}_3 = \mathbb{E}_{NT}\left(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\right)\left(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\right)' U_{it}^2, \tag{D.55}$$

$$\bar{b}_4 = \mathbb{E}_{NT} V_{it} V_{it}' \left(\widehat{U}_{it}^2 - U_{it}^2\right), \tag{D.56}$$

$$\bar{b}_5 = \mathbb{E}_{NT}\left(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\right) V_{it}\left(\widehat{U}_{it}^2 - U_{it}^2\right), \tag{D.57}$$

$$\bar{b}_6 = \mathbb{E}_{NT}\left(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\right)\left(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\right)'\left(\widehat{U}_{it}^2 - U_{it}^2\right). \tag{D.58}$$

The following lemma establishes tail bound on $\bar{b}_1$. Recall that κ_{NT} from Lemma D.3 is

$$\kappa_{NT} := \sqrt{\log^3\left(d^2 \log(NT)\right) \log NT/NT}.$$

Lemma D.8 (Higher-Order Term $\bar{b}_1$). *Under Assumptions* 4.1–4.5 *and* 4.7,

$$\|\Gamma\|_\infty = \max_{1 \le m, j \le d}\left|\mathrm{E} V_{itj} V_{itm} U_{it}^2\right| = O(1), \tag{D.59}$$

$$\|\bar{b}_1\|_\infty \lesssim_P \sqrt{\log^5\left(d^2 \log(NT)\right) \log NT/NT} \le \kappa_{NT} \log\left(d^2 NT\right) = o(1). \tag{D.60}$$

Proof. The bounds (D.59) and (D.60) follow from (B.19) and (B.20) with $Z_{1,nit} = Z_{2,nit} = V_{it}$, $\bar{N} = 1$, and $g = 1$. □

LEMMA D.9 (Higher-Order Term $\bar{b}_2$). *Under Assumptions* 4.1–4.5 *and* 4.7,

$$\|\bar{b}_2\|_\infty \lesssim_P (NT)^{-1/4}. \tag{D.61}$$

PROOF OF LEMMA D.9. *Step 1.* For $\bar{b}$ as in (D.6) and q_{NT} as in (D.18),

$$P_1^2 := \max_{1\leq j\leq d} \mathbb{E}_{NT}\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)_j^2 \leq \|\bar{b}\|_\infty \lesssim_P q_{NT}.$$

Invoking the convergence requirement (4.14) gives

$$\big(1 + \sqrt{\log^7\big(d^2\log(NT)\big)\log(NT)/NT}\big) \lesssim 1 + \kappa_{NT}\log^2\big(d^2 NT\big) \lesssim 1.$$

Invoking the bounds (B.19)–(B.20) with $Z_{1,nit} = 1$ and $Z_{2,nit} = V_{it}$ and $\bar{N} = 2$ and $g = 2$ give

$$P_2^2 := \max_{1\leq m\leq d} \mathbb{E}_{NT} V_{itm}^2 U_{it}^4 \lesssim_P \big(1 + \sqrt{\log^7\big(d^2\log(NT)\big)\log(NT)/NT}\big) \lesssim_P 1.$$

The Cauchy inequality implies

$$\begin{aligned}
&\max_{1\leq m,j\leq d}\big|\mathbb{E}_{NT}\big|\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)_j\big||V_{itm}|U_{it}^2\big| \\
&\leq \max_{1\leq j\leq d}\big(\mathbb{E}_{NT}\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)_j^2\big)^{1/2} \max_{1\leq m\leq d}\big(\mathbb{E}_{NT}V_{itm}^2 U_{it}^4\big)^{1/2} \\
&\lesssim_P \sqrt{q_{NT}\cdot 1} = o_P\big((NT)^{-1/4}\big),
\end{aligned}$$

where the last bound is established in (D.20). □

LEMMA D.10 (Higher-Order Term $\bar{b}_3$). *Under Assumptions* 4.1–4.5 *and* 4.7,

$$\|\bar{b}_3\|_\infty \lesssim_P o\big((NT)^{-1/4}\big). \tag{D.62}$$

PROOF OF LEMMA D.10. On the event $\sup_{it}|d_{i0}(X_{it}) - \widehat{d}_i(X_{it})| \leq \mathbf{d}_{NT,\infty} \leq 1$, which happens with probability $1 - o(1)$,

$$\begin{aligned}
&\max_{1\leq m,j\leq d} \mathbb{E}_{NT}\big|\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)_j\big|\big|\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)_m\big|U_{it}^2 \\
&\leq \max_{1\leq j\leq d} \mathbb{E}_{NT}\big|\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)_j\big|U_{it}^2 \\
&\leq \max_{1\leq j\leq d}\big(\mathbb{E}_{NT}\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)_j^2\big)^{1/2}\big(\mathbb{E}_{NT}U_{it}^4\big)^{1/2} \\
&\leq \sqrt{P_1^2}\big(\mathbb{E}_{NT}U_{it}^4\big)^{1/2} \lesssim_P o\big((NT)^{-1/4}\big).
\end{aligned}$$ □

Recall that the first-order estimation error is

$$R_{it}(\mathbf{d},\mathbf{l}) := l_{i0}(X_{it}) - l_i(X_{it}) - \big(d_{i0}(X_{it}) - d_i(X_{it})\big)'\beta_0.$$

Lemma D.11 (Squared Error). *Under Assumptions* 4.1–4.5, *we have that*

$$\mathbb{E}_{NT} R_{it}^2(\widehat{\mathbf{d}}, \widehat{\mathbf{l}}) \lesssim_P \mathbf{l}_{NT}^2 + o\big((NT)^{-1/2}\big). \tag{D.63}$$

Proof of Lemma D.11. *Step 1.* Consider a term $\bar{z}$ in (D.7) in a special case when

$$d_{i0}(X_{it}) := l_{i0}(X_{it}) \cdot (1, 1), \quad d = 2.$$

Then $\bar{z}$ reduces to a 2-vector

$$\bar{z} := \mathbb{E}_{NT}\big(l_{i0}(X_{it}) - \widehat{l}_i(X_{it})\big)^2 \cdot (1, 1),$$

and

$$\|\bar{z}\|_\infty = \mathbb{E}_{NT}\big(l_{i0}(X_{it}) - \widehat{l}_i(X_{it})\big)^2.$$

Invoking (G.16) with $\mathbf{d}_{NT}$ and $\mathbf{d}_{NT,\infty}$ replaced by $\mathbf{l}_{NT}$ and $\mathbf{l}_{NT,\infty}$ gives the bound.

Step 2. The following bound holds:

$$\begin{aligned}
\mathbb{E}_{NT} R_{it}^2(\widehat{\mathbf{d}}, \widehat{\mathbf{l}}) &\le 2\mathbb{E}_{NT}\big(\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)'\beta_0\big)^2 + 2\mathbb{E}_{NT}\big(l_{i0}(X_{it}) - \widehat{l}_i(X_{it})\big)^2 \\
&= 2\beta_0' \bar{b}\beta_0 + 2\mathbb{E}_{NT}\big(l_{i0}(X_{it}) - \widehat{l}_i(X_{it})\big)^2 \\
&\le 2\|\bar{b}\|_\infty \|\beta_0\|_1^2 + 2\mathbb{E}_{NT}\big(l_{i0}(X_{it}) - \widehat{l}_i(X_{it})\big)^2 \\
&\lesssim_P \|\beta_0\|_1^2\big(\mathbf{d}_{NT}^2 + \mathbf{d}_{NT,\infty}^2\sqrt{(NT)^{-1}\log(NT)\log d}\big) + \mathbf{l}_{NT}^2 + o\big((NT)^{-1/2}\big) \\
&\lesssim_P^{\text{i}} o\big((NT)^{-1/2}\big) + \mathbf{l}_{NT}^2 + o\big((NT)^{-1/2}\big),
\end{aligned}$$

where (i) follows combining $\|\beta_0\|_1 \le \bar{C}_\beta$ assumed in Assumption 4.5(a) and $q_{NT} = o((NT)^{-1/2})$, established in (D.20). □

Lemma D.12 (Higher-Order Terms $\bar{b}_4, \bar{b}_5, \bar{b}_6$ With $\widehat{U}_{it}^2 - U_{it}^2$). *Under Assumptions* 4.1–4.7,

$$\sum_{k=4}^{6} \|\bar{b}_k\|_\infty \lesssim_P \big((NT)^{-1/4} + \mathbf{l}_{NT} + \sqrt{s\log d/NT} + \mathbf{l}_{NT}^2 \log\big(d^2 NT\big)\big) =: \gamma_{NT}. \tag{D.64}$$

Proof of Lemma D.12. *Step 1.* Decompose

$$\widehat{U}_{it}^2 - U_{it}^2 = (\widehat{U}_{it} - U_{it} + U_{it})^2 - U_{it}^2 = 2U_{it}(\widehat{U}_{it} - U_{it}) + (\widehat{U}_{it} - U_{it})^2.$$

Invoking (D.23) gives

$$\begin{aligned}
\widehat{U}_{it} &= \widehat{\widetilde{Y}}_{it} - \widehat{V}_{it}'\widehat{\beta}_L = \big(\widehat{\widetilde{Y}}_{it} - \widehat{V}_{it}'\beta_0\big) + \big(\widehat{V}_{it}'\beta_0 - \widehat{V}_{it}'\widehat{\beta}_L\big) \\
&= U_{it} + R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}}) + \widehat{V}_{it}'(\beta_0 - \widehat{\beta}_L).
\end{aligned}$$

The Cauchy inequality implies

$$\begin{aligned}
\mathbb{E}_{NT}(\widehat{U}_{it} - U_{it})^2 &\le 2\mathbb{E}_{NT} R_{it}^2(\widehat{\mathbf{d}}, \widehat{\mathbf{l}}) + 2\mathbb{E}_{NT}\big(\widehat{V}_{it}'(\widehat{\beta}_L - \beta_0)\big)^2 \\
&= 2\mathbb{E}_{NT} R_{it}^2(\widehat{\mathbf{d}}, \widehat{\mathbf{l}}) + 2(\widehat{\beta}_L - \beta_0)'\widehat{Q}(\widehat{\beta}_L - \beta_0) =: U_1 + U_2,
\end{aligned}$$

where $U_1 \lesssim_P o(NT)^{-1/2} + \mathbf{l}_{NT}^2$ in established in Lemma D.11 and $U_2 \lesssim_P s \log d/NT$ is Theorem 4.1.

Step 2. Let $C(W_{it}, \eta) = (C_{mj}(W_{it}, \eta))$ be a $d \times d$ matrix. For any coordinates m and j, decompose

$$\begin{aligned}
&\mathbb{E}_{NT} C_{mj}(W_{it}, \widehat{\eta})\big(\widehat{U}_{it}^2 - U_{it}^2\big) \\
&= 2\mathbb{E}_{NT} C_{mj}(W_{it}, \widehat{\eta}) U_{it}(\widehat{U}_{it} - U_{it}) + \mathbb{E}_{NT} C_{mj}(W_{it}, \widehat{\eta})(\widehat{U}_{it} - U_{it})^2 \\
&=: 2D_{1mj}(\widehat{\eta}) + D_{2mj}(\widehat{\eta}).
\end{aligned}$$

The Cauchy inequality gives

$$\begin{aligned}
\big|D_{1mj}(\widehat{\eta})\big| &\le \big(\mathbb{E}_{NT} C_{mj}^2(W_{it}, \widehat{\eta}) U_{it}^2\big)^{1/2} \big(\mathbb{E}_{NT}(\widehat{U}_{it} - U_{it})^2\big)^{1/2} \\
&\le \max_{1 \le m, j \le d} \big(\mathbb{E}_{NT} C_{mj}^2(W_{it}, \widehat{\eta}) U_{it}^2\big)^{1/2} \big(\mathbb{E}_{NT}(\widehat{U}_{it} - U_{it})^2\big)^{1/2}.
\end{aligned}$$

The maximal inequality gives

$$\big|D_{2mj}(\widehat{\eta})\big| \le \max_{it} \max_{mj} \big|C_{mj}(W_{it}, \widehat{\eta})\big| \mathbb{E}_{NT}(\widehat{U}_{it} - U_{it})^2. \tag{D.65}$$

Step 3. If one can verify

$$\max_{1 \le m, j \le d} \mathbb{E}_{NT} C_{mj}^2(W_{it}, \widehat{\eta}) U_{it}^2 \lesssim_P 1, \tag{D.66}$$

we have that

$$\begin{aligned}
\big\|D_1(\widehat{\eta})\big\|_\infty &= \max_{1 \le m, j \le d} \big|D_{1mj}(\widehat{\eta})\big| \\
&\le O_P(1) \cdot O_P\big((NT)^{-1/4} + \mathbf{l}_{NT} + \sqrt{s \log d/NT}\big) \\
&\lesssim_P \big((NT)^{-1/4} + \mathbf{l}_{NT} + \sqrt{s \log d/NT}\big).
\end{aligned}$$

If one can verify another condition,

$$\max_{it} \max_{mj} |C_{mj}(W_{it}, \widehat{\eta}| \lesssim_P \big(\log\big(d^2 NT\big)\big), \tag{D.67}$$

we have that

$$\big\|D_2(\widehat{\eta})\big\|_\infty = O_P\big(\log\big(d^2 NT\big)\big) \cdot O_P\big(s \log d/NT + \mathbf{l}_{NT}^2 + (NT)^{-1/2}\big).$$

Step 4.1. Take $C(W_{it}, \widehat{\eta}) = V_{it} V_{it}'$, which corresponds to $\bar{b}_4$ in (D.56). Invoking (B.19) and (B.20) with $Z_{1,nit} = Z_{2,nit} = V_{it}$ and $\bar{N} = 2$ and $g = 1$ as well the assumed bound (4.14) give

$$\max_{jk} \big|\mathbb{E}_{NT} V_{itk}^2 V_{itj}^2 U_{it}^2\big| \lesssim_P \big(1 + \kappa_{NT} \log^2\big(d^2 NT\big)\big) \lesssim_P 1,$$

which verifies (D.66). By Lemma B.1 (6),

$$\max_{it} \max_{mj} |V_{itk} V_{itj}| \lesssim_P \big(\log\big(d^2 NT\big)\big),$$

which verifies (D.67).

Step 4.2. Take $C(W_{it}, \widehat{\eta}) = V_{it}(d_{i0}(X_{it}) - \widehat{d}_i(X_{it}))$, which corresponds to $\bar{b}_5$ in (D.57). In what follows, we focus on the event $\sup_{it} |d_{i0}(X_{it}) - \widehat{d}_i(X_{it})| \le \mathbf{d}_{NT,\infty} \le 1$. Invoking (B.18) with $\bar{N} = 2$ and $g = 1$ gives

$$\begin{aligned}
&\max_{1\le m,j\le d} \big|\mathbb{E}_{NT} V_{itm}^2 \big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)_j^2 U_{it}^2\big| \\
&\lesssim_P \mathbf{d}_{NT,\infty}^2 \max_{1\le m\le d} \big|\mathbb{E}_{NT} V_{itm}^2 U_{it}^2\big| \\
&\lesssim_P d_{NT,\infty}\big(1 + \sqrt{\log^5(d\log NT)\log(NT)/NT}\big) \\
&= \mathbf{d}_{NT,\infty}\big(1 + \kappa_{NT}\log(d\log NT)\big)\big) \lesssim \mathbf{d}_{NT,\infty}.
\end{aligned}$$

Likewise,

$$\max_{it}\max_{mj}\big|C(W_{it}, \widehat{\eta})\big| \le \mathbf{d}_{NT,\infty} \max_{it}\max_{1\le j\le d} |V_{itj}| \lesssim_P \big(\log\big(d^2 NT\big)\mathbf{d}_{NT,\infty}\big)$$

verifies (D.67).

Step 4.3. Take $C(W_{it}, \widehat{\eta}) = (d_{i0}(X_{it}) - \widehat{d}_i(X_{it}))(d_{i0}(X_{it}) - \widehat{d}_i(X_{it}))'$, which corresponds to $\bar{b}_6$ in (D.58). On the event $\sup_{it} |d_{i0}(X_{it}) - \widehat{d}_i(X_{it})| \le \mathbf{d}_{NT,\infty} \le 1$, the condition (D.66) becomes

$$\max_{1\le m,j\le d} \mathbb{E}_{NT}\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)_m^2\big(d_{i0}(X_{it}) - \widehat{d}_i(X_{it})\big)_j^2 U_{it}^2 \le \mathbb{E}_{NT} U_{it}^2 \lesssim_P 1.$$

Noting that $\max_{it}\max_{mj} |C(W_{it}, \widehat{\eta})| \lesssim_P \mathbf{d}_{NT,\infty}^2 \lesssim_P 1$ verifies (D.67).

Step 5. (Conclusion). Collecting the bounds and invoking $(s \vee 1)\kappa_{NT} = o(1)$ gives

$$\begin{aligned}
&o(NT)^{-1/4} + \mathbf{l}_{NT} + \sqrt{s\log d/NT} \\
&\quad + \big(\log\big(d^2NT\big)\big(o(NT)^{-1/2} + \mathbf{l}_{NT}^2 + s\log d/NT\big)\big) \lesssim \gamma_{NT}.
\end{aligned} \tag{D.68}$$

For N and T large enough,

$$\log\big(d^2NT\big)/(NT)^{-1/4} \le 1,$$

which implies $\log(d^2NT)(NT)^{-1/4} = o((NT)^{-1/4})$. Likewise,

$$\log\big(d^2NT\big)\sqrt{s\log d/NT} \le s\sqrt{\log^2\big(d^2NT\big)\log d/NT} \le (s \vee 1)\kappa_{NT} = o(1),$$

which gives (D.68). □

LEMMA D.13 (Bound on $\|\widehat{\Gamma}(\widehat{\beta}_L) - \Gamma\|_\infty$). *Under Assumptions* 4.1–4.5 *and* 4.7, *we have that*

$$\big\|\widehat{\Gamma}(\widehat{\beta}_L) - \Gamma\big\|_\infty \lesssim_P \big(\gamma_{NT} + \kappa_{NT}\log\big(d^2NT\big)\big) = o_p(1). \tag{D.69}$$

Proof. Decompose the matrix first-stage error

$$\begin{aligned}\widehat{\Gamma}(\widehat{\beta}_L) - \Gamma &= \mathbb{E}_{NT}\widehat{V}_{it}\widehat{V}_{it}'\widehat{U}_{it}^2 - \Gamma \\ &= \mathbb{E}_{NT}\widehat{V}_{it}\widehat{V}_{it}'\big(\widehat{U}_{it}^2 - U_{it}^2\big) + \mathbb{E}_{NT}\big(\widehat{V}_{it}\widehat{V}_{it}' - V_{it}V_{it}'\big)U_{it}^2 + \mathbb{E}_{NT}V_{it}V_{it}'U_{it}^2 - \Gamma \\ &= \bar{b}_6 + \bar{b}_5 + \bar{b}_5' + \bar{b}_4 + \bar{b}_3 + \bar{b}_2 + \bar{b}_2' + \bar{b}_1.\end{aligned}$$

The bound on $\bar{b}_1$ is given in (D.60), Lemma D.8. The bound on $\bar{b}_2$ is given in (D.61), Lemma D.9. The bound on $\bar{b}_3$ is given in (D.62), Lemma D.10. The bounds on $\bar{b}_4 - \bar{b}_6$ are given in (D.64), Lemma D.12. Summing the bounds gives (D.69). □

Proof of Lemma 4.3. *Step 1.* Define the following bounds:

$$\begin{aligned}\Sigma_1 &:= \big\|\widehat{\Omega}^{\text{CLIME}} - Q^{-1}\big\|_{\infty,1}\big\|\widehat{\Gamma}(\widehat{\beta}_L)\big\|_{\infty}\big\|\widehat{\Omega}^{\text{CLIME}}\big\|_{1,\infty}, \\ \Sigma_2 &:= \big\|Q^{-1}\big\|_{\infty,1}\big\|\widehat{\Gamma}(\widehat{\beta}_L) - \Gamma\big\|_{\infty}\big\|\widehat{\Omega}^{\text{CLIME}}\big\|_{1,\infty}, \\ \Sigma_3 &:= \big\|\widehat{\Omega}^{\text{CLIME}} - Q^{-1}\big\|_{\infty,1}\|\Gamma\|_{\infty}\big\|Q^{-1}\big\|_{1,\infty}\end{aligned}$$

and note that

$$\big\|\widehat{\Sigma}(\widehat{\beta}_L) - \Sigma\big\|_{\infty} = \big\|\widehat{\Omega}^{\text{CLIME}}\widehat{\Gamma}(\widehat{\beta}_L)\widehat{\Omega}^{\text{CLIME}} - Q^{-1}\Gamma Q^{-1}\big\|_{\infty} \leq \Sigma_1 + \Sigma_2 + \Sigma_3.$$

Step 2. Invoking (D.59) and (D.69) give

$$\big\|\widehat{\Gamma}(\widehat{\beta}_L)\big\|_{\infty} \leq \|\Gamma\|_{\infty} + \big\|\widehat{\Gamma}(\widehat{\beta}_L) - \Gamma\big\|_{\infty} \lesssim_P 1 + \gamma_{NT} + \kappa_{NT}\log\big(d^2NT\big) \lesssim_P 1.$$

Invoking (D.43) and (D.40) give

$$\big\|\widehat{\Omega}^{\text{CLIME}}\big\|_{1,\infty} \leq \|\widehat{\Omega}\|_{1,\infty} \leq \big\|Q^{-1}\big\|_{1,\infty} \leq \big(A_Q/(a_Q - 1)\big).$$

As a result, invoking (4.12) gives

$$\Sigma_1 = O_P\big({\lambda_Q}^{1-1/a_Q}\big) \cdot O_P(1) \cdot O_P(1).$$

Likewise,

$$\begin{aligned}\Sigma_2 &:= \big\|Q^{-1}\big\|_{\infty,1}\big\|\widehat{\Gamma}(\widehat{\beta}_L) - \Gamma\big\|_{\infty}\big\|\widehat{\Omega}^{\text{CLIME}}\big\|_{1,\infty} = O(1) \cdot O_P\big(\gamma_{NT} + \kappa_{NT}\log\big(d^2NT\big)\big) \cdot O_P(1) \\ &\lesssim_P \big(\gamma_{NT} + \kappa_{NT}\log\big(d^2NT\big)\big), \\ \Sigma_3 &:= \big\|\widehat{\Omega}^{\text{CLIME}} - Q^{-1}\big\|_{\infty,1}\|\Gamma\|_{\infty}\big\|Q^{-1}\big\|_{1,\infty} = O(1) \cdot O_P(\gamma_{NT}) \cdot O_P(1) \lesssim_P \big({\lambda_Q}^{1-1/a_Q}\big).\end{aligned}$$

Collecting the terms give

$$\begin{aligned}\big\|\widehat{\Sigma}(\widehat{\beta}_L) - \Sigma\big\|_{\infty} &= \big\|\widehat{\Omega}^{\text{CLIME}}\widehat{\Gamma}(\widehat{\beta}_L)\widehat{\Omega}^{\text{CLIME}} - Q^{-1}\Gamma Q^{-1}\big\|_{\infty} \\ &\leq \Sigma_1 + \Sigma_2 + \Sigma_3 \\ &\lesssim_P {\lambda_Q}^{1-1/a_Q} + \gamma_{NT} + \kappa_{NT}\log\big(d^2NT\big) + {\lambda_Q}^{1-1/a_Q}.\end{aligned}$$

□

D.5 *Proof of Theorem 4.3*

The proof is divided in several steps. Step 1 outlines the proof. Steps 2–5 establish (4.20). Steps 6–8 establish (4.22).

PROOF. *Step 1. (Outline).* Let $Z \sim N(0, \mathcal{C})$ and $\widehat{Z} \mid \widehat{\mathcal{C}} \sim N(0, \widehat{\mathcal{C}})$ be as defined in the Theorem. Define

$$T_{\Sigma,\beta} := \sqrt{NT}\Sigma_{jj}^{-1/2}(\widehat{\beta}_{\mathrm{DL},j} - \beta_0), \qquad T_{\widehat{\Sigma},\beta} := \sqrt{NT}\widehat{\Sigma}_{jj}^{-1/2}(\widehat{\beta}_{\mathrm{DL},j} - \beta_0)$$

and

$$T_{\Sigma} := \Sigma_{jj}^{-1/2}\mathbb{G}_{NT}V_{itj}U_{it}.$$

Define

$$\begin{aligned}
O_1(t) &:= \mathrm{P}\big(\|T_{\Sigma,\beta}\|_\infty < t\big) - \mathrm{P}\big(\|T_{\Sigma}\|_\infty < t + \delta_1\big),\\
O_2(t) &:= \mathrm{P}\big(\|T_{\Sigma}\|_\infty \le t + \delta_1\big) - \mathrm{P}\big(\|Z\|_\infty < t + \delta_1\big),\\
O_3(t) &:= \mathrm{P}\big(\|Z\|_\infty < t + \delta_1\big) - \mathrm{P}\big(\|Z\|_\infty < t\big)
\end{aligned}$$

and note that for each t

$$\mathrm{P}\big(\|T_{\Sigma,\beta}\|_\infty < t\big) - \mathrm{P}\big(\|Z\|_\infty < t\big) = \sum_{k=1}^{3} O_k(t).$$

Likewise, define

$$O_4(t) := \mathrm{P}\big(\|T_{\widehat{\Sigma},\beta}\|_\infty < t\big) - \mathrm{P}\big(\|T_{\Sigma,\beta}\|_\infty < t\big)$$

and

$$O_5(t) := \mathrm{P}\big(\|Z\|_\infty < t\big) - \mathrm{P}\big(\|\widehat{Z}\|_\infty < t \mid \widehat{\mathcal{C}}\big).$$

Note that for each t,

$$\mathrm{P}\big(\|T_{\widehat{\Sigma},\beta}\|_\infty < t\big) - \mathrm{P}\big(\|\widehat{Z}\|_\infty < t \mid \widehat{\mathcal{C}}\big) = O_4(t) + \sum_{k=1}^{3} O_k(t) + O_5(t).$$

Then (4.20) is equivalent to

$$\sup_{t\ge 0}\big|\mathrm{P}\big(\|T_{\Sigma,\beta}\|_\infty < t\big) - \mathrm{P}\big(\|Z\|_\infty < t\big)\big| \to 0 \tag{D.70}$$

and (4.22) is equivalent to

$$\sup_{t\ge 0}\big|\mathrm{P}\big(\|T_{\widehat{\Sigma},\beta}\|_\infty < t\big) - \mathrm{P}\big(\|\widehat{Z}\|_\infty < t \mid \widehat{\mathcal{C}}\big)\big| \to_P 0. \tag{D.71}$$

Step 2. We show that the elements of diagΣ are bounded from above and below. By Assumption 4.3(2), there exists a finite $\bar{\sigma}_{UV}$ such that $\max_{it} \mathrm{E}[U_{it}^2 \mid V_{it}] \le \bar{\sigma}_{UV}^2$ a.s. As a result, Assumption 4.3 gives

$$0 < \underline{\sigma}^2 \le \min_{it} \mathrm{E}\big[U_{it}^2 \mid V_{it}\big] \le \max_{it} \mathrm{E}\big[U_{it}^2 \mid V_{it}\big] \le \bar{\sigma}_{UV}^2 < \infty \quad \text{a.s.},$$

which implies $\underline{\sigma}^2 Q \preceq \Gamma \preceq \bar{\sigma}_{UV}^2 Q$, and $\underline{\sigma}^2 Q^{-1} \preceq \Sigma \preceq \bar{\sigma}_{UV}^2 Q^{-1}$. As a result,

$$0 < \underline{\sigma}^2 C_{\max}^{-1} \le c_\Sigma = \min_{1\le j\le d} \Sigma_{jj} \le C_\Sigma = \max_{1\le j\le d} \Sigma_{jj} \le \bar{\sigma}_{UV}^2 C_{\min}^{-1} < \infty.$$

Likewise, the elements of $(\operatorname{diag}\Sigma)^{-1/2}$ are bounded from above by $c_\Sigma^{-1/2}$ and from below by $C_\Sigma^{-1/2}$.

Step 3. We bound $\sup_{t\ge 0} |O_1(t)|$ with $\delta_1 = \log^{-1/2} d \log^{-1/2} NT$. Decomposition (D.47) implies

$$\|T_\Sigma\|_\infty - \|R_{NT}\|_\infty \le \|T_{\Sigma,\beta}\|_\infty \le \|T_\Sigma\|_\infty + \|R_{NT}\|_\infty,$$

and union bound gives

$$\mathrm{P}\big(\|T_{\Sigma,\beta}\|_\infty < t\big) \le \mathrm{P}\big(\|T_\Sigma\|_\infty \le t + \delta_1\big) + \mathrm{P}\big(\|R_{NT}\|_\infty \ge \delta_1\big),$$
$$\mathrm{P}\big(\|T_\Sigma\|_\infty < t\big) \le \mathrm{P}\big(\|T_{\Sigma,\beta}\|_\infty \le t + \delta_1\big) + \mathrm{P}\big(\|R_{NT}\|_\infty \ge \delta_1\big),$$

which gives

$$\sup_{t\ge 0}\big|O_1(t)\big| \le \mathrm{P}\big(\|R_{NT}\|_\infty \ge \delta_1\big) \overset{\text{i}}{=} o(1), \tag{D.72}$$

where (i) follows from

$$\|R_{NT}\|_\infty \lesssim_P \lambda_Q^{1-1/a_Q} s \log^{1/2} d + \sqrt{NT}\rho_{NT} = o_P\big(\log^{-1/2} d \log^{-1/2} NT\big)$$

given in (D.48) and (4.19).

Step 4. We verify the conditions of Lemma C.5 for the m.d.s. with

$$m = m(i,t) = T(i-1) + t, \quad M = NT$$

and

$$X_m := (\operatorname{diag}\Sigma)^{-1/2} V_m U_m, \quad m = 1, 2, \dots, M, \qquad D_M = c_\Sigma^{-1} \pi_M^{VU}.$$

To verify the condition (C.5), we invoke Assumption 4.8, which gives

$$\operatorname{Var}(X_{mj}) = \Sigma_{jj}^{-1/2} \mathrm{E} V_{mj}^2 U_m^2 \Sigma_{jj}^{-1/2} \ge \underline{\sigma}^2 \min_{it} \big\| \mathrm{E} V_{it} V_{it}' \big\|_\infty C_\Sigma^{-1} =: a_1 > 0$$

and Remark B.2,

$$\operatorname{Var}(X_{mj}) = \Sigma_{jj}^{-1/2} \mathrm{E} V_{mj}^2 U_m^2 \Sigma_{jj}^{-1/2} \le \bar{\sigma}^2 \max_{it} \big\| \mathrm{E} V_{it} V_{it}' \big\|_\infty c_\Sigma^{-1} =: A_1 < \infty.$$

By Assumption 4.8,

$$\bar{r} := (2/\kappa) \cdot \log(NT), \qquad \bar{q} := (NT)^{c_2} \log^2 d \log^2(NT)$$

obey (4.18), which implies (C.9). By Lemma C.5, there exist constants $c_2 \in (0, 1/4)$ and c_X and C_X depending on $\underline{\sigma}$, $\bar{\sigma}$, c_2, $C_{\min}$, $C_{\max}$ such that

$$\sup_{t\geq 0}\bigl|O_2(t)\bigr| = \sup_{t\geq 0}\bigl|\mathrm{P}\bigl(\|T_\Sigma\|_\infty \leq t\bigr) - \mathrm{P}\bigl(\|Z\|_\infty \leq t\bigr)\bigr| \lesssim C_X (NT)^{-c_X} + (NT)^{-c_2/2}. \tag{D.73}$$

Step 4. Bound on $\sup_{t\geq 0}|O_3(t)|$. Invoking Lemma C.2 gives

$$\begin{aligned}
&\sup_{t\geq 0}\bigl|O_3(t)\bigr| \\
&\quad = \sup_{t\geq 0}\bigl|\mathrm{P}\bigl(\|Z\|_\infty < t + \delta_1\bigr) - \mathrm{P}\bigl(\|Z\|_\infty < t\bigr)\bigr| \\
&\quad \leq \sup_{t\geq 0}\bigl|\mathrm{P}\bigl(\|Z\|_\infty < t + \delta_1\bigr) - \mathrm{P}\bigl(\|Z\|_\infty < t - \delta_1\bigr)\bigr| \\
&\quad = \sup_{t\geq 0}\mathrm{P}\bigl(\bigl|\|Z\|_\infty - t\bigr| \leq \delta_1\bigr) \leq C\delta_1\sqrt{1 \vee \log(2d/\delta_1)}.
\end{aligned}$$

Note that the R.H.S. is a nondecreasing function of δ_1 in some neighborhood of 0 and that $\sqrt{1 \vee (x+y)} \leq 1 + \sqrt{x} + \sqrt{y}$ for $x, y > 0$. Plugging in $\delta_1 = \log^{-1/2} d \log^{-1/2} NT$ gives

$$\begin{aligned}
\sup_{t\geq 0}\bigl|O_3(t)\bigr| &\leq C\log^{-1/2} d \log^{-1/2} NT\sqrt{1 \vee \bigl(\log(2d) + \log\bigl(\log^{1/2} d \log^{1/2} NT\bigr)\bigr)} \\
&\lesssim \log^{-1/2} NT + \log^{-1/2} d \log^{-1/2} NT \log^{1/2} \log(NT) = o(1).
\end{aligned} \tag{D.74}$$

Combining (D.72) and (D.73) and (D.74) give (D.70). By a standard calculation, we have $\mathrm{E}\|Z\|_\infty \lesssim \sqrt{\log 2d}$. Invoking Gaussian concentration inequality (see, e.g., Ledoux (2001, Theorem 7.1), or Comment 4 in Chernozhukov, Chetverikov, and Kato (2015, p. 56)) implies

$$\|Z\|_\infty \lesssim_P \log^{1/2}(2d) + \log^{1/2}(NT).$$

Since $\|Z\|_\infty$ and $\|T_{\Sigma,\beta}\|_\infty$ converge in distribution to the same limit,

$$\|T_{\Sigma,\beta}\|_\infty \lesssim_P \log^{1/2}(2d) + \log^{1/2}(NT). \tag{D.75}$$

Step 5.1. We bound $\sup_{t\geq 0}|O_4(t)|$. Take $\rho_j = \Sigma_{jj}^{1/2}/\widehat{\Sigma}_{jj}^{1/2}$ and let $\rho := (\rho_1, \rho_2, \ldots, \rho_d)'$ be a d-vector. Note that all Euclidean j-vectors e_j vectors obey $\|e_j\|_2 = \|e_j\|_1 = 1$ and, therefore, belong to the set in Theorem 4.2 with $K_\alpha = 1$. Let $\alpha_{NT} = (\alpha\Sigma\alpha)^{1/2}/(\alpha\widehat{\Sigma}\alpha)^{1/2}$ be as in (D.52). Invoking (D.51) and the bound (4.15) in Lemma 4.3 gives

$$\max_j|\rho_j - 1| \leq \sup_{\alpha:\|\alpha\|_2 = \|\alpha\|_1 = 1}|\alpha_{NT} - 1| \lesssim_P \gamma_{NT}.$$

In particular, it implies that the even

$$\min_{1\leq j\leq d}\rho_j > 1/2$$

occurs w.p. $1-o(1)$. For any $\rho_j > 1/2$,

$$\left|\rho_j^{-1} - 1\right| = |1-\rho_j|/|\rho_j| \leq 2|\rho_j - 1|.$$

Combining the bounds above on the event $\min_{1\leq j\leq d} \rho_j > 1/2$ give

$$\max_{1\leq j\leq d} \left|\widehat{\Sigma}_{jj}^{-1/2}/\Sigma_{jj}^{-1/2} - 1\right| = \max_{1\leq j\leq d}\left|\rho_j^{-1} - 1\right| \leq 2 \max_{1\leq j\leq d} |\rho_j - 1| \lesssim_P \gamma_{NT}. \tag{D.76}$$

Step 5.2. Let $v_1 \cdot v_2$ denote $(v_1 \cdot v_2)_j = v_{1j} \cdot v_{2j}$ for $j = 1, 2, \ldots, d$. Note that

$$T_{\widehat{\Sigma},\beta} = T_{\Sigma,\beta} \cdot \rho^{-1},$$

or, equivalently,

$$T_{\widehat{\Sigma},\beta} - T_{\Sigma,\beta} = \left(\rho^{-1} - 1\right) T_{\Sigma,\beta}.$$

Invoking (D.76) and (D.75) give

$$\|T_{\widehat{\Sigma},\beta} - T_{\Sigma,\beta}\|_\infty \leq \max_{1\leq j\leq d}\left|\rho_j^{-1} - 1\right| \|T_{\Sigma,\beta}\|_\infty = O_P(\zeta_{NT}) \cdot O_P\left(\log^{1/2} d + \log^{1/2} NT\right) =^{\text{i}} o_P(1),$$

where (i) follows from (4.21). Thus, $\|T_{\widehat{\Sigma},\beta}\|_\infty$ and $\|T_{\Sigma,\beta}\|_\infty$ converge to the same limit in distribution.

Step 6. We bound $\sup_{t\geq 0} |O_5(t)|$. Invoking Lemma C.1 with $X \sim N(0, \mathcal{C})|\widehat{\mathcal{C}}$ and $Y \sim N(0, \mathcal{C})$ and $\widehat{\Delta} = \|\mathcal{C} - \widehat{\mathcal{C}}\|_\infty$

$$\sup_{t\geq 0}\left|O_5(t)\right| \leq C'\left(\widehat{\Delta} \log^2(2d)\right)^{1/2},$$

where C depends only on the constants defined in Assumptions 4.2 and 4.3. In Step 7, we show that for ζ_{NT} in (4.15),

$$\widehat{\Delta} := \|\mathcal{C} - \widehat{\mathcal{C}}\|_\infty \lesssim_P^{\text{i}} \zeta_{NT} =^{\text{ii}} o_P\left(\log^{-2} d \log^{-1} NT\right), \tag{D.77}$$

where (i) is verified in Steps 7–8 and (ii) is directly assumed in (4.21).

Step 7. Note that

$$\|\Sigma\|_\infty = \left\|Q^{-1}\Gamma Q^{-1}\right\|_\infty \leq \left\|Q^{-1}\right\|_{\infty,1} \|\Gamma\|_\infty \left\|Q^{-1}\right\|_{1,\infty} \leq \left(A_Q/(a_Q - 1)\right)^2 \|\Gamma\|_\infty = O(1).$$

As a result,

$$\|\widehat{\Sigma}\|_\infty \leq \|\widehat{\Sigma} - \Sigma\|_\infty + \|\Sigma\|_\infty \lesssim_P 1 + \gamma_{NT} \lesssim_P 1.$$

Likewise,

$$\left\|(\text{diag}\widehat{\Sigma})^{-1/2}\right\|_{\infty,1} = \left\|(\text{diag}\widehat{\Sigma})^{-1/2}\right\|_{1,\infty} = \max_{1\leq j\leq d} \widehat{\Sigma}_{jj}^{-1/2} \lesssim_P \zeta_{NT} + c_\Sigma^{-1/2} \lesssim_P 1.$$

Step 8. Define

$$C_1 := \max_{1\leq j\leq d}\left|\widehat{\Sigma}_{jj}^{-1/2} - \Sigma_{jj}^{-1/2}\right| \|\widehat{\Sigma}\|_\infty \left\|(\text{diag}\,\widehat{\Sigma})^{-1/2}\right\|_{1,\infty},$$

$$C_2 := \left\|(\text{diag}\Sigma)^{-1/2}\right\|_{\infty,1}\|\widehat{\Sigma}-\Sigma\|_\infty\left\|(\text{diag}\widehat{\Sigma})^{-1/2}\right\|_{1,\infty},$$
$$C_3 := \left\|(\text{diag}\Sigma)^{-1/2}\right\|_{\infty,1}\|\Sigma\|_\infty \max_{1\le j\le d}\left|\widehat{\Sigma}_{jj}^{-1/2}-\Sigma_{jj}^{-1/2}\right|$$

and note that

$$\|\widehat{\mathcal{C}}-\mathcal{C}\|_\infty = \left\|(\text{diag}\widehat{\Sigma})^{-1/2}\widehat{\Sigma}(\text{diag}\widehat{\Sigma})^{-1/2}-(\text{diag}\Sigma)^{-1/2}\Sigma(\text{diag}\Sigma)^{-1/2}\right\|_\infty \le C_1+C_2+C_3.$$

Invoking (D.76) and (4.15),

$$\max_{1\le j\le d}\left|\widehat{\Sigma}_{jj}^{-1/2}-\Sigma_{jj}^{-1/2}\right| \lesssim_P \zeta_{NT}, \qquad \|\widehat{\Sigma}-\Sigma\|_\infty \lesssim_P \zeta_{NT}$$

implies that each term C_j is a product of two $O_P(1)$ terms and a single $O_P(\zeta_{NT})$ term. Thus, $C_1+C_2+C_3 \lesssim_P \zeta_{NT}$ verifies (i) in (D.77). □

Proof of Lemma 4.4. We invoke Lemma D.6 with $\bar{V}_{it} = D_{it} - d_{i0}(Z_{it})$ and $\bar{\tilde{Y}}_{it} = Y_{it} - l_{i0}(Z_{it})$ and $\bar{\beta}_0 = (\beta_0, \rho_0)$ and $g = 2$. Steps 1, 2, and 3 are established similar to the proof of Theorem 4.1. Thus, the bounds (4.27) hold for the orthogonal group lasso. As a result, $\|\widehat{\beta}_L - \beta_0\|_1 \le \sqrt{s^2\log d/NT}$ w.p. $1-o(1)$. As a result, the debiased orthogonal group lasso obeys the uniform linearization result (D.47), and Theorems 4.2 and 4.3 hold. □

Appendix E: Proofs for Section 5

Proof of Remark 5.1. To prove this, let $\|\cdot\|_{\psi_2}$ denote the Orlizs sub-Gaussian norm under the probability measure P (see van der Vaart and Wellner (1996)). Then

$$\left\|\|F_{it}\|\right\|_{\psi_2} \le \left\|\|\Pi_{it}F_{i,t-1}\|\right\|_{\psi_2} + \left\|\|QT_{it}\|\right\|_{\psi_2} \le (1-\delta)\left\|\|F_{i,t-1}\|\right\|_{\psi_2} + A'\bar{\sigma}^2,$$

where A' is a numerical constant. Iterating on this inequality exactly t times we obtain

$$\left\|\|F_{it}\|\right\|_{\psi_2} \le (1-\delta)^t\left\|\|F_{i,0}\|\right\|_{\psi_2} + A'\sum_{\bar{t}=1}^{t-1}(1-\delta)^{\bar{t}}\bar{\sigma}^2 \le A'\frac{\bar{\sigma}^2}{1-\delta}.$$ □

Proof of Remark 5.3. Step 1 shows that $\mathbf{p}_{NT} \le N^{-1/2}(2(B_{\max}+1))^{1/2}\zeta_{NT,\infty}$. Step 2 shows that w.p. $1-o(1)$,

$$\sup_{it}\left|p_i(X_{it}) - p_{i0}(X_{it})\right| \le 2\zeta_{NT,\infty}.$$

Step 1. For any δ^P and $\xi \in \bar{P}_{NT}$,

$$\left\|\delta^P - \delta_0^P\right\|_2 \le \left\|\delta^P - \delta_0^P\right\|_1 \le N^{-1/2}\zeta_{NT,\infty}, \tag{E.1}$$
$$\|\xi-\xi_0\|_2 \le \|\xi-\xi_0\|_1 \le \zeta_{NT,\infty}. \tag{E.2}$$

The Cauchy inequality gives

$$\left(p_i(X_{it}) - p_{i0}(X_{it})\right)^2 = \left(X_{it}'(\delta^P-\delta_0^P) + \xi_i - \xi_{i0}\right)^2 \le 2\left(X_{it}'(\delta^P-\delta_0^P)\right)^2 + 2(\xi_i-\xi_{i0})^2.$$

Summing over $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$ give

$$\mathbf{p}_{NT}^2 \le 2(NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathrm{E}\big(X_{it}'(\delta^P - \delta_0^P)\big)^2 + 2N^{-1}\|\xi - \xi_0\|^2$$

$$\le 2B_{\max} N^{-1} \zeta_{NT,\infty}^2 + 2N^{-1}\zeta_{NT,\infty}^2.$$

With probability $1 - o(1)$, $\max_{1\le i\le N, 1\le t\le T} \|X_{it}\|_\infty \le C_X \sqrt{\log d_X NT}$ for some finite C_X by Lemma B.1.

Step 2. The following bound holds w.p. $1 - o(1)$:

$$\begin{aligned}\sup_{it}\big|p_i(X_{it}) - p_{i0}(X_{it})\big| &\le \sup_{it}\big|X_{it}'(\delta^P - \delta_0^P)\big| + |\xi_i - \xi_{i0}| \\ &\le \sup_{it}\|X_{it}\|_\infty \big\|\delta^P - \delta_0^P\big\|_1 + \|\xi - \xi_0\|_1 \\ &\le C_X\sqrt{\log(d_X NT)} N^{-1/2}\zeta_{NT,\infty} + \zeta_{NT,\infty} \\ &\le 2\zeta_{NT,\infty},\end{aligned}$$

where the last step holds assuming N is large enough and $C_X\sqrt{\log(d_X NT)/N} \le 1$. □

Proof of Remark 5.5. Step 1 shows that $\mathbf{l}_{NT} = O(N^{-1/2}(\zeta_{NT,\infty} + \zeta_{NT,\infty}^E))$. Step 2 shows that w.p. $1 - o(1)$,

$$\sup_{it}\big|l_i(X_{it}) - l_{i0}(X_{it})\big| \le 2\bar{K}\|\beta_0\|_1 \zeta_{NT,\infty} + 2\zeta_{NT,\infty}^E.$$

Step 1. Decompose

$$l_i(X_{it}) - l_{i0}(X_{it}) = \big(d_i(X_{it}) - d_{i0}(X_{it})\big)'\beta_0 + X_{it}'\big(\delta^E - \delta_0^E\big) + \xi_i^E - \xi_{i0}^E + d_i(X_{it})'(\beta - \beta_0).$$

The Cauchy inequality gives

$$\begin{aligned}\big(l_i(X_{it}) - l_{i0}(X_{it})\big)^2 \le 4\Big(\big(\big(d_i(X_{it}) - d_{i0}(X_{it})\big)'\beta_0\big)^2 \\ + \big(X_{it}'\big(\delta^E - \delta_0^E\big)\big)^2 + \big(\xi_i^E - \xi_{i0}^E\big)^2 + \big(d_i(X_{it})'(\beta - \beta_0)\big)^2\Big).\end{aligned}$$

Note that $d_i(X_{it}) = K(X_{it})p_i(X_{it}) = K(X_{it})(X_{it}'\delta^P + \xi_i)$. Summing over $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$ give

$$\begin{aligned}\mathbf{l}_{NT}^2 \le\ & \underbrace{4\,(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \big(\delta^P - \delta_0^P\big)'\mathrm{E}\big[\big(K_{it}'\beta_0\big)^2 X_{it}X_{it}'\big]\big(\delta^P - \delta_0^P\big)}_{(\delta^P - \delta_0^P)'\Psi_D(\delta^P - \delta_0^P)} \\ & + 4(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \mathrm{E}\big(X_{it}'\big(\delta^E - \delta_0^E\big)\big)^2 + 4N^{-1}\big\|\xi^E - \xi_0^E\big\|_2^2 \\ & + 4(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \mathrm{E}\big\|d_i(X_{it})\big\|_\infty^2 \|\beta - \beta_0\|_1^2\end{aligned}$$

$$\le 4\big(B_{\max} N^{-1} \zeta_{NT,\infty}^2 + B_{\max} N^{-1} \big(\zeta_{NT,\infty}^E\big)^2 + N^{-1}\big(\zeta_{NT,\infty}^E\big)^2 + B_4 N^{-1}\big(\zeta_{NT,\infty}^E\big)^2\big)$$
$$\le 4(B_{\max} + 1 + B_4) N^{-1}\big(\zeta_{NT,\infty}^E\big)^2 + 4N^{-1} B_{\max} \zeta_{NT,\infty}^2.$$

Note that for N, T large enough such that $\|\delta^P\|_2^2 \le 2\|\delta_0^P\|_2^2 \le 2\|\delta_0^P\|_1^2$, which is bounded,

$$\begin{aligned} B_4 &= (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathrm{E}\big\| d_i(X_{it}) \big\|_\infty^2 \\ &\le \bar{K}^2 (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T \mathrm{E}\big(X_{it}' \delta^P + \xi_i\big)^2 \le 2\bar{K}^2\big(\delta^{P\prime} \Psi_X \delta^P + N^{-1}\|\xi\|^2\big) \\ &\le 2\bar{K}^2\big(B_{\max}\big\|\delta^P\big\|_2^2 + N^{-1}\|\xi\|_2^2\big) \\ &\le 4\bar{K}^2 B_{\max}\big\|\delta_0^P\big\|_2^2 + 1. \end{aligned}$$

Step 2. For N, T large enough, w.p. $1 - o(1)$,

$$\begin{aligned} &\sup_{it}\big| l_i(X_{it}) - l_{i0}(X_{it})\big| \\ &\quad \le \sup_{it}\big|K_{it}'\beta_0\big| \sup_{it}\big|p_i(X_{it}) - p_{i0}(X_{it})\big| \\ &\qquad + \sup_{it}\|X_{it}\|_\infty \big\|\delta^E - \delta_0^E\big\|_1 + \sup_i\big|\xi_i - \xi_i^E\big| \\ &\qquad + \sup_{it}\big\|K(X_{it})\big\|_\infty \big|X_{it}'\delta^E\big| \|\beta - \beta_0\|_1 \\ &\quad \le 2\bar{K}\|\beta_0\|_1 \zeta_{NT,\infty} + \|X_{it}\|_\infty\big(N^{-1/2}\zeta_{NT,\infty}^E + \bar{K}\big\|\delta^E\big\|_1 N^{-1/2}\zeta_{NT,\infty}^E\big) + \big\|\xi^E - \xi_0^E\big\|_1 \\ &\quad \le 2\bar{K}\|\beta_0\|_1 \zeta_{NT,\infty} + C_X\sqrt{\log(d_X NT)} N^{-1/2}\big(1 + \bar{K}\big\|\delta^E\big\|_1\big)\zeta_{NT,\infty}^E + \zeta_{NT,\infty}^E \\ &\quad \le 2\bar{K}\|\beta_0\|_1 \zeta_{NT,\infty} + 2\zeta_{NT,\infty}^E, \end{aligned}$$

where the last step holds assuming N is large enough and $\|\delta^E\|_1 \le \|\delta_0^E\|_1$ and

$$C_X\sqrt{\log(d_X NT)/N}\big(1 + 2\bar{K}\big\|\delta_0^E\big\|_1\big) \le 1. \qquad \square$$

Proof of Remark 5.6. Invoking Remark 5.3 and the bound (5.34) on $\zeta_{NT,\infty}^P$ in Lemma 5.1 give

$$\begin{aligned} \sqrt{NT}\mathbf{p}_{NT}^2 &\lesssim \sqrt{NT} N^{-1}\big(\zeta_{NT,\infty}^P\big)^2 \lesssim \big(S^P\big)^2 N^{-1/2} T^{1/2} T^{\nu-1} \log^{3(1-\nu)}(d_X + N) \\ &\lesssim \big(S^P\big)^2 N^{-1/2} T^{\nu-1/2} \log^{3(1-\nu)}(d_X + N) = o(1). \end{aligned}$$

In addition, the bound (5.36) on $\zeta_{NT,\infty}^E$ in Lemma 5.2 gives

$$\begin{aligned} \sqrt{NT}\mathbf{p}_{NT}\mathbf{l}_{NT} &\lesssim \sqrt{NT} N^{-1} \zeta_{NT,\infty}^P \zeta_{NT,\infty}^E \\ &\lesssim S^P \cdot S^E N^{-1/2} T^{(\nu+\nu^E)/2-1} \log^{3(1-(\nu+\nu^E)/2)}(d_X + N) = o(1). \qquad \square \end{aligned}$$

Appendix F: Tools: Tails bounds for empirical rectangular matrices under weak dependence

Lemma F.1 (Rectangular Matrix Bernstein, Theorem 1.6 in Tropp (2012)). *Consider a finite sequence $\{\Xi_m\}_{m=1}^M$ of independent, random matrices with dimensions $d_1 \times d_2$. Assume that there exist constants R_Ξ and σ_Ξ such that*

$$\mathrm{E}\Xi_m = 0, \qquad \|\Xi_m\| \leq R_\Xi \quad a.s. \tag{F.1}$$

Define

$$\sigma_\Xi^2 = \max\left(\left\|\mathrm{E}\sum_{m=1}^M \Xi_m'\Xi_m\right\|, \left\|\mathrm{E}\sum_{m=1}^M \Xi_m\Xi_m'\right\|\right). \tag{F.2}$$

Then, for all $t \geq 0$,

$$\mathrm{P}\left(\left\|\sum_{m=1}^M \Xi_m\right\| \geq t\right) \leq (d_1 + d_2)e^{-(t^2/2(\sigma_\Xi^2 + R_\Xi t/3))}. \tag{F.3}$$

Lemma F.2 (Tail Bounds for Weakly Dependent Sums, Operator Norm). *Consider the setup of Lemma B.4 with weakly dependent data $\{W_{it}\}$ and matrix-valued functions $\{\phi_i(\cdot)\}_{i=1}^N : \mathcal{W} \to \mathrm{R}^{d_1\times d_2}$. Let $q = \lfloor (2/\kappa)\log(NT)\rfloor$ be as in (B.16) and $L = \lfloor T/2q\rfloor$. For $i = 1, 2, \dots, N$ and $l = 1, 2, \dots, L$, let the data blocks $B_{i(2l-1)}$, B_{i2l} and B_{ir} be as in (B.8)–(B.10). Let the full-sized odd-block sums $\phi_i(B_{i(2l-1)})$ be as in (B.11), that is,*

$$\phi_i(B_{i(2l-1)}) = \sum_{t=(2l-2)q+1}^{t=(2l-2)q+q} \phi_i(W_{it}), \qquad \phi_i(B_{i(2l)}) = \sum_{t=(2l-1)q+1}^{t=(2lq)} \phi_i(W_{it})$$

and let $\phi_i(B^*_{i(2l-1)})$ and $\phi_i(B^*_{i(2l)})$ be their Berbee copies. In case $T \neq 2Lq$, the remainder block $\phi_i(B_{ir})$ as in (B.13), that is,

$$\phi_i(B_{ir}) := \sum_{t=2Lq+1}^{T} \phi_i(W_{it}).$$

Suppose that there exist constants R and σ such that the following conditions hold:

$$\mathrm{E}\phi_i(W_{it}) = 0, \qquad \sup_{it}\left\|\phi_i(W_{it})\right\| \leq R \quad a.s. \tag{F.4}$$

and

$$\max\left(\left\|\sum_{i=1}^N\sum_{l=1}^L \mathrm{E}\phi_i\big(B^*_{i(2l)}\big)'\phi_i\big(B^*_{i(2l)}\big)\right\|, \left\|\sum_{i=1}^N\sum_{l=1}^L \mathrm{E}\phi_i\big(B^*_{i(2l)}\big)\phi_i\big(B^*_{i(2l)}\big)'\right\|\right) \leq qNT\sigma^2, \tag{F.5}$$

$$\max\left(\left\|\sum_{i=1}^{N}\sum_{l=1}^{L}\mathrm{E}\phi_i\big(B^*_{i(2l-1)}\big)'\phi_i\big(B^*_{i(2l-1)}\big)\right\|, \left\|\sum_{i=1}^{N}\sum_{l=1}^{L}\mathrm{E}\phi_i\big(B^*_{i(2l-1)}\big)\phi_i\big(B^*_{i(2l-1)}\big)'\right\|\right) \le qNT\sigma^2, \tag{F.6}$$

$$\max\left(\left\|\sum_{i=1}^{N}\mathrm{E}\phi_i(B_{ir})\phi_i(B_{ir})'\right\|, \left\|\sum_{i=1}^{N}\mathrm{E}\phi_i(B_{ir})'\phi_i(B_{ir})\right\|\right) \le qNT\sigma^2. \tag{F.7}$$

Then, for any $t \ge 0$,

$$\mathrm{P}\left(\left\|(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi_i(W_{it})\right\| \ge 3t\right) \le 3(d_1+d_2)e^{-t^2NT/2(q\sigma^2+qRt/3)} + 2NL\gamma(q) \tag{F.8}$$

and under geometric beta-mixing condition (4.1),

$$\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi_i(W_{it})\right\| \lesssim_P \frac{1}{\sqrt{NT}}\left(\sigma\sqrt{\log(NT)\log(d_1+d_2)} + \frac{1}{\sqrt{NT}}\log(NT)R\log(d_1+d_2)\right). \tag{F.9}$$

Remark F.1. In what follows, we write $\phi(W_{it})$ in place of $\phi_i(W_{it})$, but subsume the dependence on i.

Proof of Lemma F.2. The union bound gives

$$\begin{aligned}\mathrm{P}\left(\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}\phi(W_{it})\right\| \ge 3t\right) &\le \mathrm{P}\left(\left\|\sum_{i=1}^{N}\sum_{l=1}^{L}\phi\big(B^*_{i(2l-1)}\big)\right\| \ge t\right) \\ &\quad + \mathrm{P}\left(\left\|\sum_{i=1}^{N}\sum_{l=1}^{L}\phi\big(B^*_{i(2l)}\big)\right\| \ge t\right) + \mathrm{P}\left(\left\|\sum_{i=1}^{N}\phi(B_{ir})\right\| \ge t\right) + 2NL\gamma(q).\end{aligned} \tag{F.10}$$

We first establish the bound for the odd-block sums. Define

$$m = m(i,l) = L\cdot(i-1)+l, \qquad M = NL, \qquad \Xi_m := \phi\big(B^*_{i(2l-1)}\big).$$

Since $\phi(B^*_{i(2l-1)})$ consists of q summands and W^*_{it} and W_{it} have the same marginal distributions, the bound (F.4) gives

$$\big\|\phi\big(B^*_{i(2l-1)}\big)\big\| \le qR \quad \text{a.s.},$$

which verifies (F.1) with $R_\Xi = qR$. Likewise, (F.5) directly verifies (F.2) with the bound $\sigma_\Xi^2 = qNT\sigma^2$. Invoking Lemma F.1 gives

$$\mathrm{P}\left(\left\|\sum_{i=1}^{N}\sum_{l=1}^{L}\phi\big(B^*_{i(2l-1)}\big)\right\| \geq t\right) \leq (d_1+d_2)e^{-t^2/2(qNT\sigma^2+qRt/3)}.$$

A similar bound holds for the even-numbered sums. For the remainder blocks, we take

$$m = i, \qquad M = N, \qquad \Xi_m = \phi(B_{ir}).$$

Since the remainder block has at most q elements,

$$\big\|\phi(B_{ir})\big\| \leq qR \quad \text{a.s.}$$

which implies (F.1) with $R_\Xi = qR$. Likewise, (F.7) directly verifies (F.2) with the bound $\sigma_\Xi^2 = qNT\sigma^2$. Therefore,

$$\mathrm{P}\left(\left\|\sum_{i=1}^{N}\phi(B_{ir})\right\| > t\right) \leq (d_1+d_2)e^{-t^2/2(qNT\sigma^2+qRt/3)}.$$

Invoking union bound (F.10) gives

$$\mathrm{P}\left(\left\|\sum_{i=1}^{N}\sum_{t=1}^{T}\phi(W_{it})\right\| \geq 3t\right) \leq 3(d_1+d_2)e^{-t^2/2(qNT\sigma^2+qRt/3)} + 2NL\gamma(q).$$

Plugging $t(NT)$ in place of t gives and dividing each side by NT gives

$$\begin{aligned}\mathrm{P}\left(\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi(W_{it})\right\| \geq 3t\right) &\leq 3(d_1+d_2)e^{-t^2(NT)^2/2(qNT\sigma^2+qRNTt/3)} + 2NL\gamma(q)\\ &= 3(d_1+d_2)e^{-t^2NT/2(q\sigma^2+qRt/3)} + 2NL\gamma(q),\end{aligned}$$

which coincides with (F.8). For geometric mixing, taking q as in (B.16) gives $NL\gamma(q) = o(1)$. Noting that

$$\begin{aligned}&3(d_1+d_2)e^{-t^2NT/2(q\sigma^2+qRt/3)}\\ &\quad\leq \max\big(3(d_1+d_2)e^{-(t^2NT/4q\sigma^2)}, 3(d_1+d_2)e^{-(tNT/4qR/3)}\big).\end{aligned}$$

Plugging $t = C'\sigma\sqrt{q\log(d_1+d_2)/NT}$ and taking C' large makes the first term in the max as small as desired. Plugging $t = C'\log(d_1+d_2)qR/NT$ and taking C' large makes the second terms in the max as small as desired. Therefore,

$$\begin{aligned}&\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi(W_{it})\right\|\\ &\quad\lesssim_P \frac{1}{\sqrt{NT}}\left(\sigma\sqrt{\log(NT)\log(d_1+d_2)} + \frac{1}{\sqrt{NT}}\log(NT)R\log(d_1+d_2)\right).\end{aligned}$$ □

LEMMA F.3. *Let $\gamma(X) : \mathcal{X} \to \mathrm{R}^{d_1 \times d_2}$ be a fixed matrix-valued function of a random vector X. Define the functional*

$$\phi(X) = \gamma(X) - \mathrm{E}\big[\gamma(X)\big]. \tag{F.11}$$

Let γ_{NT}^{∞} and γ_{NT} be numeric sequences obeying

$$\sup_{it} \big\|\gamma(X_{it})\big\| \leq \gamma_{NT}^{\infty} \quad a.s., \qquad (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathrm{E}\big\|\gamma(X_{it})\big\|^2 \leq \gamma_{2NT}^2. \tag{F.12}$$

Then the conditions (F.4) *and* (F.5)–(F.7) *hold with*

$$R = 2\gamma_{NT}^{\infty}, \qquad \sigma^2 = 2\gamma_{2NT}^2. \tag{F.13}$$

As a result, the bound (F.9) *in Lemma* F.2 *reduces to*

$$\left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi(X_{it}) \right\| \lesssim_P \frac{1}{\sqrt{NT}} \left(\gamma_{2NT} \sqrt{\log(NT) \log(d_1 + d_2)} + \frac{1}{\sqrt{NT}} \log(NT) \gamma_{NT}^{\infty} \log(d_1 + d_2) \right). \tag{F.14}$$

PROOF OF LEMMA F.3 . *Step 1.* Let X and $\bar{X}$ be two random vectors, and $\gamma(X)$ and $\gamma(\bar{X})$ be $d_1 \times d_2$ matrices. The following inequalities hold:

$$\begin{aligned} \big\|\mathrm{E}\gamma(X)\gamma(\bar{X})'\big\| &\leq^{\mathrm{i}} \mathrm{E}\big\|\gamma(X)\gamma(\bar{X})'\big\| \leq^{\mathrm{ii}} \mathrm{E}\big\|\gamma(X)\big\|\big\|\gamma(\bar{X})'\big\| \\ &\leq^{\mathrm{iii}} \sqrt{\mathrm{E}\big\|\gamma(X)\big\|^2 \mathrm{E}\big\|\gamma(\bar{X})'\big\|^2} \\ &\leq^{\mathrm{iv}} 1/2\big(\mathrm{E}\big\|\gamma(X)\big\|^2 + \mathrm{E}\big\|\gamma(\bar{X})'\big\|^2\big) =^{\mathrm{v}} 1/2\big(\mathrm{E}\big\|\gamma(X)\big\|^2 + \mathrm{E}\big\|\gamma(\bar{X})\big\|^2\big), \end{aligned} \tag{F.15}$$

where (i) follows from the convexity of the norm and Jensen's inequality, (ii) from submultiplicativity of operator norm $\|AB\| \leq \|A\|\|B\|$, (iii)–(iv) from Cauchy inequalities and (v) from $\|A'\| = \|A\|$. Likewise,

$$\begin{aligned} \big\|\mathrm{E}\gamma(X)\mathrm{E}\gamma(\bar{X})'\big\| &\leq^{\mathrm{i}} \big\|\mathrm{E}\gamma(X)\big\|\big\|\mathrm{E}\gamma(\bar{X})'\big\| \\ &\leq^{\mathrm{ii}} 1/2\Big(\big(\big\|\mathrm{E}\gamma(X)\big\|\big)^2 + \big(\big\|\mathrm{E}\gamma(\bar{X})'\big\|\big)^2\Big) \\ &\leq^{\mathrm{iii}} 1/2\big(\mathrm{E}\big\|\gamma(X)\big\|^2 + \mathrm{E}\big\|\gamma(\bar{X})'\big\|^2\big) =^{\mathrm{iv}} 1/2\big(\mathrm{E}\big\|\gamma(X)\big\|^2 + \mathrm{E}\big\|\gamma(\bar{X})\big\|^2\big), \end{aligned} \tag{F.16}$$

where (i) follows from $\|AB\| \leq \|A\|\|B\|$, (ii) from Cauchy inequality, (iii) from the convexity of composition $t \to t^2$ and $\cdot \to \|\cdot\|$ and Jensen's inequality, and (iv) from $\|A'\| = \|A\|$. Finally, since the R.H.S. of (F.15) and (F.16) is invariant under transposition, the same bound holds for the transposed quantities:

$$\max\big(\big\|\mathrm{E}\gamma(X)'\gamma(\bar{X})\big\|, \big\|\mathrm{E}\gamma(X)\gamma(\bar{X})'\big\|\big) \leq 1/2\big(\mathrm{E}\big\|\gamma(X)\big\|^2 + \mathrm{E}\big\|\gamma(\bar{X})\big\|^2\big),$$

$$\max\big(\big\|\mathrm{E}\gamma(X)'\mathrm{E}\gamma(\bar{X})\big\|, \big\|\mathrm{E}\gamma(X)\mathrm{E}\gamma(\bar{X})'\big\|\big) \leq 1/2\big(\mathrm{E}\big\|\gamma(X)\big\|^2 + \mathrm{E}\big\|\gamma(\bar{X})\big\|^2\big).$$

Step 2. For $\phi(X) = \gamma(X) - \mathrm{E}\gamma(X)$,

$$\begin{aligned}\mathrm{E}\phi(X)\phi(\bar{X})' &= \mathrm{E}\gamma(X)\gamma(\bar{X})' - \mathrm{E}\gamma(X)\mathrm{E}\gamma(\bar{X})' - \mathrm{E}\gamma(X)\mathrm{E}\gamma(\bar{X})' + \mathrm{E}\gamma(X)\mathrm{E}\gamma(\bar{X})' \\ &= \mathrm{E}\gamma(X)\gamma(\bar{X})' - \mathrm{E}\gamma(X)\mathrm{E}\gamma(\bar{X})'. \end{aligned} \tag{F.17}$$

Let $\{X_{mz}\}_{m,z=1}^{M,Z}$ be a double-indexed sequence. For every value of m,

$$\left(\sum_{z=1}^{Z}\gamma(X_{mz})\right)\left(\sum_{z'=1}^{Z}\gamma(X_{mz'})\right)' = \sum_{1\le z,z'\le Z}\gamma(X_{mz})\gamma(X_{mz'})'.$$

Define

$$M_1 := \mathrm{E}\sum_{m=1}^{M}\left(\sum_{z=1}^{Z}\gamma(X_{mz})\right)\left(\sum_{z=1}^{Z}\gamma(X_{mz'})\right)' = \sum_{m=1}^{M}\sum_{1\le z,z'\le Z}\mathrm{E}\gamma(X_{mz})\gamma(X_{mz'})',$$

$$M_2 := \sum_{m=1}^{M}\left(\mathrm{E}\sum_{z=1}^{Z}\gamma(X_{mz})\right)\left(\mathrm{E}\sum_{z=1}^{Z}\gamma(X_{mz'})\right)' = \sum_{m=1}^{M}\sum_{1\le z,z'\le Z}\mathrm{E}\gamma(X_{mz})\mathrm{E}\gamma(X_{mz'})'.$$

Invoking (F.17) gives

$$\sum_{m=1}^{M}\mathrm{E}\left[\sum_{z=1}^{Z}\phi(X_{mz})\right]\left[\sum_{z=1}^{Z}\phi(X_{mz'})\right]' = M_1 - M_2.$$

Step 3. The bound on $\|M_1\|$ is

$$\begin{aligned}\|M_1\| &\le \sum_{m=1}^{M}\sum_{1\le z,z'\le Z}\left\|\mathrm{E}\gamma(X_{mz})\gamma(X_{mz'})'\right\| \\ &\le 1/2\sum_{m=1}^{M}\sum_{z=1}^{Z}\sum_{z'=1}^{Z}\left(\mathrm{E}\left\|\gamma(X_{mz})\right\|^2 + \mathrm{E}\left\|\gamma(X_{mz'})\right\|^2\right) \\ &= Z/2\left(\sum_{m=1}^{M}\sum_{z=1}^{Z}\mathrm{E}\left\|\gamma(X_{mz})\right\|^2 + \sum_{m=1}^{M}\sum_{z'=1}^{Z}\mathrm{E}\left\|\gamma(X_{mz'})\right\|^2\right) \\ &= Z\sum_{m=1}^{M}\sum_{z=1}^{Z}\mathrm{E}\left\|\gamma(X_{mz})\right\|^2. \end{aligned} \tag{F.18}$$

Likewise,

$$\|M_2\| \le \sum_{m=1}^{M}\sum_{1\le z,z'\le Z}\left\|\mathrm{E}\gamma(X_{mz})\mathrm{E}\gamma(X_{mz'})'\right\|$$

$$\leq 1/2 \sum_{m=1}^{M} \sum_{z=1}^{Z} \sum_{z'=1}^{Z} \left(\mathrm{E}\left\|\gamma(X_{mz})\right\|^2 + \mathrm{E}\left\|\gamma(X_{mz'})\right\|^2\right)$$

$$= Z \sum_{m=1}^{M} \sum_{z=1}^{Z} \mathrm{E}\left\|\gamma(X_{mz})\right\|^2. \tag{F.19}$$

As a result,

$$\|M_1 - M_2\| \leq \|M_1\| + \|M_2\| \leq 2Z \sum_{m=1}^{M} \sum_{z=1}^{Z} \mathrm{E}\left\|\gamma(X_{mz})\right\|^2.$$

Because the bounds (F.15) and (F.16) are invariant to transpositions of $\gamma(X)$ and/or $\gamma(\bar{X})$,

$$\left\| \sum_{m=1}^{M} \mathrm{E}\left[\sum_{z=1}^{Z} \phi(X_{mz})'\right]\left[\sum_{z=1}^{Z} \phi(X_{mz'})\right] \right\| \leq 2Z \sum_{m=1}^{M} \sum_{z=1}^{Z} \mathrm{E}\left\|\gamma(X_{mz})\right\|^2. \tag{F.20}$$

Step 4. We first verify the condition (F.5) for the odd-numbered full-sized blocks. We note that the L.H.S. of (F.5) is a special case of the L.H.S. of (F.20) with

$$m = m(i, l) = L \cdot (i-1) + l, \qquad M = NL, \qquad Z = q,$$

$$X_{mz} := X_{i,(2l-2)q+z},$$

$$\phi(B_{i(2l-1)}) = \sum_{t=(2l-2)q+1}^{t=(2l-2)q+q} \phi(X_{it}) = \sum_{z=1}^{q} \phi(X_{mz}).$$

As a result,

$$\left\| \sum_{i=1}^{N} \sum_{l=1}^{L} \mathrm{E}\phi\big(B^*_{i(2l-1)}\big)\phi\big(B^*_{i(2l-1)}\big)' \right\| \leq 2q \sum_{i=1}^{N} \sum_{l=1}^{L} \sum_{z=1}^{q} \mathrm{E}\left\|\gamma\big(X^*_{i(2l-1),z}\big)\right\|^2 \tag{F.21}$$

$$= 2q \sum_{i=1}^{N} \sum_{l=1}^{L} \sum_{z=1}^{q} \mathrm{E}\left\|\gamma(X_{i(2l-1),z})\right\|^2 \tag{F.22}$$

$$\leq 2q \sum_{i=1}^{N} \sum_{t=1}^{T} \mathrm{E}\left\|\gamma(X_{it})\right\|^2 = 2qNT\gamma^2_{2NT}. \tag{F.23}$$

A similar argument for even-numbered, full-sized blocks and $\phi(B^*_{i(2l)}) = \sum_{t=(2l-1)q+1}^{t=(2l)q} \phi(X^*_{it})$ verifies condition (F.6) of Lemma F.2. Finally, if the remainder block is nonempty, that is, $T - 2Lq \neq 0$, invoking (F.20) with

$$m = i, \qquad M = N, \qquad Z := T - 2Lq$$

and noting that

$$\left\| \sum_{i=1}^{N} \mathrm{E}\phi(B_{ir})\phi(B_{ir})' \right\| \leq 2q \sum_{i=1}^{N} \sum_{z=1}^{T-2Lq} \mathrm{E}\left\| \gamma(X_{iz}) \right\|^2$$

$$\leq 2q \sum_{i=1}^{N} \sum_{t=1}^{T} \mathrm{E}\left\| \gamma(X_{it}) \right\|^2 = 2qNT\gamma_{2NT}^2, \tag{F.24}$$

which verifies condition (F.7) of Lemma F.2. Finally, the condition (F.4) follows from

$$\left\| \phi\left(B_{i(2l-1)}^*\right) \right\| \leq qR \quad \text{a.s.}, \qquad \left\| \phi\left(B_{i(2l)}^*\right) \right\| \leq qR \quad \text{a.s.}, \qquad \left\| \phi(B_{ir}) \right\| \leq qR,$$

since each block has at most q summands. Plugging $R = 2\gamma_{NT}^{\infty}$ and $q = 2\gamma_{2NT}^2$ into (F.9) gives (F.14). □

Corollaries F.1 and F.2 are special cases of Lemma F.3 with various cases of the γ-function.

COROLLARY F.1 (Covariance Matrix Moments). *Let $\psi(X) : \mathcal{X} \to \mathrm{R}^{d\times 1}$ be a fixed-vector function of a random vector X. Define*

$$\gamma(X) = \psi(X)\psi(X)' \tag{F.25}$$

and the ϕ-function

$$\phi(X) := \gamma(X) - \mathrm{E}\big[\gamma(X)\big] = \psi(X)\psi(X)' - \mathrm{E}\big[\psi(X)\psi(X)'\big].$$

Let the numeric sequences ψ_{NT}^{∞} and ψ_{4NT} obey

$$\sup_{it} \left\| \psi(X_{it}) \right\| \leq \psi_{NT}^{\infty} \quad a.s., \tag{F.26}$$

$$(NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathrm{E}\left\| \psi(X_{it}) \right\|^4 \leq \psi_{4NT}^4. \tag{F.27}$$

Then the bound (F.12) holds with $\gamma_{NT}^{\infty} := (\psi_{NT}^{\infty})^2$ and $\gamma_{2NT}^2 := \psi_{4NT}^4$. As a result, the rate (F.14) reduces to

$$\left\| (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi(X_{it}) \right\|$$

$$\lesssim_P \sqrt{\psi_{4NT}^4 \log(NT) \log(2d)/NT + \log(NT)\left(\psi_{NT}^{\infty}\right)^2 \log(2d)/NT}. \tag{F.28}$$

PROOF OF COROLLARY F.1. Noting that

$$\|\gamma(X_{it})\|_\infty \le \|\psi(X_{it})\|_\infty^2 \le (\psi_{NT}^\infty)^2$$

and

$$\|\gamma(X_{it})\|^2 = \|\psi(X_{it})\psi(X_{it})'\|^2 \le \|\psi(X_{it})\|^2 \|\psi(X_{it})'\|^2 = \|\psi(X_{it})\|^4.$$

Therefore,

$$(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \mathrm{E}\|\gamma(X_{it})\|^2 \le (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \mathrm{E}\|\psi(X_{it})\|^4 \le \psi_{4NT}^4.$$

Application of Lemma F.3 yields the result. □

COROLLARY F.2 (Product Moments). *Let $\psi(X): \mathcal{X} \to \mathrm{R}^{d\times 1}$ be a fixed-vector function of a random vector X and $\xi(X)$ be a random variable. Define*

$$\gamma(X) := \psi(X)\cdot \xi(X).$$

Let the numeric sequences ψ_{NT}^∞, ξ_{NT}^∞, and ψ_{4NT}, ξ_{4NT} obey

$$\sup_{it}\|\psi(X_{it})\| \le \psi_{NT}^\infty \quad a.s., \qquad \sup_{it}|\xi(X_{it})| \le \xi_{NT}^\infty \quad a.s., \tag{F.29}$$

$$(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \mathrm{E}\|\psi(X_{it})\|^4 \le \psi_{4NT}^4, \qquad (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \mathrm{E}\xi^4(X_{it}) \le \xi_{4NT}^4. \tag{F.30}$$

Then the bound (F.12) *holds with*

$$\gamma_{NT}^\infty := (\psi_{NT}^\infty)\cdot \xi_{NT}^\infty, \qquad \gamma_{2NT}^2 := 1/2(\psi_{4NT}^4 + \xi_{4NT}^4).$$

As a result, the rate (F.14) *reduces to*

$$\left\|(NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \phi(X_{it})\right\| \lesssim_P \sqrt{(\psi_{4NT}^4+\xi_{4NT}^4)\log(NT)\log(d+1)/NT} + \log(NT)\psi_{NT}^\infty\xi_{NT}^\infty\log(d+1)/NT. \tag{F.31}$$

PROOF OF COROLLARY F.2. Noting that

$$\|\gamma(X_{it})\|_\infty \le \|\psi(X_{it})\|_\infty |\xi(X_{it})| \le \psi_{NT}^\infty \xi_{NT}^\infty$$

and

$$\|\gamma(X_{it})\|^2 = \|\psi(X_{it})\|^2 \xi^2(X_{it}).$$

The Cauchy inequality gives

$$\mathrm{E}\|\gamma(X_{it})\|^2 = \mathrm{E}\|\psi(X_{it})\|^2\xi^2(X_{it}) \le 1/2\big(\mathrm{E}\|\psi(X_{it})\|^4 + \mathrm{E}\xi^4(X_{it})\big).$$

Therefore,

$$(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\|\gamma(X_{it})\|^2 \le (NT)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\|\psi(X_{it})\|^4 + \sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\xi^4(X_{it})\right)$$
$$\le (\psi_{4NT}^4 + \xi_{4NT}^4)/2.$$

Application of Lemma F.3 yields the result. □

Appendix G: Additional results on orthogonal OLS

Assumption G.1 (Tail Bound on Empirical Covariance Matrix in ℓ_2 Norm). *For some sequence $v_{NT} = o(1)$, in the regime where $d \to \infty$, we have that*

$$\|\widetilde{Q} - Q\| \lesssim_P v_{NT}. \tag{G.1}$$

Remark G.1. Suppose Assumptions 4.1–4.3 hold and $\sup_{it}\|V_{it}\|_\infty \le R$ a.s. and

$$\max_{itj}\mathrm{E}V_{itj}^4 \le \sigma_{4V}^4.$$

We invoke Corollary F.1 with $\psi(W_{it}) = V_{it}$ and $\psi_{NT}^\infty := \sqrt{d}R$ and (F.27) with $\psi_{4NT}^4 = d^2\sigma_{4V}^4$. As a result, the rate bound (F.28) reduces to

$$v_{NT} = \sqrt{d^2\log(2d)\log(NT)/NT} + dR\log(2d)\log(NT)/NT.$$

Further improvement of this rate may be possible under additional structure on V_{it}; see, for example, Theorem 1 and Corollary 3 in Banna, Merlevede, and Youssef (2016).

Let $D_{NT} \times L_{NT}$ be a sequence of realization sets such that the following conditions hold. Let $\mathbf{d}_{NT}, \mathbf{l}_{NT}, \mathbf{d}_{NT,4}, \mathbf{l}_{NT,4}$ be the numeric sequences obeying the following bounds:

$$\sup_{\mathbf{d}\in D_{NT}} (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathrm{E}\|d_i(X_{it}) - d_{i0}(X_{it})\|^2\right)^{1/2} \le \mathbf{d}_{NT},$$

$$\sup_{\mathbf{d}\in D_{NT}} (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathrm{E}\|d_i(X_{it}) - d_{i0}(X_{it})\|^4\right)^{1/4} \le \mathbf{d}_{NT,4},$$

$$\sup_{\mathbf{l}\in L_{NT}} (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathrm{E}\big(l_i(X_{it}) - l_{i0}(X_{it})\big)^4\right)^{1/4} \le \mathbf{l}_{NT,4}.$$

Define the following rates:

$$r_{2NT} := \mathbf{d}_{NT}\mathbf{l}_{NT} + \sqrt{\frac{(\mathbf{d}_{NT,4}^4 + \mathbf{l}_{NT,4}^4)\log(NT)\log(d+1)}{NT}}$$
$$+ \sqrt{d}\log(d+1)\log(NT)/NT, \tag{G.2}$$

$$\chi_{NT} := \mathbf{d}_{NT}^2 + \sqrt{\mathbf{d}_{NT,4}^4\log(2d)\log(NT)/NT} + d\log(2d)\log(NT)/NT. \tag{G.3}$$

ASSUMPTION G.2. *We suppose that the true parameter vector has bounded ℓ_2-norm:*

$$\|\beta_0\|_2 \leq \bar{C}_\beta$$

for some finite constant $\bar{C}_\beta$; We suppose that the reduced form estimators obey: $\widehat{\mathbf{l}}(\cdot) \in L_{NT}$ and $\widehat{\mathbf{d}}(\cdot) \in D_{NT}$ if such that $\mathbf{d}_{NT}, \mathbf{d}_{NT,4}, \mathbf{l}_{NT}, \mathbf{l}_{NT,4}$ decay sufficiently fast:

$$r_{2NT} + \chi_{NT} = o\big((NT)^{-1/2}\big). \tag{G.4}$$

Furthermore, the reduced form estimates are bounded as

$$\sup_{\mathbf{d} \in D_{NT}} \big\|d_i(X_{it})\big\| \leq \sqrt{d}D, \qquad \sup_{\mathbf{l} \in L_{NT}} \big|l_i(X_{it})\big| \leq L \quad \forall i.$$

THEOREM G.1 (Orthogonal Least Squares). *Suppose Assumptions* 4.1–4.3, G.1, *and* G.2 *hold. Then the following statements hold:*

1. *The orthogonal least squares estimator converges at the rate $\sqrt{d/NT}$:*

$$\|\widehat{\beta}_{\text{OLS}} - \beta_0\|_2 \lesssim_P \sqrt{d/NT}. \tag{G.5}$$

2. *For any deterministic sequence $\{\alpha\} = \{\alpha_{N,T}\}$ with $\|\alpha_{N,T}\| = 1$, the estimator $\alpha'\widehat{\beta}_{\text{OLS}}$ of $\alpha'\beta_0$ is asymptotically linear:*

$$\sqrt{NT}\alpha'(\widehat{\beta}_{\text{OLS}} - \beta_0) = \alpha' Q^{-1}\mathbb{G}_{NT} V_{it} U_{it} + o_P(1). \tag{G.6}$$

3. *If the Lindeberg condition holds for each $M > 0$,*

$$\limsup_{NT\to\infty} \sup_{\|\alpha\|_2=1} (NT)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} \mathrm{E}\big[\big(\alpha' V_{it} U_{it}\big)^2 1\big\{\big|\alpha' V_{it} U_{it}\big| > M\sqrt{NT}\big\}\big] = 0,$$

then the orthogonal least squares estimator is asymptotically Gaussian:

$$\lim_{NT\to\infty} \sup_{\|\alpha\|_2=1} \sup_{t\in\mathbb{R}} \left| \mathrm{P}\left(\frac{\sqrt{NT}\alpha'(\widehat{\beta}_{\text{OLS}} - \beta_0)}{\|\alpha'\Sigma\|^{1/2}} < t \right) - \Phi(t) \right| = 0. \tag{G.7}$$

LEMMA G.1 (First-Order Terms, ℓ_2-Norm). *Let $\bar{a}$, $\bar{m}$, $\bar{f}$, $\bar{e}$ be as in* (D.2)–(D.5). *Under Assumptions* 4.1–4.3, *the following bounds hold:*

$$\|\bar{a}\| \lesssim_P (\sqrt{d/NT}\mathbf{d}_{NT}), \tag{G.8}$$

$$\|\bar{m}\| \lesssim_P (\sqrt{d/NT}\mathbf{l}_{NT}), \tag{G.9}$$

$$\|\bar{f}\| \lesssim_P \big(\mathbf{d}_{NT}(NT)^{-1/2}\big), \tag{G.10}$$

$$\|\bar{e}\| \lesssim_P \big(\sqrt{d/NT}(\mathbf{d}_{NT} + \mathbf{l}_{NT})\big). \tag{G.11}$$

PROOF OF LEMMA G.1. Define

$$\xi^V_{NT} := \sqrt{d/NT}\mathbf{d}_{NT}, \qquad \xi^B_{NT} = 0,$$

and the A-function as

$$A(W_{it},\eta)=V_{it}\big(d_{i0}(X_{it})-d_i(X_{it})\big). \tag{G.12}$$

Define $B_{Ak}(\eta)$ and $V_{Ak}(\eta)$ with $\eta=\mathbf{d}$ as in (A.8)–(A.9). Consider any $\eta=\eta_{NT}\in D_{NT}$ in what follows. Since V_{it} obeys the martingale difference property (D.13), it follows that $\|B_{Ak}(\eta_{NT})\|=0$. Furthermore, for any $1\le j,j'\le d$,

$$\mathrm{E}\big[(\alpha' V_{it})(\alpha' V_{(it')})\big(d_{i0}(X_{it})-d_i(X_{it})\big)_j\big(d_{i0}(X_{it'})-d_i(X_{it'})\big)_{j'}\big]=0. \tag{G.13}$$

Combining (G.13) and Assumption 4.3,

$$\begin{aligned}
&\mathrm{E}\big[\|\alpha' V_{Ak}(\eta_{NT})\|^2\big]\\
&=^{\text{i}} (NT_k)^{-2}\sum_{i=1}^{N}\sum_{t\in\mathcal{M}_k}\sum_{j=1}^{d}\mathrm{E}\big[(\alpha' V_{it})^2\big(d_{i0}(X_{it})-d_i(X_{it})\big)_j^2\big]\\
&\le (NT_k)^{-2}\sup_{\mathbf{d}\in D_{NT}}\sum_{i=1}^{N}\sum_{t\in\mathcal{M}_k}\sum_{j=1}^{d}\mathrm{E}\big[\mathrm{E}\big[\|V_{it}\|^2|\Phi_{it},X_{it}\big]\big(d_{i0}(X_{it})-d_i(X_{it})\big)_j^2\big]\\
&\le (NT_k)^{-2}\sup_{\mathbf{d}\in D_{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\big\|d_{i0}(X_{it})-d_i(X_{it})\big\|^2 d\sigma_V^2\\
&\le^{\text{ii}} (d/NT_k)\sigma_V^2(T/T_k)\mathbf{d}_{NT}^2,
\end{aligned}$$

where (i) follows from (G.13) and (ii) follows from definition of $\mathbf{d}_{NT}$. By Assumption 4.5, we have that $\mathrm{P}(\widehat{\mathbf{d}}_k\in D_{NT},\forall k=1,\dots,K)\to 1$. Moreover, since the number of cross-fit folds is finite, the size T_k of each fold obeys

$$1\lesssim T_k/T\le 1.$$

We conclude by Lemma A.6 that (G.8) holds. Repeating the same argument for

$$A(W_{it},\eta)=V_{it}\big(l_{i0}(X_{it})-l_i(X_{it})\big)\quad\text{and}\quad A(W_{it},\eta)=U_{it}\big(d_{i0}(X_{it})-d_i(X_{it})\big)$$

establishes claims (G.9) and (G.10). Finally, (D.12) holds by definition of $\bar{e}=\bar{m}-\bar{a}'\beta_0$ and Holder inequalities. □

In the lemma below, abusing the notation, we treat l_i as some generic vector-valued function.

Lemma G.2 (Second-Order Bias). *Let $d_{i0}(X_{it})$ be a d_1-vector and $l_{i0}(X_{it})$ be a d_2-vector. Suppose that*

$$\sup_{\mathbf{d}\in D_{NT}}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\big(\mathrm{E}\big\|d_i(X_{it})-d_{i0}(X_{it})\big\|^2\big)^{1/2}\le \mathbf{d}_{NT},$$

$$\sup_{\mathbf{l}\in D_{NT}}(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\big(\mathrm{E}\big\|l_i(X_{it})-l_{i0}(X_{it})\big\|^2\big)^{1/2}\le \mathbf{l}_{NT}.$$

Consider the A-function as

$$A(W_{it},\eta)=\big(d_{i0}(X_{it})-d_i(X_{it})\big)\big(l_{i0}(X_{it})-l_i(X_{it})\big),\quad \eta=(\mathbf{d},\mathbf{l}) \tag{G.14}$$

and its bias $B_{Ak}(\eta)$ as in (A.8). *Then we have the bias bound*:

$$\sup_{\eta\in(D_{NT},L_{NT})}\|B_{Ak}(\eta)\|_2\le \mathbf{d}_{NT}\mathbf{l}_{NT}(T/T_k).$$

Proof of Lemma G.2. Take $\alpha\in\mathcal{S}^{d-1}$. Let $X_{it}(\alpha):=\alpha'(d_{i0}(X_{it})-d_i(X_{it}))$ and $Y_{itj}:=(l_{i0}(X_{it})-l_i(X_{it}))_j$ and

$$a_{it}^2=\|d_{i0}(X_{it})-d_i(X_{it})\|^2,\qquad b_{it}^2=\|l_{i0}(X_{it})-l_i(X_{it})\|^2=\sum_{j=1}^d b_{itj}^2.$$

Recognize that

$$\begin{aligned}\alpha' B_{Akj}(\eta)&=(NT_k)^{-1}\sum_{i=1}^N\sum_{t\in\mathcal{M}_k}\mathrm{E}\alpha'\big(d_{i0}(X_{it})-d_i(X_{it})\big)\big(l_{i0}(X_{it})-l_i(X_{it})\big)_j\\&=(NT_k)^{-1}\sum_{i=1}^N\sum_{t\in\mathcal{M}_k}\mathrm{E}X_{it}(\alpha)Y_{itj}.\end{aligned}$$

The Cauchy inequality gives

$$|\mathrm{E}X_{it}(\alpha)Y_{itj}|\le\sqrt{\mathrm{E}X_{it}^2(\alpha)\mathrm{E}Y_{itj}^2}\le\sqrt{\mathrm{E}\|d_{i0}(X_{it})-d_i(X_{it})\|^2\mathrm{E}Y_{itj}^2}=:\sqrt{a_{it}^2b_{itj}^2}.$$

Summing over i and t and invoking Cauchy inequality give

$$\begin{aligned}\alpha' B_{Akj}(\eta)&\le(NT_k)^{-1}\sum_{i=1}^N\sum_{t\in\mathcal{M}_k}\sqrt{a_{it}^2b_{itj}^2}\le(NT_k)^{-1}\sqrt{\sum_{i=1}^N\sum_{t\in\mathcal{M}_k}a_{it}^2}\sqrt{\sum_{i=1}^N\sum_{t\in\mathcal{M}_k}b_{itj}^2},\\ \|\alpha' B_{Ak}(\eta)\|^2&=\sum_{j=1}^d|\alpha' B_{Akj}(\eta)|^2\le(NT_k)^{-1}\Big(\sum_{i=1}^N\sum_{t\in\mathcal{M}_k}a_{it}^2\Big)(NT_k)^{-1}\Big(\sum_{j=1}^d\sum_{i=1}^N\sum_{t\in\mathcal{M}_k}b_{itj}^2\Big)\\&\le(NT_k)^{-1}\Big(\sum_{i=1}^N\sum_{t=1}^T a_{it}^2\Big)(NT_k)^{-1}\Big(\sum_{i=1}^N\sum_{t=1}^T b_{it}^2\Big)\\&\le \mathbf{l}_{NT}^2\mathbf{d}_{NT}^2(T/T_k)^2.\end{aligned}$$

□

Next, we invoke Lemmas F.2 and F.3 and Corollaries F.1–F.2 from Appendix F.

Lemma G.3 (Second-Order Covariance Term). *Define*

$$\zeta_{NT}^B=\mathbf{d}_{NT}^2,\qquad \zeta_{NT}^V=\sqrt{\mathbf{d}_{NT,4}^4\log(2d)\log NT/NT+dD\log(2d)\log(NT)/NT}.$$

Under Assumptions 4.1–4.3, *the following bounds hold for the term* $\bar{b}$ *defined in* (D.6):

$$\|\bar{b}\| \lesssim_P \zeta^B_{NT} + \zeta^V_{NT} = \chi_{NT}. \tag{G.15}$$

PROOF OF LEMMA G.3. Define the A-function as

$$A(W_{it},\eta) = \big(d_{i0}(X_{it}) - d_i(X_{it})\big)\big(d_{i0}(X_{it}) - d_i(X_{it})\big)', \quad \eta = \mathbf{d} = \mathbf{l}.$$

Let $B_{Ak}(\eta)$ and $V_{Ak}(\eta)$ be defined according to (A.8)–(A.9). Invoking Lemma G.2 with $\mathbf{l} = \mathbf{d}$ gives $\|B_{Ak}(\eta_{NT})\|_\infty = O(\zeta^B_{NT})$ for any partition k. Note that

$$\begin{aligned} V_{Ak}(\eta_{NT}) &= (NT_k)^{-1}\sum_{i=1}^N \sum_{t\in\mathcal{M}_k} \Big(\big(d_{i0}(X_{it}) - d_i(X_{it})\big)\big(d_{i0}(X_{it}) - d_i(X_{it})\big)' \\ &\quad - \mathrm{E}\big[\big(d_{i0}(X_{it}) - d_i(X_{it})\big)\big(d_{i0}(X_{it}) - d_i(X_{it})\big)'\big]\Big) \\ &=: (NT_k)^{-1}\sum_{i=1}^N \sum_{t\in\mathcal{M}_k} \phi_i(X_{it}). \end{aligned}$$

Define

$$\begin{aligned} \psi_i(X_{it}) &= \big(d_{i0}(X_{it}) - d_i(X_{it})\big), \\ \gamma_i(X_{it}) &= \psi_i(X_{it})\psi_i(X_{it})' = \big(d_{i0}(X_{it}) - d_i(X_{it})\big)\big(d_{i0}(X_{it}) - d_i(X_{it})\big)', \\ \phi_i(X_{it}) &= \gamma_i(X_{it}) - \mathrm{E}\big[\gamma_i(X_{it})\big]. \end{aligned}$$

Note that $\psi_i(X_{it}) = (d_{i0}(X_{it}) - d_i(X_{it}))$ obeys the conditions (F.26) and (F.27) with

$$\psi^\infty_{NT} := \sqrt{d}D, \qquad \psi_{NT,4} := \mathbf{d}_{NT,4}.$$

As a result, the bound (F.28) reduces to ζ^V_{NT} for each partition k and $T = T_k$. Since $T_k/T \asymp 1$, the bound follows. □

LEMMA G.4 (Second-Order Covariance Term, cont.). *Suppose Assumptions* 4.1–4.3 *and* G.2 *hold. Let* $\bar{z}$ *and* $\bar{g}$ *be as defined in* (D.7) *and* (D.8). *Then*

$$\|\bar{z}\| \lesssim_P r_{2NT}, \tag{G.16}$$

$$\|\bar{g}\| \lesssim_P r_{2NT} + \chi_{NT}. \tag{G.17}$$

PROOF OF LEMMA G.4. Define the A-function as

$$A(W_{it},\eta) = \big(d_{i0}(X_{it}) - d_i(X_{it})\big)\big(l_{i0}(X_{it}) - l_i(X_{it})\big), \quad \eta = (\mathbf{d},\mathbf{l}).$$

Let $B_{Ak}(\eta)$ and $V_{Ak}(\eta)$ be defined according to (A.8)–(A.9). Let

$$\begin{aligned} \zeta^B_{NT} &= \mathbf{d}_{NT}\mathbf{l}_{NT}, \\ \zeta^V_{NT} &= \sqrt{\frac{(\mathbf{d}^4_{NT,4} + \mathbf{l}^4_{NT,4})\log(NT)\log(d+1)}{NT}} + \sqrt{d}D\log(2d)\log(NT)/NT. \end{aligned}$$

Invoking Lemma G.2 with $d_1 = d$ and $d_2 = 1$ give $\|B_{Ak}(\eta_{NT})\|_\infty = O(\zeta^B_{NT})$ for any partition k. Note that

$$\begin{aligned} V_{Ak}(\eta_{NT}) &= (NT_k)^{-1} \sum_{i=1}^{N} \sum_{t \in \mathcal{M}_k} \Big(\big(d_{i0}(X_{it}) - d_i(X_{it})\big)\big(l_{i0}(X_{it}) - l_i(X_{it})\big) \\ &\quad - \mathrm{E}\big[\big(d_{i0}(X_{it}) - d_i(X_{it})\big)\big(l_{i0}(X_{it}) - l_i(X_{it})\big)'\big]\Big) \\ &=: (NT_k)^{-1} \sum_{i=1}^{N} \sum_{t \in \mathcal{M}_k} \phi_i(X_{it}). \end{aligned}$$

Define

$$\begin{aligned} \psi_i(X_{it}) &= \big(d_{i0}(X_{it}) - d_i(X_{it})\big), \quad \xi_i(X_{it}) = l_{i0}(X_{it}) - l_i(X_{it}), \\ \gamma_i(X_{it}) &= \psi_i(X_{it})\xi_i(X_{it}) = \big(d_{i0}(X_{it}) - d_i(X_{it})\big)\big(l_{i0}(X_{it}) - l_i(X_{it})\big), \\ \phi_i(X_{it}) &= \gamma_i(X_{it}) - \mathrm{E}\big[\gamma_i(X_{it})\big]. \end{aligned}$$

Note that $\psi_i(X_{it}) = (d_{i0}(X_{it}) - d_i(X_{it}))$ and $\xi_i(X_{it}) = l_{i0}(X_{it}) - l_i(X_{it})$ obey the conditions (F.29) and (F.30) with

$$\psi^\infty_{NT} := \sqrt{d}D, \qquad \psi_{NT,4} := \mathbf{d}_{NT,4}, \qquad \xi^\infty_{NT} := L, \qquad \xi_{NT,4} := \mathbf{l}_{NT,4}.$$

As a result, the bound (F.31) reduces to ζ^V_{NT} for each partition k and $T = T_k$. Since $T_k/T \asymp 1$, the bound (G.16) follows. Recognizing that $\bar{g} = \bar{z} - \bar{b}'\beta_0$ and invoking $\|\beta_0\| \leq C_\beta$ as in Assumption G.2 give

$$\|\bar{g}\| \leq \|\bar{z}\| + \|\bar{b}\beta_0\| \leq \|\bar{z}\| + \|\bar{b}\|\|\beta_0\|,$$

(G.17) follows. □

Proof of Theorem G.1. *Step 0.* Let $R_{it}(\widehat{d}, \widehat{l})$ be as defined in (D.1). Let $\bar{a}, \bar{b}, \bar{e}, \bar{f}, \bar{g}$ be as defined in (D.2), (D.6), ..., (D.8). As shown in the proof of Lemma D.3, the Gram matrix estimation error

$$\widehat{Q} - \widetilde{Q} = \mathbb{E}_{NT}\widehat{V}_{it}\widehat{V}_{it}' - \mathbb{E}_{NT}V_{it}V_{it}' = \bar{a} + \bar{a}' + \bar{b}$$

and gradient estimation error

$$\widehat{S} - S = \mathbb{E}_{NT}\widehat{V}_{it}\big(U_{it} + R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}})\big) - V_{it}U_{it}] = \bar{e} + \bar{f} + \bar{g}.$$

We have that

$$\|\widehat{Q} - Q\| \leq \|\widehat{Q} - \widetilde{Q}\| + \|\widetilde{Q} - Q\| \lesssim_P^{\mathrm{i}} (\chi_{NT} + \sqrt{d/NT}\mathbf{d}_{NT} + v_{NT}) = o(1),$$

where (i) follows from Lemmas G.1–G.4 and Assumption G.1.Furthermore, by lemmas

$$\|\widehat{S} - S\| = \|\bar{e} + \bar{f} + \bar{g}\| \lesssim_P (r_{2NT} + \chi_{NT}) = o(1/\sqrt{NT}),$$

where we used Assumptions 4.1–4.3 and G.2 to conclude that $r_{2NT} + \chi_{NT} = o(1/\sqrt{NT})$.

Step 1. Since Q is invertible by assumption, $\widehat{Q}$ is also invertible w.p. $1-o(1)$ by Step 0. Therefore, we can decompose $\widehat{\beta}_{\text{OLS}} - \beta_0$ as

$$\begin{aligned}
\widehat{\beta}_{\text{OLS}} - \beta_0 &= \widehat{Q}^{-1}\mathbb{E}_{NT}[\widehat{V}_{it}\widehat{\widetilde{Y}}_{it}] - \widehat{Q}^{-1}\widehat{Q}'\beta_0 = \widehat{Q}^{-1}\mathbb{E}_{NT}[\widehat{V}_{it}\widehat{\widetilde{Y}}_{it}] - \widehat{Q}^{-1}\big(\mathbb{E}_{NT}\widehat{V}_{it}\widehat{V}_{it}'\big)\beta_0 \\
&= \widehat{Q}^{-1}\mathbb{E}_{NT}\big[\widehat{V}_{it}\big(\widehat{\widetilde{Y}}_{it} - \widehat{V}_{it}'\beta_0\big)\big] \\
&= \widehat{Q}^{-1}\mathbb{E}_{NT}\big[\widehat{V}_{it}\big(U_{it} + R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}})\big)\big] \\
&= \widehat{Q}^{-1}\mathbb{E}_{NT}V_{it}U_{it} + \widehat{Q}^{-1}\mathbb{E}_{NT}\big[\widehat{V}_{it}\big(U_{it} + R_{it}(\widehat{\mathbf{d}}, \widehat{\mathbf{l}})\big) - V_{it}U_{it}\big].
\end{aligned}$$

Therefore, the following bound holds by triangle and Holder inequalities:

$$\|\widehat{\beta}_{\text{OLS}} - \beta_0\| \leq \big\|\widehat{Q}^{-1}\big\|\|\mathbb{E}_{NT}V_{it}U_{it}\| + \big\|\widehat{Q}^{-1}\big\|\|\widehat{S} - S\| =: \big\|\widehat{Q}^{-1}\big\|(L_1 + L_2).$$

The first term L_1 is bounded as

$$\begin{aligned}
\mathrm{E}\|\mathbb{E}_{NT}V_{it}U_{it}\|^2 &= \sum_{j=1}^{d}\mathrm{E}\big(\mathbb{E}_{NT}(V_{it})_jU_{it}\big)^2 \\
&=^{\mathrm{i}} (NT)^{-2}\sum_{j=1}^{d}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\big((V_{it})_jU_{it}\big)^2 \\
&\leq (NT)^{-2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\|V_{it}\|^2\sup_{it}\mathrm{E}\big[U_{it}^2 \mid V_{it}\big] \\
&\leq \bar{\sigma}^2(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{trace}\big(\mathrm{E}V_{it}V_{it}'\big) \\
&= \bar{\sigma}^2\,\mathrm{trace}(Q) \leq^{\mathrm{ii}} (d/NT)C_{\max},
\end{aligned}$$

where (i) follows from the m.d.s. property in Lemma B.3, and (ii) from $\max\mathrm{eig}(Q) \leq C_{\max}$. The Markov inequality gives $L_1 \lesssim_P (\sqrt{d/NT})$. The second term $L_2 := \|\widehat{S} - S\|$ is $o_P(1/\sqrt{NT})$ by Step 0. Step 0 implies $\max\mathrm{eig}(\widehat{Q}^{-1}) < 2C_{\min}^{-1}$ w.p. $1-o(1)$. Therefore, the rate bound (G.5) follows.

Step 2. From Step 1,

$$\begin{aligned}
\alpha'(\widehat{\beta}_{\text{OLS}} - \beta_0) &= \alpha'\widehat{Q}^{-1}\mathbb{E}_{NT}\big[\widehat{V}_{it}(U_{it} + R_{it}(\widehat{d}, \widehat{l})\big] \\
&= \alpha' Q^{-1}\mathbb{E}_{NT}V_{it}U_{it} \\
&\quad + \alpha'\big(\widehat{Q}^{-1} - Q^{-1}\big)\mathbb{E}_{NT}V_{it}U_{it} + \alpha'\widehat{Q}^{-1}\big[\mathbb{E}_{NT}\big[\widehat{V}_{it}(U_{it} + R_{it}(\widehat{d}, \widehat{l}) - V_{it}U_{it}\big]\big] \\
&=: \alpha' Q^{-1}\mathbb{E}_{NT}V_{it}U_{it} + S_1(\alpha) + S_2(\alpha).
\end{aligned}$$

The bound on $S_1(\alpha)$ follows:

$$\begin{aligned}
\big|S_1(\alpha)\big| &\leq \big\|\widehat{Q}^{-1} - Q^{-1}\big\|\|\mathbb{E}_{NT}V_{it}U_{it}\| \\
&\leq \big\|\widehat{Q}^{-1}\big\|\|\widehat{Q} - Q\|\big\|Q^{-1}\big\|\|\mathbb{E}_{NT}V_{it}U_{it}\|
\end{aligned}$$

$$= O_P(1) \cdot o_P(1) \cdot O_P(1) \cdot O_P\big((NT)^{-1/2}\big) = o_P\big((NT)^{-1/2}\big),$$

where $O_P(\cdot)$ and $o_P(\cdot)$ bounds are established in Steps 0–2. The bound on $S_2(\alpha)$ follows from:

$$\big|S_2(\alpha)\big| \leq \|\alpha\| C_{\min}^{-1} \|\widehat{S} - S\| \lesssim_P (r_{2NT} + \chi_{NT}) = o_P\big((NT)^{-1/2}\big),$$

where we are using the results of Step 0. As a result,

$$\sqrt{NT}\alpha'(\widehat{\beta}_{\text{OLS}} - \beta_0) = \alpha' Q^{-1} \mathbb{G}_{NT} V_{it} U_{it} + o_P(1),$$

which gives (G.6).

Step 3. The proof of pointwise normality follows similar to Step 1 of the proof of Theorem 4.2, where the step (D.51) is replaced by

$$\big|\alpha'\big(\alpha'\Sigma\alpha\big)^{-1/2} R_{NT}\big| \leq \|\alpha\|_2 O(1) \|R_{NT}\|_2 \lesssim_P \sqrt{NT}(r_{2NT} + \chi_{NT}) = o_P(1). \qquad \square$$

References

Banna, Marwa, Florence Merlevede, and Pierre Youssef (2016), "Bernstein-type inequality for a class of dependent random matrices." *Random Matrices: Theory and Applications*, 5, 1650006. [59]

Cai, Tony, Weidong Liu, and Xi Luo (2011), "A constrained l_1 minimization approach to sparse precision matrix estimation." *Journal of the American Statistical Association*, 106, 594–607. [33, 34]

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato (2015), "Comparison and anti-concentration bounds for maxima of Gaussian random vectors." *Probability Theory and Related Fields*, 162, 47–70. [19, 46]

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. [19, 20]

Chernozhukov, Victor, Denis Chetverikov, Kengo Kato, and Yuta Koike (2019), "Improved central limit theorem and bootstrap approximations in high dimensions." [19]

Dudley, Richard Mansfield and Walter Philipp (1983), "Invariance principles for sums of Banach space valued random elements and empirical processes." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 62, 509–552. [3]

Fan, Xiequan, Ion Grama, and Quansheng Liu (2012), "Large deviation exponential inequalities for supermartingales." *Electronic Communications in Probability*, 17, 1–8. doi:10.1214/ECP.v17-2318. [14]

Hansen, Bruce (2019), "A weak law of large numbers under weak mixing." [37]

Kock, Anders Bredahl, and Haihan Tang (2019), "Uniform inference in high-dimensional dynamic panel data models." *Econometric Theory*, 35, 295–359. [12, 14, 15]

Ledoux, Michel (2001), *Concentration of Measure Phenomenon*. American Mathematical Society. [46]

Lounici, Karim, Massimiliano Pontil, Sara van de Geer, and Alexandre B. Tsybakov (2011), "Oracle inequalities and optimal inference under group sparsity." *Annals of Statistics*, 39, 2164–2204. doi:10.1214/11-AOS896. [27]

McLeish, Donald Leslie (1974), "Dependent central limit theorems and invariance principles." *Annals of Probability*, 2, 620–628. [37]

Negahban, Sahand, Pradeep Ravikumar, Martin James Wainwright, and Bin Yu (2012), "A unified framework for high-dimensional analysis of m-estimators with decomposable regularizers." *Statistical Science*, 27, 538–557. [30]

Rigollet, Philippe and Jan-Christian Hutter (2017), "High-dimensional statistics." [8]

Schwarz, Gideon (1980), "Finitely determined processes—an indiscrete approach." *Journal of Mathematical Analysis and Applications*, 76, 146–158. [2]

Tropp, Joel Aaron (2012), "User-Friendly Tail Bounds for Sums of Random Matrices." *Foundations Computational Mathematics*, 12, 389–434. [51]

van der Vaart, Aad and Jon August Wellner (1996), *Weak Convergence and Empirical Processes*. Springer. [16, 17, 48]

Villani, Cédric (2007), *Topics in Optimal Transportation*, Vol. 58. American Mathematical Soc. [2]

Wainwright, Martin James (2019), *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. [9, 10]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 27 June, 2020; final version accepted 3 December, 2022; available online 20 December, 2022.