Local Interaction Autoregressive Model for High-Dimensional Time Series Data

Jingyang Li[1], Yang Chen[2]

[1] *Fudan University;* [2] *University of Michigan, Ann Arbor*

Supplementary Material

This supplementary material contains proofs of the main theorems, auxiliary lemmas, and additional methodological and numerical details. It is organized as follows. Section S1 presents the proofs of the main theoretical results in the order in which they appear in the main paper, including the coefficient estimation theory, the separable estimator, auto-covariance estimation, and neighborhood-selection consistency. Section S2 collects the auxiliary probabilistic and matrix-analytic lemmas used throughout the proofs. Section S3 provides additional methodological and numerical details, including the non-identifiability issue for non-nested neighborhoods, estimation and bandwidth selection for the general lag-P model, and supplementary simulation results. Section S4 describes the tensor extension of the proposed LIAR framework and presents the tensor TEC application, including the tensor neighborhood-selection results and prediction comparison.

S1 Proofs of Main Theorems

The auxiliary lemmas used in the proofs are collected in Section S2. We present the proofs of the main theorems in the same order as they appear in the main text.

S1.1 Proof of Theorem 1

The proof is the one-block version of the martingale central limit theorem argument used in the proof of Theorem 2. Since Theorem 1 also follows immediately from the joint result in Theorem 2, we present the short marginal derivation here for completeness.

By Theorem 2,

$$\sqrt{T}\,\mathsf{vecstack}\{\widehat{\boldsymbol{m}}_{\boldsymbol{i}} - \boldsymbol{m}_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{S}\} \Rightarrow N(0, \boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{D}^{-1}).$$

Taking the block corresponding to a fixed location $\boldsymbol{i}$, the limiting covariance is

$$\boldsymbol{\Gamma}_0(\mathcal{J}_{\boldsymbol{i}};\mathcal{J}_{\boldsymbol{i}})^{-1}\{[\boldsymbol{\Sigma}]_{\boldsymbol{i},\boldsymbol{i}}\boldsymbol{\Gamma}_0(\mathcal{J}_{\boldsymbol{i}};\mathcal{J}_{\boldsymbol{i}})\}\boldsymbol{\Gamma}_0(\mathcal{J}_{\boldsymbol{i}};\mathcal{J}_{\boldsymbol{i}})^{-1} = \sigma_{\boldsymbol{i}}^2\boldsymbol{\Gamma}_0(\mathcal{J}_{\boldsymbol{i}};\mathcal{J}_{\boldsymbol{i}})^{-1},$$

where $[\boldsymbol{\Sigma}]_{\boldsymbol{i},\boldsymbol{i}} = \mathrm{Var}([\boldsymbol{E}_t]_{\boldsymbol{i}}) = \sigma_{\boldsymbol{i}}^2$. Hence

$$\sqrt{T}(\widehat{\boldsymbol{m}}_{\boldsymbol{i}} - \boldsymbol{m}_{\boldsymbol{i}}) \Rightarrow N\left(0, \sigma_{\boldsymbol{i}}^2\boldsymbol{\Gamma}_0(\mathcal{J}_{\boldsymbol{i}};\mathcal{J}_{\boldsymbol{i}})^{-1}\right),$$

which proves Theorem 1.

S1.2 Proof of Theorem 2

Proof. For each $\boldsymbol{i}$, $\widehat{\boldsymbol{m}_{\boldsymbol{i}}}$ admits the following closed-form solution:

$$\widehat{\boldsymbol{m}_{\boldsymbol{i}}} = (\boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{Y}_{\boldsymbol{i}})^{-1} \boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{z}_{\boldsymbol{i}} = \boldsymbol{m}_{\boldsymbol{i}} + (\boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{Y}_{\boldsymbol{i}})^{-1} \boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{\nu}_{\boldsymbol{i}}.$$

Therefore,

$$\mathsf{vecstack}\{\widehat{\boldsymbol{m}_{\boldsymbol{i}}} - \boldsymbol{m}_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{S}\} = \mathsf{vecstack}\{(\boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{Y}_{\boldsymbol{i}})^{-1} \boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{\nu}_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{S}\}.$$

Let $\boldsymbol{\gamma} = \mathsf{vecstack}\{\boldsymbol{\gamma}_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{S}\}$, then we have

$$\begin{aligned} T^{-1/2} \boldsymbol{\gamma}^\top \mathsf{vecstack}\{\boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{\nu}_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{S}\} &= T^{-1/2} \sum_{\boldsymbol{i}} \boldsymbol{\gamma}_{\boldsymbol{i}}^\top \boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{\nu}_{\boldsymbol{i}} \\ &= \sum_{t=2}^{T} \underbrace{T^{-1/2} \sum_{\boldsymbol{i}} \boldsymbol{\gamma}_{\boldsymbol{i}}^\top \boldsymbol{x}_{t-1}^{(\boldsymbol{i})} [\boldsymbol{E}_t]_{\boldsymbol{i}}}_{:= Z_{t,T}}. \end{aligned}$$

Let $\mathcal{A}_{t-1}$ denote the σ-field generated by previous innovations. Then

$$\mathbb{E}(Z_{t,T} \mid \mathcal{A}_{t-1}) = 0,$$

and

$$\mathbb{E}(Z_{t,T}^2 \mid \mathcal{A}_{t-1}) = T^{-1} \sum_{\boldsymbol{i},\boldsymbol{j} \in \mathcal{S}} \boldsymbol{\Sigma}_{\boldsymbol{i},\boldsymbol{j}} \boldsymbol{\gamma}_{\boldsymbol{i}}^\top \boldsymbol{x}_{t-1}^{(\boldsymbol{i})} \boldsymbol{x}_{t-1}^{(\boldsymbol{j})\top} \boldsymbol{\gamma}_{\boldsymbol{j}}.$$

Denote

$$V_{TT}^2 = T^{-1} \sum_{t=2}^{T} \sum_{\boldsymbol{i},\boldsymbol{j} \in \mathcal{S}} \boldsymbol{\Sigma}_{\boldsymbol{i},\boldsymbol{j}}\, \boldsymbol{\gamma}_{\boldsymbol{i}}^\top\, \boldsymbol{x}_{t-1}^{(\boldsymbol{i})} \boldsymbol{x}_{t-1}^{(\boldsymbol{j})\top}\, \boldsymbol{\gamma}_{\boldsymbol{j}}.$$

On the other hand, we have

$$s_{TT} := \mathbb{E}\Big(\sum_{t=2}^{T} Z_{t,T} \Big)^2 = T^{-1}(T-1) \sum_{\boldsymbol{i},\boldsymbol{j}} \boldsymbol{\Sigma}_{\boldsymbol{i},\boldsymbol{j}}\, \boldsymbol{\gamma}_{\boldsymbol{i}}^\top\, \boldsymbol{\Gamma}_0(\mathcal{J}_{\boldsymbol{i}}; \mathcal{J}_{\boldsymbol{j}})\, \boldsymbol{\gamma}_{\boldsymbol{j}}.$$

Then from Lemma 6 (under Regime 1), or from Lemma 2 (under Regime 2 or 3), we have $V_{TT}s_{TT}^{-1} \xrightarrow{p} 1$. Finally, for all $\epsilon > 0$, we can show using the moment condition of $[\boldsymbol{E}_t]_{\boldsymbol{i}}$,

$$\lim_{T\to\infty} s_{TT}^{-2} \sum_{t=2}^{T} \mathbb{E}\big(Z_{t,T}^2 \cdot 1(|Z_{t,T}| \geq \epsilon s_{TT} \big| \mathcal{A}_{t-1})\big) = 0.$$

So we conclude from Theorem 5.3.4 in Fuller (2009),

$$T^{-1/2}\mathsf{vecstack}\{\boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{\nu}_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{S}\} \implies N(0, \boldsymbol{C}).$$

From Lemma 6 (under Regime 1), or from Lemma 2 (under Regime 2 or 3), we have $\frac{1}{T}\boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{Y}_{\boldsymbol{i}} \xrightarrow{p} \boldsymbol{\Gamma}_0(\mathcal{J}_{\boldsymbol{i}}; \mathcal{J}_{\boldsymbol{i}})$. So we conclude

$$\sqrt{T}\mathsf{vecstack}\{\widehat{\boldsymbol{m}}_{\boldsymbol{i}} - \boldsymbol{m}_{\boldsymbol{i}} : \boldsymbol{i} \in \mathcal{S}\} \Rightarrow N(0, \boldsymbol{D}^{-1}\boldsymbol{C}\boldsymbol{D}^{-1}).$$

where $\boldsymbol{D} = \mathsf{blkdiag}\{\boldsymbol{\Gamma}_0(\mathcal{J}_{\boldsymbol{i}}; \mathcal{J}_{\boldsymbol{i}}) : \boldsymbol{i} \in \mathcal{S}\}$. □

S1.3 Proof of Theorem 3

The proof of this theorem relies on Lemma 5. Since $\widehat{\boldsymbol{M}}_R^{\mathsf{blk}}$ is the best rank R approximation of $\widehat{\boldsymbol{M}}^{\mathsf{blk}}$. And we have $\|\widehat{\boldsymbol{M}}^{\mathsf{blk}} - \boldsymbol{M}^{\mathsf{blk}}\|_{\mathrm{F}} \to 0$ as $T \to \infty$, so the condition in Lemma 5 is satisfied for sufficiently large T. So we conclude

$$\begin{aligned}\widehat{\boldsymbol{M}}_R^{\mathsf{blk}} - \boldsymbol{M}^{\mathsf{blk}} = \widehat{\boldsymbol{M}}^{\mathsf{blk}} - \boldsymbol{M}^{\mathsf{blk}} - \boldsymbol{U}_\perp^{\mathsf{blk}}\boldsymbol{U}_\perp^{\mathsf{blk}\top}(\widehat{\boldsymbol{M}}^{\mathsf{blk}} - \boldsymbol{M}^{\mathsf{blk}}) \\ \times \boldsymbol{V}_\perp^{\mathsf{blk}}\boldsymbol{V}_\perp^{\mathsf{blk}\top} + \boldsymbol{R},\end{aligned}$$

where $\|\boldsymbol{R}\|_{\mathrm{F}} \leq \frac{40\|\widehat{\boldsymbol{M}}^{\mathsf{blk}} - \boldsymbol{M}^{\mathsf{blk}}\|_{\mathrm{F}}^2}{\lambda_{\min}(\boldsymbol{M}^{\mathsf{blk}})}$, and thus $\sqrt{T}\|\boldsymbol{R}\|_{\mathrm{F}} = o_p(1)$. Next we vectorize both sides and we obtain

$$\begin{aligned}\mathsf{vec}(\widehat{\boldsymbol{M}}_R^{\mathsf{blk}} - \boldsymbol{M}^{\mathsf{blk}}) = \big(\boldsymbol{I} - \boldsymbol{V}_\perp^{\mathsf{blk}}\boldsymbol{V}_\perp^{\mathsf{blk}\top} \otimes \boldsymbol{U}_\perp^{\mathsf{blk}}\boldsymbol{U}_\perp^{\mathsf{blk}\top}\big) \\ \times \mathsf{vec}(\widehat{\boldsymbol{M}}^{\mathsf{blk}} - \boldsymbol{M}^{\mathsf{blk}}) + \mathsf{vec}(\boldsymbol{R}).\end{aligned}$$

Together with Theorem 2, we obtain the result.

S1.4 Proof of Theorem 4

Proof. The result is an immediate consequence of the first part of Lemma 2. Indeed, Lemma 2 implies that, under Assumption 1,

$$\max_{\boldsymbol{i},\boldsymbol{j}\in\mathcal{S}} \left|[\widehat{\boldsymbol{\Gamma}}_0]_{\boldsymbol{i},\boldsymbol{j}} - [\boldsymbol{\Gamma}_0]_{\boldsymbol{i},\boldsymbol{j}}\right| = O_p\left\{\left(\frac{\log(M\vee N\vee T)}{T}\right)^{1/2} \mu_{2q}^2 J_\square (1-\delta)^{-4}\right\}.$$

Since μ_{2q} and δ are treated as constants, and since

$$\|\widehat{\boldsymbol{\Gamma}}_0 - \boldsymbol{\Gamma}_0\|_{\ell_\infty} = \max_{\boldsymbol{i},\boldsymbol{j}\in\mathcal{S}} \left|[\widehat{\boldsymbol{\Gamma}}_0]_{\boldsymbol{i},\boldsymbol{j}} - [\boldsymbol{\Gamma}_0]_{\boldsymbol{i},\boldsymbol{j}}\right|,$$

we obtain

$$\|\widehat{\boldsymbol{\Gamma}}_0 - \boldsymbol{\Gamma}_0\|_{\ell_\infty} = O_p\left\{\left(\frac{\log(M\vee N\vee T)}{T}\right)^{1/2} J_\square\right\}.$$

This proves Theorem 4. □

S1.5 Proof of Theorem 5

Proof. For each $\boldsymbol{i} \in \mathcal{S}$, recall we have $\mathcal{J}_{\boldsymbol{i}}[k_0] = \mathcal{J}_{\boldsymbol{i}}$. And $\{\widehat{k}_{\boldsymbol{i}} \neq k_0\} = \{\widehat{k}_{\boldsymbol{i}} < k_0\} \cup \{\widehat{k}_{\boldsymbol{i}} > k_0\}$. For $k < k_0$, we have

$$\begin{aligned}
\mathsf{RSS}_{\boldsymbol{i}}(k) &= \boldsymbol{z}_{\boldsymbol{i}}^\top (\boldsymbol{I} - \boldsymbol{Y}_{\boldsymbol{i}}[k](\boldsymbol{Y}_{\boldsymbol{i}}[k]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k])^{-1} \boldsymbol{Y}_{\boldsymbol{i}}[k]^\top) \boldsymbol{z}_{\boldsymbol{i}} \\
&= (\boldsymbol{Y}_{\boldsymbol{i}}[k_0]\boldsymbol{m}_{\boldsymbol{i}} + \boldsymbol{\nu}_{\boldsymbol{i}})^\top (\boldsymbol{I} - \boldsymbol{Y}_{\boldsymbol{i}}[k](\boldsymbol{Y}_{\boldsymbol{i}}[k]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k])^{-1} \boldsymbol{Y}_{\boldsymbol{i}}[k]^\top) \\
&\quad \times (\boldsymbol{Y}_{\boldsymbol{i}}[k_0]\boldsymbol{m}_{\boldsymbol{i}} + \boldsymbol{\nu}_{\boldsymbol{i}}).
\end{aligned}$$

Since $k < k_0$, $\boldsymbol{Y}_{\boldsymbol{i}}[k]$ is a submatrix of $\boldsymbol{Y}_{\boldsymbol{i}}[k_0]$, and we can split $\boldsymbol{Y}_{\boldsymbol{i}}[k_0]\boldsymbol{m}_{\boldsymbol{i}}$ as

$$\boldsymbol{Y}_{\boldsymbol{i}}[k_0]\boldsymbol{m}_{\boldsymbol{i}} \;=\; \boldsymbol{Y}_{\boldsymbol{i}}[k]\boldsymbol{b}_1 \;+\; \boldsymbol{S}\boldsymbol{b}_2,$$

where $\|\boldsymbol{b}_2\|_{\ell_2} = \|\boldsymbol{M}_{\boldsymbol{i}}(\mathcal{J}_{\boldsymbol{i}} \backslash \mathcal{J}_{\boldsymbol{i}}[k])\|_{\mathrm{F}}$. Let $\mathcal{I}_1 := \mathcal{J}_{\boldsymbol{i}}[k]$ and $\mathcal{I}_2 := \mathcal{J}_{\boldsymbol{i}} \backslash \mathcal{J}_{\boldsymbol{i}}[k]$. Define

$$\widehat{\boldsymbol{G}} := \frac{1}{T-1} \begin{bmatrix} \boldsymbol{Y}_{\boldsymbol{i}}[k]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k] & \boldsymbol{Y}_{\boldsymbol{i}}[k]^\top \boldsymbol{S} \\ \boldsymbol{S}^\top \boldsymbol{Y}_{\boldsymbol{i}}[k] & \boldsymbol{S}^\top \boldsymbol{S} \end{bmatrix} := \begin{bmatrix} \widehat{\boldsymbol{\Gamma}}_0(\mathcal{I}_1;\mathcal{I}_1) & \widehat{\boldsymbol{\Gamma}}_0(\mathcal{I}_1;\mathcal{I}_2) \\ \widehat{\boldsymbol{\Gamma}}_0(\mathcal{I}_2;\mathcal{I}_1) & \widehat{\boldsymbol{\Gamma}}_0(\mathcal{I}_2;\mathcal{I}_2) \end{bmatrix}.$$

Then we denote

$$\boldsymbol{G} := \mathbb{E}\widehat{\boldsymbol{G}} = \begin{bmatrix} \boldsymbol{\Gamma}_0(\mathcal{I}_1;\mathcal{I}_1) & \boldsymbol{\Gamma}_0(\mathcal{I}_1;\mathcal{I}_2) \\ \boldsymbol{\Gamma}_0(\mathcal{I}_2;\mathcal{I}_1) & \boldsymbol{\Gamma}_0(\mathcal{I}_2;\mathcal{I}_2) \end{bmatrix}.$$

Since $\big(\boldsymbol{I} - \boldsymbol{Y}_{\boldsymbol{i}}[k](\boldsymbol{Y}_{\boldsymbol{i}}[k]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k])^{-1} \boldsymbol{Y}_{\boldsymbol{i}}[k]^\top\big) \boldsymbol{Y}_{\boldsymbol{i}}[k_0] = 0$, we have

$$\mathsf{RSS}_{\boldsymbol{i}}(k) = (\boldsymbol{S}\boldsymbol{b}_2 + \boldsymbol{\nu}_{\boldsymbol{i}})^\top (\boldsymbol{I} - \boldsymbol{Y}_{\boldsymbol{i}}[k](\boldsymbol{Y}_{\boldsymbol{i}}[k]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k])^{-1} \boldsymbol{Y}_{\boldsymbol{i}}[k]^\top)(\boldsymbol{S}\boldsymbol{b}_2 + \boldsymbol{\nu}_{\boldsymbol{i}}).$$

On the other hand, we have

$$\mathsf{RSS}_{\boldsymbol{i}}(k_0) = \boldsymbol{\nu}_{\boldsymbol{i}}^\top (\boldsymbol{I} - \boldsymbol{Y}_{\boldsymbol{i}}[k_0](\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1}\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top)\boldsymbol{\nu}_{\boldsymbol{i}}.$$

Therefore,

$$\begin{aligned}\mathsf{RSS}_{\boldsymbol{i}}(k) - \mathsf{RSS}_{\boldsymbol{i}}(k_0) \geq\ & \boldsymbol{b}_2^\top \boldsymbol{S}^\top (\boldsymbol{I} - \boldsymbol{Y}_{\boldsymbol{i}}[k](\boldsymbol{Y}_{\boldsymbol{i}}[k]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k])^{-1}\boldsymbol{Y}_{\boldsymbol{i}}[k]^\top)\boldsymbol{S}\boldsymbol{b}_2 \\ & + 2\boldsymbol{b}_2^\top \boldsymbol{S}^\top (\boldsymbol{I} - \boldsymbol{Y}_{\boldsymbol{i}}[k](\boldsymbol{Y}_{\boldsymbol{i}}[k]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k])^{-1}\boldsymbol{Y}_{\boldsymbol{i}}[k]^\top)\boldsymbol{\nu}_{\boldsymbol{i}}.\end{aligned}$$

We denote $\widehat{\boldsymbol{M}} = \widehat{\boldsymbol{G}}^{-1}$, $\boldsymbol{M} = \boldsymbol{G}^{-1}$. And $\widehat{\boldsymbol{M}}$ is partitioned with the same shape as $\widehat{\boldsymbol{G}}$, $\widehat{\boldsymbol{M}} = \begin{bmatrix} \widehat{\boldsymbol{M}}_{1,1} & \widehat{\boldsymbol{M}}_{1,2} \\ \widehat{\boldsymbol{M}}_{2,1} & \widehat{\boldsymbol{M}}_{2,2} \end{bmatrix}$. For the first term, we have

$$\boldsymbol{S}^\top (\boldsymbol{I} - \boldsymbol{Y}_{ij}[k](\boldsymbol{Y}_{ij}[k]^\top \boldsymbol{Y}_{ij}[k])^{-1}\boldsymbol{Y}_{ij}[k]^\top)\boldsymbol{S} = (T-1)(\widehat{\boldsymbol{M}}_{2,2})^{-1}.$$

And

$$\|\widehat{\boldsymbol{M}} - \boldsymbol{M}\| = \|\widehat{\boldsymbol{G}}^{-1} - \boldsymbol{G}^{-1}\| \leq \|\widehat{\boldsymbol{G}}^{-1}\| \cdot \|\widehat{\boldsymbol{G}} - \boldsymbol{G}\| \cdot \|\boldsymbol{G}^{-1}\|. \tag{S1.1}$$

Notice from Lemma 2, with probability tending to 1,

$$\begin{aligned}\|\widehat{\boldsymbol{G}} - \boldsymbol{G}\| &\leq |\mathcal{J}_{\boldsymbol{i}}| \cdot \|\widehat{\boldsymbol{G}} - \boldsymbol{G}\|_{\ell_\infty} \\ &= O_p\big((T^{-1}\log(M \vee N \vee T))^{1/2}|\mathcal{J}_{\boldsymbol{i}}| J_\square \mu_{2q}^2 (1-\delta)^{-4}\big).\end{aligned}$$

Since $\boldsymbol{G}$ is a submatrix of $\boldsymbol{\Gamma}_0$, we have $\lambda_{\min}(\boldsymbol{G}) \geq \lambda_{\min}(\boldsymbol{\Gamma}_0) \geq \kappa_1$. Also, as long as $(T^{-1}\log(M \vee N \vee T))^{1/2}\mu_{2q}^2 J J_\square (1-\delta)^{-4} \lesssim \kappa_1$, we have $\lambda_{\min}(\widehat{\boldsymbol{G}}) \geq$

$\lambda_{\min}(\boldsymbol{G}) - \frac{\kappa_1}{2} \geq \frac{\kappa_1}{2}$. Therefore $\|\widehat{\boldsymbol{M}} - \boldsymbol{M}\| \leq c_0\kappa_1^{-1}$ for some $c_0 \in (0,1)$. So

$$\begin{aligned}&\lambda_{\min}\{\boldsymbol{S}^\top(\boldsymbol{I} - \boldsymbol{Y_i}[k](\boldsymbol{Y_i}[k]^\top\boldsymbol{Y_i}[k])^{-1}\boldsymbol{Y_i}[k]^\top)\boldsymbol{S}\}\\ &= T\lambda_{\min}\big((\widehat{\boldsymbol{M}}_{2,2})^{-1}\big) = T\lambda_{\max}^{-1}(\widehat{\boldsymbol{M}}_{2,2}).\end{aligned}$$

Since $\widehat{\boldsymbol{M}}_{2,2}$ is a submatrix of $\widehat{\boldsymbol{M}}$, we have $\lambda_{\max}(\widehat{\boldsymbol{M}}_{2,2}) \leq \lambda_{\max}(\widehat{\boldsymbol{M}}) \leq \lambda_{\max}(\boldsymbol{M}) + \|\widehat{\boldsymbol{M}} - \boldsymbol{M}\| \leq \kappa_1^{-1} + c_0\kappa_1^{-1}$. And thus

$$\lambda_{\min}(\boldsymbol{S}^\top(\boldsymbol{I} - \boldsymbol{Y_i}[k](\boldsymbol{Y_i}[k]^\top\boldsymbol{Y_i}[k])^{-1}\boldsymbol{Y_i}[k]^\top)\boldsymbol{S}) \geq (1+c_0)^{-1}T\kappa_1,$$

which also implies

$$\boldsymbol{b}_2^\top\boldsymbol{S}^\top(\boldsymbol{I} - \boldsymbol{Y_i}[k](\boldsymbol{Y_i}[k]^\top\boldsymbol{Y_i}[k])^{-1}\boldsymbol{Y_i}[k]^\top)\boldsymbol{S}\boldsymbol{b}_2 \geq (1+c_0)^{-1}T\kappa_1\|\boldsymbol{b}_2\|_{\ell_2}^2.$$

On the other hand, we have from Cauchy-Schwarz inequality and second part in Lemma 2,

$$\begin{aligned}&2\boldsymbol{b}_2^\top\boldsymbol{S}^\top(\boldsymbol{I} - \boldsymbol{Y}_{ij;k}(\boldsymbol{Y}_{ij;k}^\top\boldsymbol{Y}_{ij;k})^{-1}\boldsymbol{Y}_{ij;k}^\top)\boldsymbol{\epsilon}_{ij}\\ \leq\ &\frac{1}{2}(1+c_0)^{-1}T\kappa_1\|\boldsymbol{b}_2\|_{\ell_2}^2 + C_1\kappa_1^{-1}\mu_{2q}^4(1-\delta)^{-4}J_\square^2\log(M\vee N\vee T).\end{aligned}$$

From Lemma 3, we have $\mathsf{RSS}_{\boldsymbol{i}}(k_0) \lesssim T\mu_{2q}^2$ with probability tending to 1. Therefore, for sufficiently large T, we have

$$\begin{aligned}\mathsf{RSS}_{\boldsymbol{i}}(k) - \mathsf{RSS}_{\boldsymbol{i}}(k_0) &\geq \frac{1}{2}(1+c_0)^{-1}T\kappa_1\|\boldsymbol{b}_2\|_{\ell_2}^2\\ &\quad - C_1\kappa_1^{-1}\mu_{2q}^4(1-\delta)^{-4}J_\square^2\log(M\vee N\vee T)\\ &\geq c_1\mathsf{RSS}_{\boldsymbol{i}}(k_0)\frac{\kappa_1}{\mu_{2q}^2}\|\boldsymbol{b}_2\|_{\ell_2}^2\end{aligned}$$

for some $c_1 \in (0,1)$. And using $\log(1+x) \geq \frac{1}{2}x$ for $x \in (0,2)$, we have

$$\log \mathsf{RSS}_{\boldsymbol{i}}(k) - \log \mathsf{RSS}_{\boldsymbol{i}}(k_0) \geq \min\{c_2 \frac{\kappa_1}{\mu_{2q}^2} \|\boldsymbol{b}_2\|_{\ell_2}^2, \log 3\}.$$

Therefore for sufficiently large T, we have

$$\begin{aligned}\mathsf{BIC}_{\boldsymbol{i}}(k) - \mathsf{BIC}_{\boldsymbol{i}}(k_0) \geq \min\{c_2 \frac{\kappa_1}{\mu_{2q}^2} \|\boldsymbol{b}_2\|_{\ell_2}^2, \log 3\} \\ - D_0 \frac{|\mathcal{J}_{\boldsymbol{i}}|}{T} \log(M \vee N \vee T) > 0,\end{aligned}$$

under the margin condition that

$$\|\boldsymbol{b}_2\|_{\ell_2} = \|\boldsymbol{M}_{\boldsymbol{i}}(\mathcal{J}_{\boldsymbol{i}} \backslash \mathcal{J}_{\boldsymbol{i}}[k])\|_{\mathrm{F}} \geq \|\boldsymbol{M}_{\boldsymbol{i}}(\mathcal{J}_{\boldsymbol{i}} \backslash \mathcal{J}_{\boldsymbol{i}}[k_0 - 1])\|_{\mathrm{F}}.$$

Now we show, for all $\boldsymbol{i} \in \mathcal{S}$ and $k > k_0$, $\mathbb{P}(\mathsf{BIC}_{\boldsymbol{i}}(k) < \mathsf{BIC}_{\boldsymbol{i}}(k_0)) \to 0$. We have

$$\mathsf{RSS}_{\boldsymbol{i}}(k) = \inf_{\boldsymbol{\gamma}} \|\boldsymbol{z}_{\boldsymbol{i}} - \boldsymbol{Y}_{\boldsymbol{i}}[k]\boldsymbol{\gamma}\|_{\ell_2}^2.$$

We can write $\boldsymbol{Y}_{\boldsymbol{i}}[k]\boldsymbol{\gamma} = \boldsymbol{Y}_{\boldsymbol{i}}[k_0]\boldsymbol{\gamma}_1 + \boldsymbol{S}\boldsymbol{\gamma}_2$. Denote

$$\widetilde{\boldsymbol{S}} = \big(I - \boldsymbol{Y}_{\boldsymbol{i}}[k_0]\big(\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0]\big)^{-1} \boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top\big)\boldsymbol{S}.$$

Then

$$\begin{aligned}\mathsf{RSS}_{\boldsymbol{i}}(k) &= \inf_{\boldsymbol{\gamma}} \|\boldsymbol{z}_{\boldsymbol{i}} - \boldsymbol{Y}_{\boldsymbol{i}}[k]\boldsymbol{\gamma}\|_{\ell_2}^2 = \inf_{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2} \|\boldsymbol{z}_{\boldsymbol{i}} - \boldsymbol{Y}_{\boldsymbol{i}}[k_0]\boldsymbol{\gamma}_1 - \boldsymbol{S}\boldsymbol{\gamma}_2\|_{\ell_2}^2 \\ &= \inf_{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2} \|\boldsymbol{z}_{\boldsymbol{i}} - \boldsymbol{Y}_{\boldsymbol{i}}[k_0]\Big(\boldsymbol{\gamma}_1 + \big(\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0]\big)^{-1} \boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{\gamma}_2\Big) - \widetilde{\boldsymbol{S}}\boldsymbol{\gamma}_2\|_{\ell_2}^2 \\ &= \inf_{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2} \|\boldsymbol{z}_{\boldsymbol{i}} - \boldsymbol{Y}_{\boldsymbol{i}}[k_0]\boldsymbol{\gamma}_1 - \widetilde{\boldsymbol{S}}\boldsymbol{\gamma}_2\|_{\ell_2}^2.\end{aligned}$$

An explicit minimizer is

$$\boldsymbol{\gamma}_1 = \big(\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0]\big)^{-1} \boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{z}_{\boldsymbol{i}}, \qquad \boldsymbol{\gamma}_2 = \big(\widetilde{\boldsymbol{S}}^\top \widetilde{\boldsymbol{S}}\big)^{-1} \widetilde{\boldsymbol{S}}^\top \boldsymbol{z}_{\boldsymbol{i}}.$$

Therefore,

$$\mathsf{RSS}_{\boldsymbol{i}}(k) = \mathsf{RSS}_{\boldsymbol{i}}(k_0) - \|\widetilde{\boldsymbol{S}}\big(\widetilde{\boldsymbol{S}}^\top \widetilde{\boldsymbol{S}}\big)^{-1} \widetilde{\boldsymbol{S}}^\top \boldsymbol{\nu}_{\boldsymbol{i}}\|_{\ell_2}^2.$$

Next we bound $\|\widetilde{\boldsymbol{S}}(\widetilde{\boldsymbol{S}}^\top \widetilde{\boldsymbol{S}})^{-1} \widetilde{\boldsymbol{S}}^\top \boldsymbol{\nu}_{\boldsymbol{i}}\|_{\ell_2}^2$:

$$\|\widetilde{\boldsymbol{S}}(\widetilde{\boldsymbol{S}}^\top \widetilde{\boldsymbol{S}})^{-1} \widetilde{\boldsymbol{S}}^\top \boldsymbol{\nu}_{\boldsymbol{i}}\|_{\ell_2}^2 = \boldsymbol{\nu}_{\boldsymbol{i}}^\top \widetilde{\boldsymbol{S}}(\widetilde{\boldsymbol{S}}^\top \widetilde{\boldsymbol{S}})^{-1} \widetilde{\boldsymbol{S}}^\top \boldsymbol{\nu}_{\boldsymbol{i}} \le \lambda_{\max}\big((\widetilde{\boldsymbol{S}}^\top \widetilde{\boldsymbol{S}})^{-1}\big) \|\widetilde{\boldsymbol{S}}^\top \boldsymbol{\nu}_{\boldsymbol{i}}\|_{\ell_2}^2.$$

Since $\widetilde{\boldsymbol{S}}^\top \boldsymbol{\nu}_{\boldsymbol{i}} \in \mathbb{R}^{|\mathcal{J}_{\boldsymbol{i}}[k]| - |\mathcal{J}_{\boldsymbol{i}}|}$, we have

$$\|\widetilde{\boldsymbol{S}}^\top \boldsymbol{\nu}_{\boldsymbol{i}}\|_{\ell_2}^2 \le \big(|\mathcal{J}_{\boldsymbol{i}}[k]| - |\mathcal{J}_{\boldsymbol{i}}|\big) \|\widetilde{\boldsymbol{S}}^\top \boldsymbol{\nu}_{\boldsymbol{i}}\|_{\ell_\infty}^2.$$

Recall $\widetilde{\boldsymbol{S}}^\top \boldsymbol{\nu}_{\boldsymbol{i}} = \boldsymbol{S}^\top\big(\boldsymbol{I} - \boldsymbol{Y}_{\boldsymbol{i}}[k_0](\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1} \boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top\big)\boldsymbol{\nu}_{\boldsymbol{i}}$. Let an arbitrary row of $\boldsymbol{S}^\top$ be $\boldsymbol{x}^\top$. Then

$$\begin{aligned} &\boldsymbol{x}^\top\big(\boldsymbol{I} - \boldsymbol{Y}_{\boldsymbol{i}}[k_0](\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1} \boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top\big)\boldsymbol{\nu}_{\boldsymbol{i}} \\ &= \boldsymbol{x}^\top \boldsymbol{\nu}_{\boldsymbol{i}} - \boldsymbol{x}^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0](\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1} \boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{\nu}_{\boldsymbol{i}}. \end{aligned} \tag{S1.2}$$

From Lemma 2, $\boldsymbol{x}^\top \boldsymbol{\nu}_{\boldsymbol{i}} = O_p\big((T \log(M \vee N \vee T))^{1/2} \mu_{2q}^2 J_\square^{1/2} (1-\delta)^{-3/2}\big)$. Moreover,

$$\begin{aligned} &\boldsymbol{x}^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0](\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1} \boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{\nu}_{\boldsymbol{i}} \\ &= \boldsymbol{x}^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0]\Big((\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1} - (\mathbb{E}\, \boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1}\Big) \boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{\nu}_{\boldsymbol{i}} \\ &\quad + \boldsymbol{x}^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0](\mathbb{E}\, \boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1} \boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{\nu}_{\boldsymbol{i}}. \end{aligned}$$

Since $\frac{1}{T}\mathbb{E}\,\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0]$ is a submatrix of $\boldsymbol{\Gamma}_0$, we have

$$\lambda_{\max}\big((\mathbb{E}\,\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1}\big) = \lambda_{\min}^{-1}(\mathbb{E}\,\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0]) \le T^{-1}\kappa_1^{-1}.$$

Therefore, from Lemma 2 and $\max_{\boldsymbol{i}}[\boldsymbol{\Gamma}_0]_{\boldsymbol{i};\boldsymbol{i}} \le \|[\boldsymbol{E}_t]_{\boldsymbol{i}}\|_{2q}^2 \le \mu_{2q}^2$, we have

$$\begin{aligned}
&|\boldsymbol{x}^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0](\mathbb{E}\,\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1}\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{\nu}_{\boldsymbol{i}}| \\
&\lesssim (T\log(M\vee N\vee T))^{1/2}\mu_{2q}^2\,|\mathcal{J}_{\boldsymbol{i}}|\,J_\square^{1/2}(1-\delta)^{-2}\,\kappa_1^{-1}\mu_{2q}^2.
\end{aligned}$$

Next we bound $\|(\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1} - (\mathbb{E}\,\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1}\|$ similarly as in (S1.1):

$$\begin{aligned}
&\|(\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1} - (\mathbb{E}\,\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1}\| \\
&\qquad\lesssim T^{-1}(T^{-1}\log(M\vee N\vee T))^{1/2}\mu_{2q}^2\,|\mathcal{J}_{\boldsymbol{i}}|\,J_\square(1-\delta)^{-4}\kappa_1^{-2}.
\end{aligned}$$

Thus

$$\begin{aligned}
&\boldsymbol{x}^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0]\Big((\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1} - (\mathbb{E}\,\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1}\Big)\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{\nu}_{\boldsymbol{i}} \\
&\lesssim T^{-1}(T^{-1}\log(M\vee N\vee T))^{1/2}\mu_{2q}^2\,|\mathcal{J}_{\boldsymbol{i}}|\,J_\square(1-\delta)^{-3}\kappa_1^{-2}\cdot|\mathcal{J}_{\boldsymbol{i}}|\cdot T\mu_{2q}^2 \\
&\qquad\qquad\cdot(T\log(M\vee N\vee T))^{1/2}\mu_{2q}^2\,J_\square^{1/2}(1-\delta)^{-2} \\
&= J_\square^{3/2}\,|\mathcal{J}_{\boldsymbol{i}}|^2\,\log(M\vee N\vee T)\,\mu_{2q}^6\,\kappa_1^{-2}(1-\delta)^{-6}.
\end{aligned}$$

Returning to (S1.2), we conclude

$$\|\widetilde{\boldsymbol{S}}^\top \boldsymbol{\nu}_{\boldsymbol{i}}\|_{\ell_\infty} = O_p\big((T\log(M\vee N\vee T))^{1/2}\mu_{2q}^4\,|\mathcal{J}_{\boldsymbol{i}}|\,J_\square^{1/2}(1-\delta)^{-2}\kappa_1^{-1}\big).$$

Hence

$$\mathsf{RSS}_{\boldsymbol{i}}(k_0) - \mathsf{RSS}_{\boldsymbol{i}}(k) \lesssim \left(|\mathcal{J}_{\boldsymbol{i}}[k]| - |\mathcal{J}_{\boldsymbol{i}}|\right) \log(M \vee N \vee T)\, \mu_{2q}^8\, |\mathcal{J}_{\boldsymbol{i}}|^2 \times J_{\square}(1-\delta)^{-4}\, \kappa_1^{-3}.$$

From Lemma 3, $\mathsf{RSS}_{\boldsymbol{i}}(k) \gtrsim T\mu_{2q}^2$. Therefore

$$\frac{\mathsf{RSS}_{\boldsymbol{i}}(k_0) - \mathsf{RSS}_{\boldsymbol{i}}(k)}{\mathsf{RSS}_{\boldsymbol{i}}(k)} \lesssim \frac{|\mathcal{J}_{\boldsymbol{i}}[k]| - |\mathcal{J}_{\boldsymbol{i}}|}{T} \log(M \vee N \vee T)\, \mu_{2q}^6\, |\mathcal{J}_{\boldsymbol{i}}|^2 \times J_{\square}(1-\delta)^{-4}\, \kappa_1^{-3}.$$

Using $\log(1+x) \le x$ for $x > 0$, we have

$$\begin{aligned} &\log \mathsf{RSS}_{\boldsymbol{i}}(k_0) - \log \mathsf{RSS}_{\boldsymbol{i}}(k) \\ \lesssim\ & \frac{|\mathcal{J}_{\boldsymbol{i}}[k]| - |\mathcal{J}_{\boldsymbol{i}}|}{T} \log(M \vee N \vee T)\, \mu_{2q}^6\, |\mathcal{J}_{\boldsymbol{i}}|^2 \\ & \times J_{\square}(1-\delta)^{-4}\, \kappa_1^{-3}. \end{aligned}$$

Therefore

$$\begin{aligned} \mathsf{BIC}_{\boldsymbol{i}}(k) - \mathsf{BIC}_{\boldsymbol{i}}(k_0) &\ge D_0\, \frac{|\mathcal{J}_{\boldsymbol{i}}[k]| - |\mathcal{J}_{\boldsymbol{i}}|}{T} \log(M \vee N \vee T) \\ &\quad + \log \mathsf{RSS}_{\boldsymbol{i}}(k) - \log \mathsf{RSS}_{\boldsymbol{i}}(k_0) \\ &> 0, \end{aligned}$$

as long as we choose $D_0 \gtrsim \mu_{2q}^6\, |\mathcal{J}_{\boldsymbol{i}}|^2\, J_{\square}\, (1-\delta)^{-4}\, \kappa_1^{-3}$.

□

S2 Auxiliary Lemmas

The auxiliary results in this section are used to support the proofs in Section S1. Lemma 1 provides the functional-dependence concentration inequality that is used as the main probabilistic tool. Lemma 2 applies this concentration result to obtain entrywise bounds for the sample auto-covariance matrix and the noise-design cross terms; these bounds are used in the proofs of Theorems 2, 4, and 5. Lemma 3 controls the residual sum of squares over candidate neighborhoods and is used in the proof of Theorem 5. Lemma 4 records a locality-induced sparsity property of powers of the autoregressive coefficient matrix, which is used in the proof of Lemma 2. Lemma 5 is the perturbation result for the best rank-R projection and is used in the proof of Theorem 3. Finally, Lemma 6 provides the fixed-dimensional law of large numbers used under Regime 1 in the proof of Theorem 2.

We start with some notations. Let $\mathbf{\Gamma} \in \mathbb{R}^{MN \times MN}$. For multi-indices $\boldsymbol{i} = (i_1, i_2), \boldsymbol{j} = (j_1, j_2) \in \mathcal{S} = [M] \times [N]$, define

$$[\mathbf{\Gamma}]_{\boldsymbol{i},\boldsymbol{j}} := [\mathbf{\Gamma}]_{(i_2-1)M+i_1,(j_2-1)M+j_1}, \tag{S2.3}$$

where the last equality uses column-major linear indexing. The sub-matrix $\mathbf{\Gamma}(\mathcal{J}_1; \mathcal{J}_2)$ is then formed by extracting the elements of $\mathbf{\Gamma}$ corresponding to

the multi-indices in $\mathcal{J}_1$ and $\mathcal{J}_2$, arranged using column-major ordering.

Following Wu (2005), we introduce the following functional dependent measure. Let z_i be a stationary process of the form $z_i = g(\mathcal{F}_i)$, where g is a measurable function and $\mathcal{F}_i = (\cdots, e_{-1}, e_0, e_1, \cdots)$ with independent and identically distributed random variables $\{e_i\}$. The functional dependent measure is defined as

$$\theta_{i,q} = \|z_i - z_i^*\|_q = \|g(\mathcal{F}_i) - g(\mathcal{F}_i^*)\|_q,$$

where $z_i^* = g(\mathcal{F}_i^*)$ is the coupled process of z_i, $\mathcal{F}_i^* = (\cdots, e_{-1}, e_0^*, e_1, \cdots)$ with $\{e_0, e_0^*\}$ being independent and identically distributed. Here $\|\cdot\|_q := (\mathbb{E}|\cdot|)^{1/q}$ for some $q \geq 1$. Then we have the following from Theorem 2 in Liu et al. (2013) that plays the crucial role in our proof.

Lemma 1. *Let $S_n = \sum_{i=1}^n z_n$ and $\Theta_{m,q} = \sum_{i=m}^\infty \theta_{i,q}$. Assume for each m, $\Theta_{m,q} = O(m^{-\alpha})$ with $\alpha > \frac{1}{2} - \frac{1}{q}$ and $q > 2$. Then we have for all $x > 0$,*

$$\mathbb{P}(|S_n| \geq x) \leq C_1 \frac{\Theta_{0,q}^q n}{x^q} + C_3 \exp(-C_2 \frac{x^2}{\Theta_{0,q}^2 n}),$$

for $C_1, C_2, C_3 > 0$ depending only on q.

Recall $[\widehat{\boldsymbol{\Gamma}}_0]_{\boldsymbol{i},\boldsymbol{j}} = \frac{1}{T}\sum_{t=1}^T [\boldsymbol{X}_t]_{\boldsymbol{i}}[\boldsymbol{X}_t]_{\boldsymbol{j}}$ and thus $\boldsymbol{\Gamma}_0 = \mathbb{E}\widehat{\boldsymbol{\Gamma}}_0$. We denote

$$\boldsymbol{x_i} = ([\boldsymbol{X}_1]_{\boldsymbol{i}}, \cdots, [\boldsymbol{X}_{T-1}]_{\boldsymbol{i}})^\top,$$

and recall $\boldsymbol{\nu_i} = ([\boldsymbol{E}_2]_{\boldsymbol{i}}, \cdots, [\boldsymbol{E}_T]_{\boldsymbol{i}})^\top$. We first establish the following concentration inequality.

Lemma 2. *Under Assumption 1, there exists absolute constant $C > 0$, such that*

1. *For $\boldsymbol{i}, \boldsymbol{j} \in \mathcal{S}$, we have for $x > 0$,*

$$\mathbb{P}\bigg(\big|[\widehat{\boldsymbol{\Gamma}}_0]_{\boldsymbol{i},\boldsymbol{j}} - [\boldsymbol{\Gamma}_0]_{\boldsymbol{i},\boldsymbol{j}}\big| \geq x \bigg) \lesssim CT\bigg(\frac{\mu_{2q}^2 J_\square (1-\delta)^{-4}}{Tx} \bigg)^q + C\exp\bigg(-\frac{Tx^2}{(\mu_{2q}^2 J_\square (1-\delta)^{-4})^2} \bigg).$$

And this leads to

$$\max_{\boldsymbol{i},\boldsymbol{j}\in\mathcal{S}} \big|[\widehat{\boldsymbol{\Gamma}}_0]_{\boldsymbol{i},\boldsymbol{j}} - [\boldsymbol{\Gamma}_0]_{\boldsymbol{i},\boldsymbol{j}}\big| = O_p\big((T^{-1}\log(M\vee N\vee T))^{1/2} \mu_{2q}^2 J_\square (1-\delta)^{-4} \big).$$

2. *For $\boldsymbol{i}, \boldsymbol{j} \in \mathcal{S}$, we have for $x > 0$,*

$$\mathbb{P}(\big|\boldsymbol{\nu}_{\boldsymbol{i}}^\top \boldsymbol{x}_{\boldsymbol{j}}\big| \geq x) \leq CT\left(\frac{\mu_{2q}^2 J_\square^{1/2} (1-\delta)^{-2}}{x} \right)^q + C\exp\left(-\frac{x^2}{T\cdot(\mu_{2q}^2 J_\square^{1/2}(1-\delta)^{-2})^2} \right).$$

And this leads to

$$\max_{\boldsymbol{i},\boldsymbol{j}\in\mathcal{S}} \big|\boldsymbol{\nu}_{\boldsymbol{i}}^\top \boldsymbol{x}_{\boldsymbol{j}}\big| = O_p\big((T\log(M\vee N\vee T))^{1/2} \mu_{2q}^2 J_\square^{1/2} (1-\delta)^{-2} \big).$$

Proof. We shall use Lemma 1 to prove this. Denote $\mu_q = \max_{\boldsymbol{i}} \|[\boldsymbol{E}_0]_{\boldsymbol{i}}\|_q$. Using innovation representation, we have

$$\boldsymbol{x}_t = \sum_{l=0}^{\infty} \boldsymbol{M}^l \boldsymbol{\epsilon}_{t-l}.$$

As a result, for each $\boldsymbol{i} = (i_1, i_2) \in \mathcal{S}$,

$$(\boldsymbol{e}_{i_2} \otimes \boldsymbol{e}_{i_1})^\top \boldsymbol{x}_t = \sum_{l=0}^{\infty} (\boldsymbol{e}_{i_2} \otimes \boldsymbol{e}_{i_1})^\top \boldsymbol{M}^l \boldsymbol{\epsilon}_{t-l}$$
$$= \sum_{l=0}^{\infty} \sum_{\boldsymbol{j}} (\boldsymbol{e}_{i_2} \otimes \boldsymbol{e}_{i_1})^\top \boldsymbol{M}^l (\boldsymbol{e}_{j_2} \otimes \boldsymbol{e}_{j_1})(\boldsymbol{e}_{j_2} \otimes \boldsymbol{e}_{j_1})^\top \boldsymbol{\epsilon}_{t-l}.$$

Following the definition of $J_\square$ and Lemma 4, we have $\|(\boldsymbol{e}_{i_2} \otimes \boldsymbol{e}_{i_1})^\top \boldsymbol{M}^l\|_{\ell_0} \leq \min\{J_\square l^2, MN\}$. And thus

$$\begin{aligned}
\max_{\boldsymbol{i}} \|(\boldsymbol{e}_{i_2} \otimes \boldsymbol{e}_{i_1})^\top \boldsymbol{x}_t\|_{2q} &\leq \max_{\boldsymbol{i}} \Bigg\| \sum_{l=0}^{\infty} \sum_{\boldsymbol{j}} (\boldsymbol{e}_{i_2} \otimes \boldsymbol{e}_{i_1})^\top \boldsymbol{M}^l (\boldsymbol{e}_{j_2} \otimes \boldsymbol{e}_{j_1}) \\
&\qquad \times (\boldsymbol{e}_{j_2} \otimes \boldsymbol{e}_{j_1})^\top \boldsymbol{\epsilon}_{t-l} \Big\|_{2q} \\
&\leq \max_{\boldsymbol{i}} \sum_{l=0}^{\infty} \sum_{\boldsymbol{j}} |[\boldsymbol{M}^l]_{\boldsymbol{i},\boldsymbol{j}}| \cdot \|[\boldsymbol{E}_{t-l}]_{\boldsymbol{j}}\|_{2q} \\
&\leq \max_{\boldsymbol{i}} \sum_{l=0}^{\infty} J_\square^{1/2} l \|(\boldsymbol{e}_{i_2} \otimes \boldsymbol{e}_{i_1})^\top \boldsymbol{M}^l\|_{\ell_2} \cdot \mu_{2q} \\
&\leq J_\square^{1/2} \mu_{2q} \sum_{l \geq 0} l\delta^l \leq C J_\square^{1/2} \mu_{2q} (1-\delta)^{-2} \\
&< +\infty, \qquad \text{(S2.4)}
\end{aligned}$$

where in the second-to-last line, we use the fact $\|\boldsymbol{M}^l\| \leq \delta^l$.

On the other hand, we can bound $\|(\boldsymbol{e}_{i_2} \otimes \boldsymbol{e}_{i_1})^\top \boldsymbol{x}_t - (\boldsymbol{e}_{i_2} \otimes \boldsymbol{e}_{i_1})^\top \boldsymbol{x}_t^*\|_{2q}$

as follows using the innovation representation:

$$\begin{aligned}
&\|(\boldsymbol{e}_{i_2}\otimes\boldsymbol{e}_{i_1})^\top\boldsymbol{x}_t-(\boldsymbol{e}_{i_2}\otimes\boldsymbol{e}_{i_1})^\top\boldsymbol{x}_t^*\|_{2q}\\
&\quad=\left\|\sum_{\boldsymbol{j}}(\boldsymbol{e}_{i_2}\otimes\boldsymbol{e}_{i_1})^\top\boldsymbol{M}^t(\boldsymbol{e}_{j_2}\otimes\boldsymbol{e}_{j_1})(\boldsymbol{e}_{j_2}\otimes\boldsymbol{e}_{j_1})^\top(\boldsymbol{\epsilon}_0-\boldsymbol{\epsilon}_0^*)\right\|_{2q}\\
&\leq C\mu_{2q}J_\square^{1/2}t\delta^t. \qquad \text{(S2.5)}
\end{aligned}$$

Now we can bound $\|[\boldsymbol{X}_t]_{\boldsymbol{i}}[\boldsymbol{X}_t]_{\boldsymbol{j}}-[\boldsymbol{X}_t]^*_{\boldsymbol{i}}[\boldsymbol{X}_t]^*_{\boldsymbol{j}}\|_q$ using (S2.4) and (S2.5):

$$\begin{aligned}
&\|[\boldsymbol{X}_t]_{\boldsymbol{i}}[\boldsymbol{X}_t]_{\boldsymbol{j}}-[\boldsymbol{X}_t]^*_{\boldsymbol{i}}[\boldsymbol{X}_t]^*_{\boldsymbol{j}}\|_q\\
&\leq\|(\boldsymbol{e}_{i_2}\otimes\boldsymbol{e}_{i_1})^\top\boldsymbol{x}_t\cdot(\boldsymbol{e}_{j_2}\otimes\boldsymbol{e}_{j_1})^\top(\boldsymbol{x}_t-\boldsymbol{x}_t^*)\|_q\\
&\quad+\|(\boldsymbol{e}_{i_2}\otimes\boldsymbol{e}_{i_1})^\top(\boldsymbol{x}_t-\boldsymbol{x}_t^*)\cdot(\boldsymbol{e}_{j_2}\otimes\boldsymbol{e}_{j_1})^\top\boldsymbol{x}_t^*\|_q\\
&\leq\|(\boldsymbol{e}_{i_2}\otimes\boldsymbol{e}_{i_1})^\top\boldsymbol{x}_t\|_{2q}\cdot\|(\boldsymbol{e}_{j_2}\otimes\boldsymbol{e}_{j_1})^\top(\boldsymbol{x}_t-\boldsymbol{x}_t^*)\|_{2q}\\
&\quad+\|(\boldsymbol{e}_{j_2}\otimes\boldsymbol{e}_{j_1})^\top\boldsymbol{x}_t^*\|_{2q}\cdot\|(\boldsymbol{e}_{i_2}\otimes\boldsymbol{e}_{i_1})^\top(\boldsymbol{x}_t-\boldsymbol{x}_t^*)\|_{2q}\\
&\leq C\mu_{2q}^2J_\square(1-\delta)^{-2}t\delta^t.
\end{aligned}$$

Therefore

$$\begin{aligned}
\Theta_{m,q}&:=\sum_{t=m}^\infty\|[\boldsymbol{X}_t]_{\boldsymbol{i}}[\boldsymbol{X}_t]_{\boldsymbol{j}}-[\boldsymbol{X}_t]^*_{\boldsymbol{i}}[\boldsymbol{X}_t]^*_{\boldsymbol{j}}\|_q\\
&\leq C\mu_{2q}^2J_\square(1-\delta)^{-4}\delta^m=O(m^{-\alpha})
\end{aligned}$$

for any $\alpha>1$.

From Lemma 1, we conclude

$$\mathbb{P}\bigg(\big|[\widehat{\mathbf{\Gamma}}_0]_{\boldsymbol{i},\boldsymbol{j}} - [\mathbf{\Gamma}_0]_{\boldsymbol{i},\boldsymbol{j}}\big| \geq x\bigg) \lesssim T\bigg(\frac{\mu_{2q}^2 J_\square (1-\delta)^{-4}}{Tx}\bigg)^q + \exp\bigg(-\frac{Tx^2}{(\mu_{2q}^2 J_\square (1-\delta)^{-4})^2}\bigg).$$

By setting $x = T^{-1/2}\mu_{2q}^2 J_\square (1-\delta)^{-4}(\log(M \vee N \vee T))^{1/2}$, we conclude

$$\max_{\boldsymbol{i},\boldsymbol{j}\in\mathcal{S}} \big|[\widehat{\mathbf{\Gamma}}_0]_{\boldsymbol{i},\boldsymbol{j}} - [\mathbf{\Gamma}_0]_{\boldsymbol{i},\boldsymbol{j}}\big| = O_p\big(T^{-1/2}\mu_{2q}^2 J_\square (1-\delta)^{-4}(\log(M \vee N \vee T))^{1/2}\big).$$

Similarly, we can show

$$\mathbb{P}\big(\big|\boldsymbol{\nu}_{\boldsymbol{i}}^\top \boldsymbol{x}_{\boldsymbol{j}}\big| \geq x\big) \lesssim T\left(\frac{\mu_{2q}^2 J_\square^{1/2}(1-\delta)^{-2}}{x}\right)^q + \exp\left(-\frac{x^2}{T\cdot(\mu_{2q}^2 J_\square^{1/2}(1-\delta)^{-2})^2}\right),$$

and

$$\max_{\boldsymbol{i},\boldsymbol{j}\in\mathcal{S}} \big|\boldsymbol{\nu}_{\boldsymbol{i}}^\top \boldsymbol{x}_{\boldsymbol{j}}\big| = O_p\big((T\log(M \vee N \vee T))^{1/2}\mu_{2q}^2 J_\square^{1/2}(1-\delta)^{-2}\big).$$

□

Lemma 3. *Under Assumption 1, we have for each $\boldsymbol{i}$ and $k \geq k_0$,*

$$\max_{\boldsymbol{i}} \Big|\mathsf{RSS}_{\boldsymbol{i}}(k) - (T-1)\mathbf{\Sigma}_{\boldsymbol{i};\boldsymbol{i}}\Big| = O_p\big((T\log(M \vee N \vee T))^{1/2}\mu_{2q}^2\big)$$

holds as $T \to \infty$ and $\frac{\max_{\boldsymbol{i}} |\mathcal{J}_{\boldsymbol{i}}[K_0]|^2 \cdot J_\square^2 \log(M\vee N\vee T)}{T} \leq \widetilde{c}$ for some $\widetilde{c} > 0$ depending only on $\kappa_1, \delta, \mu_{2q}$.

Proof. When $k \geq k_0$, we can decompose $\mathsf{RSS}_{\boldsymbol{i}}(k)$ as

$$\mathsf{RSS}_{\boldsymbol{i}}(k) = \boldsymbol{\nu}_{\boldsymbol{i}}^\top \boldsymbol{\nu}_{\boldsymbol{i}} - \boldsymbol{\nu}_{\boldsymbol{i}}^\top \boldsymbol{Y}_{\boldsymbol{i}}[k](\boldsymbol{Y}_{\boldsymbol{i}}[k]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k])^{-1}\boldsymbol{Y}_{\boldsymbol{i}}[k]^\top \boldsymbol{\nu}_{\boldsymbol{i}}.$$

We consider the event

$$\mathcal{E}_1 = \Big\{ \max_{\boldsymbol{i},\boldsymbol{j}\in\mathcal{S}} \big|[\widehat{\boldsymbol{\Gamma}}_0]_{\boldsymbol{i},\boldsymbol{j}} - [\boldsymbol{\Gamma}_0]_{\boldsymbol{i},\boldsymbol{j}}\big| \lesssim (T^{-1}\log(M\vee N\vee T))^{1/2}\mu_{2q}^2 J_{\square}(1-\delta)^{-4}, \\ \max_{\boldsymbol{i},\boldsymbol{j}\in\mathcal{S}} \big|\boldsymbol{\nu}_{\boldsymbol{i}}^\top \boldsymbol{x}_{\boldsymbol{j}}\big| \lesssim \big(T\log(M\vee N\vee T)\big)^{1/2}\mu_{2q}^2 J_{\square}^{1/2}(1-\delta)^{-2} \Big\}.$$

From Lemma 2, $\mathbb{P}(\mathcal{E}_1)\to 0$. We shall continue our proof under $\mathcal{E}_1$. From Lemma 1, we obtain

$$\max_{\boldsymbol{i}} \big|\boldsymbol{\nu}_{\boldsymbol{i}}^\top\boldsymbol{\nu}_{\boldsymbol{i}} - (T-1)\sigma_{\boldsymbol{i}}^2\big| = O_p\big((T\log(M\vee N\vee T))^{1/2}\mu_{2q}^2\big). \tag{S2.6}$$

Recall $\mathcal{J}_{\boldsymbol{i}}[k]$ is the index set corresponding with the columns of $\boldsymbol{Y}_{ij;k}$, then for all i,j, under $\mathcal{E}_1$,

$$\begin{aligned}
&(T-1)^{-1}\lambda_{\min}(\boldsymbol{Y}_{\boldsymbol{i}}[k]^\top\boldsymbol{Y}_{\boldsymbol{i}}[k]) \\
\geq\ &\lambda_{\min}\big(\boldsymbol{\Gamma}_0(\mathcal{J}_{\boldsymbol{i}}[k];\mathcal{J}_{\boldsymbol{i}}[k])\big) \\
&- \|(T-1)^{-1}\boldsymbol{Y}_{\boldsymbol{i}}[k]^\top\boldsymbol{Y}_{\boldsymbol{i}}[k] - \boldsymbol{\Gamma}_0(\mathcal{J}_{\boldsymbol{i}}[k];\mathcal{J}_{\boldsymbol{i}}[k])\| \\
\geq\ &\kappa_1 - |\mathcal{J}_{\boldsymbol{i}}[k]|(T^{-1}\log(M\vee N\vee T))^{1/2}\mu_{2q}^2 J_{\square}(1-\delta)^{-4} \\
\geq\ &\kappa_1(1-c_1),
\end{aligned}$$

for some $c_1\in(0,1)$ as long as

$$|\mathcal{J}_i[k]|(T^{-1}\log(M\vee N\vee T))^{1/2}\mu_{2q}^2 J_{\square}(1-\delta)^{-4} \leq c_1\kappa_1.$$

And under $\mathcal{E}_1$,

$$\|\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top\boldsymbol{\nu}_{\boldsymbol{i}}\|_{\ell_\infty} \lesssim (T\log(M\vee N\vee T))^{1/2}\mu_{2q}^2 J_{\square}^{1/2}(1-\delta)^{-2}.$$

Therefore, with

$$\Pi_{\boldsymbol{i},k_0} = \boldsymbol{Y}_{\boldsymbol{i}}[k_0](\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top \boldsymbol{Y}_{\boldsymbol{i}}[k_0])^{-1}\boldsymbol{Y}_{\boldsymbol{i}}[k_0]^\top,$$

$$|\boldsymbol{\nu}_{\boldsymbol{i}}^\top \Pi_{\boldsymbol{i},k_0}\boldsymbol{\nu}_{\boldsymbol{i}}| \lesssim |\mathcal{J}_{\boldsymbol{i}}| J_\square (1-\delta)^{-4}\mu_{2q}^4\kappa_1^{-1}\log(M\vee N\vee T).$$

Together with (S2.6), and

$$|\mathcal{J}_{\boldsymbol{i}}| J_\square (1-\delta)^{-4}\mu_{2q}^4\kappa_1^{-1}\log(M\vee N\vee T) \lesssim (T\log(M\vee N\vee T))^{1/2}\mu_{2q}^2,$$

we finish the proof. □

Lemma 4. *For any $p > 0$, we have*

$$\|(\boldsymbol{e}_{i_2}\otimes \boldsymbol{e}_{i_1})^\top \boldsymbol{M}^p\|_{\ell_0} \le \min\{J_\square p^2, MN\}.$$

Proof. We prove this for $p = 2$ and it can be similarly extended to larger p. Recall all the support $\mathcal{J}_{\boldsymbol{i}}$ are included in the same rectangle $\{i_1-k_1,\cdots,i_1+k_1\}\times\{i_2-k_2,\cdots,i_2+k_2\}$ of size $J_\square = (2k_1+1)(2k_2+1)$. We denote $\boldsymbol{M}$ by its columns: $\boldsymbol{M} = \begin{bmatrix}\widetilde{\boldsymbol{m}}_{11} & \cdots & \widetilde{\boldsymbol{m}}_{MN}\end{bmatrix}$. Then

$$(\boldsymbol{e}_{i_2}\otimes \boldsymbol{e}_{i_1})^\top \boldsymbol{M}^2(\boldsymbol{e}_{j_2}\otimes \boldsymbol{e}_{j_1}) = \boldsymbol{m}_{\boldsymbol{i}}^\top \boldsymbol{m}_{\boldsymbol{j}} = \sum_{\boldsymbol{k}} [\boldsymbol{M}]_{\boldsymbol{i},\boldsymbol{k}}[\boldsymbol{M}]_{\boldsymbol{k},\boldsymbol{j}}.$$

The summation is non-vanishing only when $\mathcal{J}_{\boldsymbol{i}}\cap\mathcal{J}_{\boldsymbol{j}} \neq \emptyset$. When $|i_1-j_1| > 2k_1$ or $|i_2-j_2| > 2k_2$, the intersection is empty. And thus there are at most $(4k_1+1)(4k_2+2) \le 4J_\square$ $\boldsymbol{j}$s making the term non-vanishing. □

Lemma 5. *Let $\boldsymbol{M} \in \mathbb{R}^{P_1 \times P_2}$ be a rank R matrix. Let $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ be its compact singular value decomposition and denote $\lambda_{\min} > 0$ the smallest non-zero singular value of $\boldsymbol{M}$. Then for any matrix $\widehat{\boldsymbol{M}}$ such that $\|\widehat{\boldsymbol{M}} - \boldsymbol{M}\|_{\mathrm{F}} \leq \frac{\lambda_{\min}}{8}$, we have that the best rank R approximation of $\widehat{\boldsymbol{M}}$ under Frobenius norm (denoted by $\mathsf{SVD}_R(\widehat{\boldsymbol{M}})$) satisfies*

$$\mathsf{SVD}_R(\widehat{\boldsymbol{M}}) - \boldsymbol{M} = \boldsymbol{Z} - \boldsymbol{U}_\perp \boldsymbol{U}_\perp^\top \boldsymbol{Z} \boldsymbol{V}_\perp \boldsymbol{V}_\perp^\top + \boldsymbol{R},$$

where $\boldsymbol{Z} = \widehat{\boldsymbol{M}} - \boldsymbol{M}$ and $\boldsymbol{U}_\perp$ ($\boldsymbol{V}_\perp$ resp.) is the orthogonal complement of $\boldsymbol{U}$ ($\boldsymbol{V}$ resp.), and the remainder term $\boldsymbol{R}$ satisfies

$$\|\boldsymbol{R}\|_{\mathrm{F}} \leq 40 \frac{\|\widehat{\boldsymbol{M}} - \boldsymbol{M}\|_{\mathrm{F}}^2}{\lambda_{\min}}.$$

Proof. See the proof of Lemma 18 in Shen et al. (2025). □

Lemma 6. *Under Regime 1, suppose $\rho(\boldsymbol{M}) < 1$. And let $\boldsymbol{E}_t$ be i.i.d. with mean zero and $\mathbb{E}|[\boldsymbol{E}_t]_{\boldsymbol{i}}|^{2+\widetilde{q}} < +\infty$ for each $\boldsymbol{i} \in \mathcal{S}$ and some $\widetilde{q} > 0$. Then we have*

$$\frac{1}{T-1} \sum_{t=1}^{T-1} \mathsf{vec}(\boldsymbol{X}_t) \mathsf{vec}(\boldsymbol{X}_t)^\top \xrightarrow{p} \boldsymbol{\Gamma}_0$$

as $T \to \infty$.

Proof. See Theorem 8.2.3 in Fuller (2009). □

S3 Additional Methodological and Numerical Details

This section collects additional methodological details and numerical results.

S3.1 Non-identifiability under non-nested neighborhoods

In this section, we give the formal statement of the ambiguity for neighborhood selection within non-nested candidates.

Proposition 1. *Let $K_1 > 0, K_2 \geq 0$. Then there exists a distribution $\mathcal{P}$, and $\boldsymbol{X}_t \overset{i.i.d.}{\sim} \mathcal{P}$, such that there exists $k_1 < K_1, k_2 > K_2$, and $\widehat{\boldsymbol{M}} \in \mathbb{R}^{\mathcal{S}}$, the following holds:*

- $(2k_1+1)(2k_2+1) < (2K_1+1)(2K_2+1)$,
- $\mathit{supp}(\widehat{\boldsymbol{M}}) = \widetilde{\mathcal{J}}$,
- $y_t = \langle \boldsymbol{X}_t, \widehat{\boldsymbol{M}} \rangle$.

Proof. We set $k_1 = 0, k_2 = K_2 + 1$ (see Figure 1). Let $\boldsymbol{X}$ be supported on only two entries $\{(i_0 - K_1, j_0), (i_0, j_0 - K_2 - 1)\}$ and that $\boldsymbol{X}_t(i_0 - K_1, j_0) = \boldsymbol{X}_t(i_0, j_0 - K_2 - 1) \sim \mathcal{P}_0$, where $\mathcal{P}_0$ can be an arbitrary distribution. And let $\widehat{\boldsymbol{M}} = \boldsymbol{M}(i_0 - K_1, j_0)\boldsymbol{e}_{i_0}\boldsymbol{e}_{j_0-K_2-1}^\top$. Then we have $y_t = \langle \boldsymbol{X}_t, \boldsymbol{M} \rangle = \boldsymbol{X}_t(i_0 - K_1, j_0) \cdot \boldsymbol{M}(i_0 - K_1, j_0) = \langle \boldsymbol{X}_t, \widehat{\boldsymbol{M}} \rangle$. □

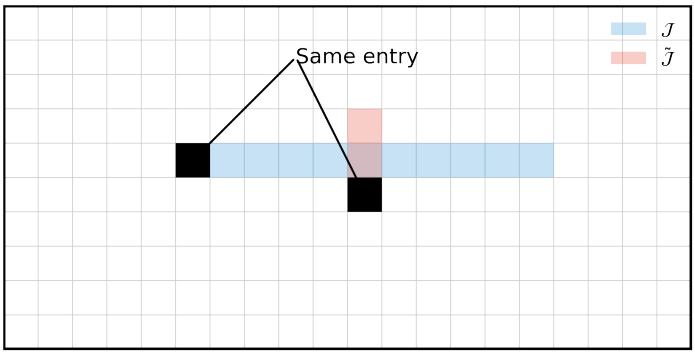


Figure 1: An example with the ground-truth bandwidths $K_1 = 5, K_2 = 0$ (in black), and $k_1 = 0, k_2 = 1$ (in red).

S3.2 Extension to the general lag-P model

This subsection collects the extension of LIAR to the general lag-P setting.

S3.2.1 Estimation for the general lag-P model

We introduce the estimation for the general lag P model. We work on the following model:

$$\boldsymbol{X}_t = \sum_{\boldsymbol{i} \in \mathcal{S}} \sum_{p \in [P]} \langle \boldsymbol{X}_{t-p}, \boldsymbol{M}_{p,\boldsymbol{i}} \rangle \ \boldsymbol{e}_{i_1} \boldsymbol{e}_{i_2}^\top + \boldsymbol{E}_t, \tag{S3.7}$$

For each $\boldsymbol{i} \in \mathcal{S}$, we denote $\boldsymbol{x}_t^{(\boldsymbol{i})} := \boldsymbol{X}_t(\mathcal{J}_{\boldsymbol{i}}), \boldsymbol{m}_p^{(\boldsymbol{i})} = \boldsymbol{M}_{p,\boldsymbol{i}}(\mathcal{J}_{\boldsymbol{i}}) \in \mathbb{R}^{|\mathcal{J}_{\boldsymbol{i}}|}$. Define the stacked coefficient and per-time design vectors

$$\boldsymbol{m}_{\boldsymbol{i}} := \begin{bmatrix} \boldsymbol{m}_1^{(\boldsymbol{i})} \\ \vdots \\ \boldsymbol{m}_P^{(\boldsymbol{i})} \end{bmatrix}, \quad \boldsymbol{r}_t^{(\boldsymbol{i})} := \begin{bmatrix} \boldsymbol{x}_{t-1}^{(\boldsymbol{i})} \\ \vdots \\ \boldsymbol{x}_{t-P}^{(\boldsymbol{i})} \end{bmatrix} \in \mathbb{R}^{P|\mathcal{J}_{\boldsymbol{i}}|}$$

Then the scalar regression for entry $\boldsymbol{i}$ reads, for $t = P+1, \ldots, T$, $[\boldsymbol{X}_t]_{\boldsymbol{i}} = \langle \boldsymbol{r}_t^{(\boldsymbol{i})}, \boldsymbol{m}_{\boldsymbol{i}} \rangle + [\boldsymbol{E}_t]_{\boldsymbol{i}}$. Stacking over t gives the standard linear model $\boldsymbol{z}_{\boldsymbol{i}} = \boldsymbol{Y}_{\boldsymbol{i}} \boldsymbol{m}_{\boldsymbol{i}} + \boldsymbol{\nu}_{\boldsymbol{i}}$, where

$$\boldsymbol{z}_{\boldsymbol{i}} = \begin{bmatrix} [\boldsymbol{X}_{P+1}]_{\boldsymbol{i}} \\ \vdots \\ [\boldsymbol{X}_T]_{\boldsymbol{i}} \end{bmatrix}, \qquad \boldsymbol{\nu}_{\boldsymbol{i}} = \begin{bmatrix} [\boldsymbol{E}_{P+1}]_{\boldsymbol{i}} \\ \vdots \\ [\boldsymbol{E}_T]_{\boldsymbol{i}} \end{bmatrix} \in \mathbb{R}^{T-P},$$

$$\boldsymbol{Y}_{\boldsymbol{i}} = \begin{bmatrix} \boldsymbol{r}_{P+1}^{(\boldsymbol{i})\top} \\ \vdots \\ \boldsymbol{r}_T^{(\boldsymbol{i})\top} \end{bmatrix} \in \mathbb{R}^{(T-P)\times P|\mathcal{J}_{\boldsymbol{i}}|}. \tag{S3.8}$$

The least square estimator is

$$\widehat{\boldsymbol{m}_{\boldsymbol{i}}} = \arg\min_{\boldsymbol{m}} \frac{1}{2} \|\boldsymbol{z}_{\boldsymbol{i}} - \boldsymbol{Y}_{\boldsymbol{i}} \boldsymbol{m}\|_{\ell_2}^2 = (\boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{Y}_{\boldsymbol{i}})^{-1} \boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{z}_{\boldsymbol{i}}.$$

Notice this procedure can be done in parallel for different $\boldsymbol{i} \in \mathcal{S}$. We summarize the procedure in Algorithm 1.

Algorithm 1 Parallel Least Square for General Lag P

Input: Data $\{\boldsymbol{X}_t\}_{t=1}^T$, local neighborhood $\{\mathcal{J}_{\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathcal{S}}$, lag parameter P

ParFor $\boldsymbol{i} \in \mathcal{S}$

Collect the covariate $\boldsymbol{Y}_{\boldsymbol{i}}$ and response $\boldsymbol{z}_{\boldsymbol{i}}$ defined in (S3.8)

Solve least square $\widehat{\boldsymbol{m}}_{\boldsymbol{i}} = (\boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{Y}_{\boldsymbol{i}})^{-1}\boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{z}_{\boldsymbol{i}}$

End ParFor

Output: $\{\widehat{\boldsymbol{m}}_{\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathcal{S}}$

Estimation of Coefficients for SP-LIAR. For SP-LIAR, $\mathcal{J}_{\boldsymbol{i}}$ are rectangles by construction, and we denote $\boldsymbol{M}_p^{[\boldsymbol{i}]} := \boldsymbol{M}_{p,\boldsymbol{i}}[\mathcal{J}_{\boldsymbol{i}}]$ as a sub-matrix of $\boldsymbol{M}_{p,\boldsymbol{i}}$. Arrange these matrices into the block matrix

$$\boldsymbol{M}_p^{\mathsf{blk}} = \begin{bmatrix} \boldsymbol{M}_p^{[1,1]} & \cdots & \boldsymbol{M}_p^{[1,N]} \\ \vdots & & \vdots \\ \boldsymbol{M}_p^{[M,1]} & \cdots & \boldsymbol{M}_p^{[M,N]} \end{bmatrix}.$$

Then under SP-LIAR, $\boldsymbol{M}_p^{\mathsf{blk}}$ has rank at most R. This motivates us to consider the best rank R approximation of the following estimator of $\boldsymbol{M}_p^{\mathsf{blk}}$, where $\widehat{\boldsymbol{m}}_{\boldsymbol{i}} = \begin{bmatrix} \widehat{\boldsymbol{m}}_1^{(\boldsymbol{i})\top} & \cdots & \widehat{\boldsymbol{m}}_P^{(\boldsymbol{i})\top} \end{bmatrix}^\top$:

$$\widehat{\boldsymbol{M}}_p^{\mathsf{blk}} = \begin{bmatrix} \mathsf{mat}(\widehat{\boldsymbol{m}}_p^{(1,1)}) & \cdots & \mathsf{mat}(\widehat{\boldsymbol{m}}_p^{(1,N)}) \\ \vdots & & \vdots \\ \mathsf{mat}(\widehat{\boldsymbol{m}}_p^{(M,1)}) & \cdots & \mathsf{mat}(\widehat{\boldsymbol{m}}_p^{(M,N)}) \end{bmatrix},$$

where mat is the inverse operation of vec. Our algorithm is summarized in Algorithm 2.

Algorithm 2 Parallel Least Square for SP-LIAR

Input: Local neighborhood $\mathcal{J}_{\boldsymbol{i}}$, data $\{\boldsymbol{X}_t\}_{t=1}^T$, lag parameter P, rank R

Run Algorithm 1 for $\{\widehat{\boldsymbol{m}}_{\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathcal{S}}$

Perform rank R SVD approximation on $\widehat{\boldsymbol{M}}_p^{\mathsf{blk}}$ and get: $\widehat{\boldsymbol{M}}_{p,R}^{\mathsf{blk}} = \mathsf{SVD}_R(\widehat{\boldsymbol{M}}_p^{\mathsf{blk}})$

Output: $\{\widehat{\boldsymbol{M}}_{p,R}^{(\boldsymbol{i})}\}_{p\in[P],\boldsymbol{i}\in\mathcal{S}}$, where $\widehat{\boldsymbol{M}}_{p,R}^{(\boldsymbol{i})}$ be the (i_1,i_2)-th block of $\widehat{\boldsymbol{M}}_{p,R}^{\mathsf{blk}}$

S3.2.2 Bandwidth selection for the general lag-P model

We also restrict the search to a nested sequence indexed by k for each location $\boldsymbol{i} = (i_1, i_2)$:

$$\mathcal{J}_{\boldsymbol{i}}[1] \subsetneq \cdots \subsetneq \mathcal{J}_{\boldsymbol{i}}[k_0] \subsetneq \cdots \subsetneq \mathcal{J}_{\boldsymbol{i}}[K_0] \subset \mathcal{S},$$

where $\mathcal{J}_{\boldsymbol{i}}[k_0] = \mathcal{J}_{\boldsymbol{i}}$, and K_0 is a prescribed cap. Write $s_k = |\mathcal{J}_{\boldsymbol{i}}[k]|$ (depends on k, and implicitly on $\boldsymbol{i}$). For any level k, let $\boldsymbol{x}_t[k] := \boldsymbol{X}_t\big(\mathcal{J}_{\boldsymbol{i}}[k]\big) \in \mathbb{R}^{s_k}$ and stack P lags into one vector $\boldsymbol{r}_t[k] = \begin{bmatrix}\boldsymbol{x}_t[k]^\top & \cdots & \boldsymbol{x}_{t-P+1}[k]^\top\end{bmatrix}^\top \in \mathbb{R}^{Ps_k}$. Then we form the design matrix $\boldsymbol{Y}[k] \in \mathbb{R}^{(T-P)\times Ps_k}$:

$$\boldsymbol{Y}[k] = \begin{bmatrix}\boldsymbol{r}_{P+1}[k] & \cdots & \boldsymbol{r}_T[k]\end{bmatrix}^\top.$$

Also recall $\boldsymbol{z} = \begin{bmatrix}[\boldsymbol{X}_{P+1}]_{\boldsymbol{i}} & \cdots & [\boldsymbol{X}_T]_{\boldsymbol{i}}\end{bmatrix}^\top$. The residual sum of squares is

$$\mathsf{RSS}_{\boldsymbol{i}}(k) := \|\boldsymbol{z} - \boldsymbol{Y}[k](\boldsymbol{Y}[k]^\top\boldsymbol{Y}[k])^{-1}\boldsymbol{Y}[k]^\top\boldsymbol{z}\|_{\ell_2}^2.$$

And we use the Bayesian information criterion for the bandwidth selection.

$$\mathsf{BIC}_{\boldsymbol{i}}(k) = \log \mathsf{RSS}_{\boldsymbol{i}}(k) + D_0 \, \frac{|\mathcal{J}_{\boldsymbol{i}}[k]| \cdot P}{T} \, \log(M \vee N \vee T). \tag{S3.9}$$

For each $\boldsymbol{i} \in \mathcal{S}$, we select $\widehat{k}_{\boldsymbol{i}}$ by $\widehat{k}_{\boldsymbol{i}} = \arg\min_{k \le K_0} \ \mathsf{BIC}_{\boldsymbol{i}}(k)$.

S3.3 Additional Simulation Results

S3.3.1 Visualization of the CLT

To illustrate the asymptotic normality result in Theorem 1 of the main paper, we conduct an oracle diagnostic under the same data-generating mechanism as the estimation-accuracy experiment in Section 5. Specifically, we consider the lag-one LIAR model with $P = 1$, $(M, N) = (10, 10)$, and a square neighborhood with radius $K = 3$. The local coefficients are generated as in Section 5 and the resulting transition matrix is rescaled to have spectral radius smaller than one. The innovations are Gaussian with the same noise level as in Section 5, and the first 500 observations are discarded as burn-in.

We focus on the interior location $\boldsymbol{i}_0 = (5, 5)$. Its true neighborhood $\mathcal{J}_{\boldsymbol{i}_0}$ is not truncated, so $|\mathcal{J}_{\boldsymbol{i}_0}| = (2K + 1)^2 = 49$. Let j_0 denote the coordinate of $\boldsymbol{m}_{\boldsymbol{i}_0}$ corresponding to the self-lag entry at $\boldsymbol{i}_0$. Under the column-major ordering used for $\mathcal{J}_{\boldsymbol{i}_0}$, this coordinate is $j_0 = 25$ in one-based indexing.

For each sample size $T \in \{500, 1000, 2000, 5000\}$, we generate $B = 500$

independent replications. In each replication, we estimate $\boldsymbol{m}_{\boldsymbol{i}_0}$ by least squares using the oracle true neighborhood $\mathcal{J}_{\boldsymbol{i}_0}$, rather than a BIC-selected neighborhood. This construction isolates the least-squares CLT from the additional effect of neighborhood selection. We then compute the studentized statistic

$$Z_{\boldsymbol{i}_0,j_0} = \frac{\widehat{\boldsymbol{m}}_{\boldsymbol{i}_0,j_0} - \boldsymbol{m}_{\boldsymbol{i}_0,j_0}}{\widehat{\sigma}_{\boldsymbol{i}_0}\left\{\left[(\boldsymbol{Y}_{\boldsymbol{i}_0}^\top \boldsymbol{Y}_{\boldsymbol{i}_0})^{-1}\right]_{j_0,j_0}\right\}^{1/2}},$$

where

$$\widehat{\sigma}_{\boldsymbol{i}_0}^2 = \frac{\|\boldsymbol{z}_{\boldsymbol{i}_0} - \boldsymbol{Y}_{\boldsymbol{i}_0}\widehat{\boldsymbol{m}}_{\boldsymbol{i}_0}\|_{\ell_2}^2}{T - 1 - |\mathcal{J}_{\boldsymbol{i}_0}|}.$$

Figure 2 compares the empirical distributions of $Z_{\boldsymbol{i}_0,j_0}$ with the standard normal benchmark across sample sizes. The distributions are centered around zero and become increasingly close to the standard normal density as T increases.

S3.3.2 Additional numerical results for neighborhood selection

After the oracle CLT diagnostic above, we next consider a larger-sample version of the BIC selection experiment in Section 5. The data-generating mechanism is unchanged: the true neighborhood radius is $K = 3$, and BIC selects over $k \in \{0, 1, 2, 3, 4, 5\}$. Only the sample size is increased to $T \in \{4000, 5000, 6000\}$.

Figure 3 reports the per-pixel success rates for $M = N \in \{10, 15, 20\}$.

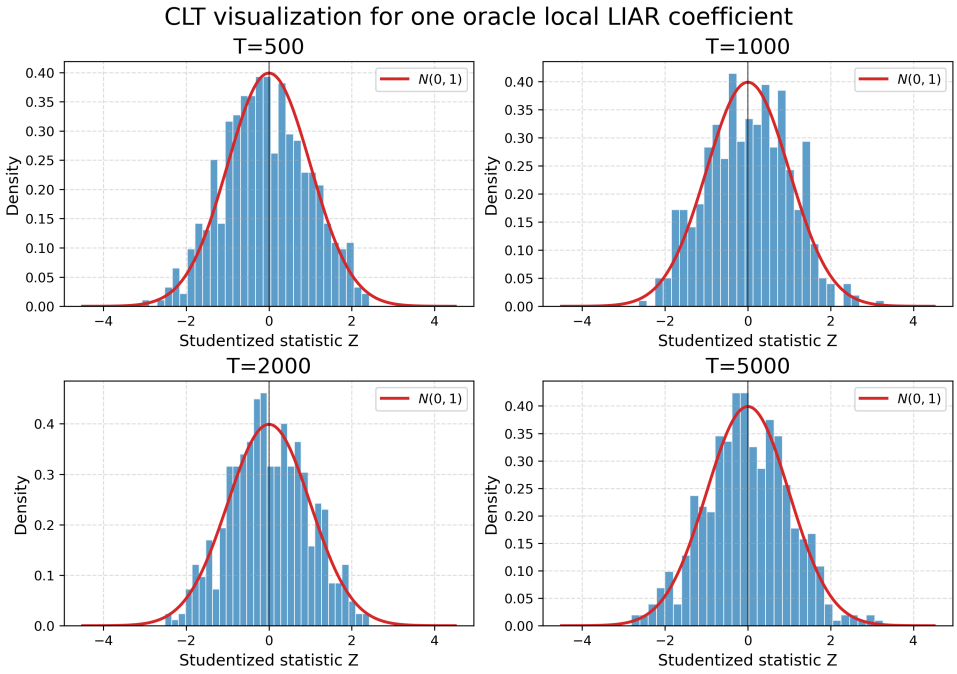


Figure 2: Visualization of the CLT for the representative coefficient $\boldsymbol{m}_{\boldsymbol{i}_0,j_0}$. The empirical distributions of $Z_{\boldsymbol{i}_0,j_0}$ are compared with the standard normal benchmark across sample sizes.

Compared with the moderate-sample results in the main paper, the success rates further improve as T increases. These results are consistent with the neighborhood-selection consistency result in Theorem 5.

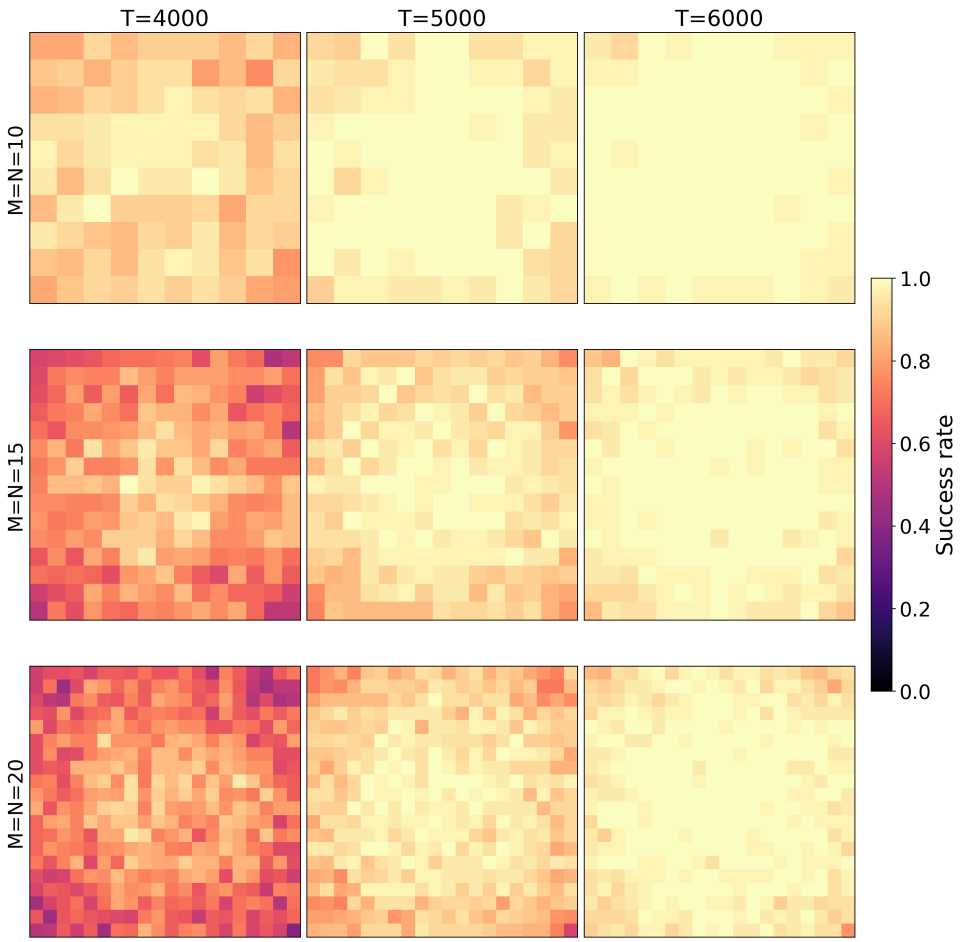


Figure 3: Additional neighborhood-selection results for $M = N \in \{10, 15, 20\}$ and $T \in \{4000, 5000, 6000\}$.

S4 Tensor Local Interaction Autoregression and TEC Application

This section collects the tensor extension of LIAR and the tensor TEC application. We first formulate the tensor local interaction autoregression and its entrywise least-squares estimator, and then apply the same construction to TEC data organized by two spatial modes and one within-hour temporal mode.

S4.1 Tensor LIAR formulation and estimation

We consider a local interaction autoregressive model for tensor-valued time series. The modeling idea mirrors the matrix case: each tensor entry evolves from past values in a local neighborhood, while the neighborhood itself may vary across entries and across tensor modes.

We use calligraphic-font bold-face letters, such as $\boldsymbol{\mathcal{M}}$ and $\boldsymbol{\mathcal{X}}$, to denote tensors. An order-d tensor is a d-way array; for example, $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{N_1 \times \cdots \times N_d}$ has size N_j along its jth mode. Let $\mathcal{S}_d = [N_1] \times \cdots \times [N_d]$ be the tensor index set. For a tensor time series $\{\boldsymbol{\mathcal{X}}_t\}_{t \geq 0} \subset \mathbb{R}^{\mathcal{S}_d}$, the tensor LIAR takes the form

$$\boldsymbol{\mathcal{X}}_t = \sum_{\boldsymbol{i} \in \mathcal{S}_d} \langle \boldsymbol{\mathcal{X}}_{t-1}, \boldsymbol{\mathcal{M}}_{\boldsymbol{i}} \rangle \cdot \boldsymbol{e}_{i_1} \circ \cdots \circ \boldsymbol{e}_{i_d} + \boldsymbol{\mathcal{E}}_t,$$

where $\boldsymbol{i} = (i_1, \ldots, i_d) \in \mathcal{S}_d$, and $\boldsymbol{\mathcal{M}}_{\boldsymbol{i}}$ is supported on a local neighborhood $\mathcal{J}_{\boldsymbol{i}}$ of $\boldsymbol{i}$. For example, a rectangular neighborhood with radius vector $(k_{1,\boldsymbol{i}}, \ldots, k_{d,\boldsymbol{i}})$ contains indices whose ℓth coordinate lies within $k_{\ell,\boldsymbol{i}}$ of i_ℓ, truncated at the tensor boundary.

Vectorizing both sides gives

$$\boldsymbol{x}_t = \underbrace{\sum_{\boldsymbol{i} \in \mathcal{S}_d} (\boldsymbol{e}_{i_d} \otimes \cdots \otimes \boldsymbol{e}_{i_1}) \mathsf{vec}(\boldsymbol{\mathcal{M}}_{\boldsymbol{i}})^\top}_{=: \boldsymbol{M}} \boldsymbol{x}_{t-1} + \boldsymbol{\epsilon}_t,$$

where $\boldsymbol{x}_t = \mathsf{vec}(\boldsymbol{\mathcal{X}}_t)$ and $\boldsymbol{\epsilon}_t = \mathsf{vec}(\boldsymbol{\mathcal{E}}_t)$. Denote the local covariate vector and coefficient vector by

$$\boldsymbol{x}_t^{(\boldsymbol{i})} = \boldsymbol{\mathcal{X}}_t(\mathcal{J}_{\boldsymbol{i}}) \in \mathbb{R}^{|\mathcal{J}_{\boldsymbol{i}}|}, \qquad \boldsymbol{m}_{\boldsymbol{i}} = \boldsymbol{\mathcal{M}}_{\boldsymbol{i}}(\mathcal{J}_{\boldsymbol{i}}) \in \mathbb{R}^{|\mathcal{J}_{\boldsymbol{i}}|}.$$

Then, for each entry $\boldsymbol{i}$ and $t = 2, \ldots, T$, the tensor model reduces to the scalar regression

$$[\boldsymbol{\mathcal{X}}_t]_{\boldsymbol{i}} = \langle \boldsymbol{x}_{t-1}^{(\boldsymbol{i})}, \boldsymbol{m}_{\boldsymbol{i}} \rangle + [\boldsymbol{\mathcal{E}}_t]_{\boldsymbol{i}}.$$

Stacking over time gives the standard linear model

$$\boldsymbol{z}_{\boldsymbol{i}} = \boldsymbol{Y}_{\boldsymbol{i}} \boldsymbol{m}_{\boldsymbol{i}} + \boldsymbol{\nu}_{\boldsymbol{i}},$$

where

$$\boldsymbol{z}_{\boldsymbol{i}} = \begin{bmatrix} [\boldsymbol{\mathcal{X}}_2]_{\boldsymbol{i}} & \cdots & [\boldsymbol{\mathcal{X}}_T]_{\boldsymbol{i}} \end{bmatrix}^\top, \qquad \boldsymbol{\nu}_{\boldsymbol{i}} = \begin{bmatrix} [\boldsymbol{\mathcal{E}}_2]_{\boldsymbol{i}} & \cdots & [\boldsymbol{\mathcal{E}}_T]_{\boldsymbol{i}} \end{bmatrix}^\top \in \mathbb{R}^{T-1},$$

$$\boldsymbol{Y}_{\boldsymbol{i}} = \begin{bmatrix} \boldsymbol{x}_1^{(\boldsymbol{i})} & \cdots & \boldsymbol{x}_{T-1}^{(\boldsymbol{i})} \end{bmatrix}^\top \in \mathbb{R}^{(T-1) \times |\mathcal{J}_{\boldsymbol{i}}|}.$$

The least-squares estimator is

$$\widehat{\boldsymbol{m}}_{\boldsymbol{i}} = \arg\min_{\boldsymbol{m}} \frac{1}{2}\|\boldsymbol{z}_{\boldsymbol{i}} - \boldsymbol{Y}_{\boldsymbol{i}}\boldsymbol{m}\|_{\ell_2}^2 = (\boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{Y}_{\boldsymbol{i}})^{-1}\boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{z}_{\boldsymbol{i}}.$$

Algorithm 3 summarizes the parallel implementation.

Algorithm 3 Parallel least squares for tensor LIAR

Input: Local neighborhoods $\{\mathcal{J}_{\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathcal{S}_d}$, data $\{\boldsymbol{\mathcal{X}}_t\}_{t=1}^T$

ParFor $\boldsymbol{i} \in \mathcal{S}_d$

Collect the covariate $\boldsymbol{Y}_{\boldsymbol{i}}$ and response $\boldsymbol{z}_{\boldsymbol{i}}$

Solve least squares $\widehat{\boldsymbol{m}}_{\boldsymbol{i}} = (\boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{Y}_{\boldsymbol{i}})^{-1}\boldsymbol{Y}_{\boldsymbol{i}}^\top \boldsymbol{z}_{\boldsymbol{i}}$

End ParFor

Output: $\{\widehat{\boldsymbol{m}}_{\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathcal{S}_d}$

Theoretical results for coefficient estimation, auto-covariance estimation, and neighborhood selection extend to this tensor setting by the same proof strategy as in the matrix case. The framework also allows tensor autoregressions with no locality along some modes, by choosing $\mathcal{J}_{\boldsymbol{i}}$ to span the full range along those dimensions while remaining local along others.

S4.2 Tensor TEC application

The TEC application uses the tensor formulation above to separate coarse spatial variation from within-hour temporal variation. Starting from the same GNSS-derived TEC product (Sun et al., 2023), we first aggregate the spatial grid to $4^\circ \times 4^\circ$, yielding a 46×91 spatial grid. We then stack,

for each clock hour, the twelve 5-minute frames into a third mode, which represents the within-hour slot. This produces hourly tensors of size $46 \times 91 \times 12$. Indexing these tensors by hour gives a tensor time series of shape $46 \times 91 \times 12 \times 720$ for the 30-day period in June 2019.

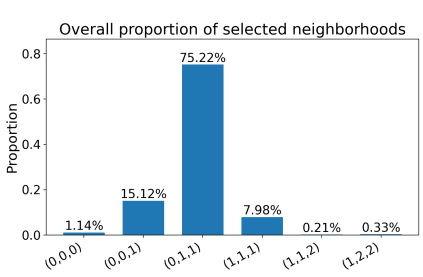
We use the first 80% of hours for training and the remaining 20% for testing. Test RMSE is computed from one-step-ahead full-tensor predictions over the test period, averaging squared prediction errors over all test hours and tensor entries. Neighborhood sizes are selected entrywise by BIC over the candidate set

$$\{(0,0,0),(0,0,1),(0,1,1),(1,1,1),(1,1,2),(1,2,2),(2,2,2)\}.$$

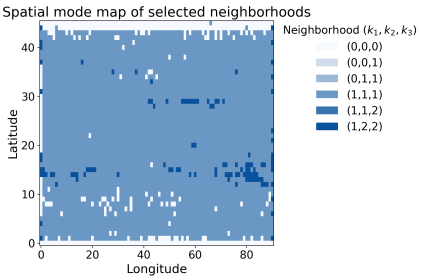
Figure 4a reports the selection frequencies across all locations and within-hour slots. The selections concentrate on small neighborhoods, with $(0,1,1)$ most common and $(0,0,1)$ the primary alternative. This concentration suggests that the hourly TEC tensor dynamics are driven mainly by short-range interactions along the spatial and within-hour dimensions, rather than by dense full-tensor dependence.

To examine when and where these selected neighborhoods arise, Figure 4c shows the proportion of each selected neighborhood versus within-hour slot, from 5 to 60 minutes. The slot-wise proportions are fairly stable across the hour, with only mild changes near hour boundaries. This stabil-

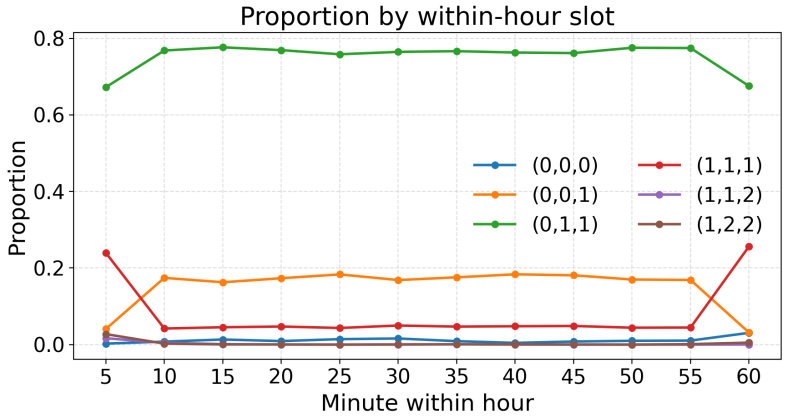
ity suggests that the learned local dependence pattern is persistent within an hour, with only modest changes near transitions between adjacent hourly blocks. Figure 4b displays the spatial mode map, defined as the most frequently selected neighborhood per pixel over the 12 within-hour slots.



(a) Selection frequencies of neighborhoods (k_1, k_2, k_3) over all locations and slots.



(b) Spatial mode map: per pixel, the most frequently selected neighborhood over 12 slots.



(c) Proportion of each neighborhood versus within-hour slot, from 5 to 60 minutes.

Figure 4: Neighborhood selection in the tensor TEC series using entrywise BIC. Top: (a) overall selection frequencies and (b) spatial mode map. Bottom: (c) proportions by within-hour slot.

As a benchmark, we fit the pixel-wise autoregression LIAR-P. Its test RMSE is 1.29, whereas LIAR with BIC-selected neighborhoods attains a lower test RMSE of 1.12. Thus, the tensor neighborhoods selected by BIC capture spatio-temporal interactions that are useful for one-step-ahead full-tensor prediction.

Bibliography

Fuller, W. A. (2009). *Introduction to statistical time series*. John Wiley & Sons.

Liu, W., H. Xiao, and W. B. Wu (2013). Probability and moment inequalities under dependence. *Statistica Sinica*, 1257–1272.

Shen, Y., J. Li, J.-F. Cai, and D. Xia (2025). Computationally efficient and statistically optimal robust high-dimensional linear regression. *The Annals of Statistics 53*(1), 374–399.

Sun, H., Y. Chen, S. Zou, J. Ren, Y. Chang, Z. Wang, and A. Coster (2023). Complete global total electron content map dataset based on a video imputation algorithm vista. *Scientific Data 10*(1), 236.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences 102*(40), 14150–14154.