Submitted to Bernoulli

Supplement to "On Consistency and Sparsity for High-Dimensional Functional Time Series with Application to Autoregressions"

SHAOJUN GUO[1] and XINGHAO QIAO[2]

[1] *Institute of Statistics and Big Data, Renmin University of China, Beijing, 100872, P.R. China. E-mail:* sjguo@ruc.edu.cn

[2] *Department of Statistics, London School of Economics, London, WC2A 2AE, U.K. E-mail:* x.qiao@lse.ac.uk

This supplementary material contains the discussion of several potential extensions in Section A, examples for infinite-dimensional functional data satisfying Condition 2 in Section B, proofs of main theorems in Section C, additional technical proofs in Section D, derivations of functional stability measure for the illustrative VFAR(1) example in Section E, some derivations for VFAR models in Section F, details of the algorithms to fit sparse VFAR models in Section G and additional empirical results in Section H.

A. Discussion of potential extensions

We identify several important directions for future research. The first topic considers the functional extension of high-dimensional factor models (Bai and Ng, 2002; Lam and Yao, 2012), where the observations, $\mathbf{X}_t(\cdot)$'s, can be decomposed as the sum of two unobservable and mutually orthogonal components

$$\mathbf{X}_t(u) = \mathbf{W}_t(u) + \boldsymbol{\epsilon}_t(u),\ t = 1, \ldots, n,\ u \in \mathcal{U}. \tag{A.1}$$

Here $\mathbf{W}_t(\cdot) = \mathbf{B}\mathbf{f}_t(\cdot)$ are the common components driven by r (much smaller than p) functional factors $\mathbf{f}_t(\cdot) = \big(f_{t1}(\cdot), \ldots, f_{tr}(\cdot)\big)^{\mathrm{T}}$, $\mathbf{B} \in \mathbb{R}^{p\times r}$ is the factor loading matrix and $\boldsymbol{\epsilon}_t(\cdot) = \big(\epsilon_{t1}(\cdot), \ldots, \epsilon_{tp}(\cdot)\big)^{\mathrm{T}}$ are idiosyncratic components. For each $h \in \mathbb{Z}$ and $\{\mathbf{X}_t(\cdot)\}_{t=1}^n$, denote $\boldsymbol{\Sigma}_h^X(u, v) = \text{Cov}\{\mathbf{X}_t(u), \mathbf{X}_{t+h}(v)\}$ and its sample estimator by $\widehat{\boldsymbol{\Sigma}}_h^X(u, v)$. To estimate such functional factor model (A.1), we discuss two different approaches. (i) The first one is based on the following integrated covariance decomposition,

$$\int\int \boldsymbol{\Sigma}_0^X(u, v)dudv = \mathbf{B}\Big\{\int\int \boldsymbol{\Sigma}_0^f(u, v)dudv\Big\}\mathbf{B}^{\mathrm{T}} + \int\int \boldsymbol{\Sigma}_0^{\epsilon}(u, v)dudv. \tag{A.2}$$

Intuitively, by imposing some eigenvalue conditions on two terms on the right-hand side of (A.2) similar to those in Fan, Liao and Mincheva (2013), the above decomposition is asymptotically identified as $p \to \infty$ and hence $\mathbf{B}$ can be recovered by performing an eigenanalysis of $\iint \widehat{\boldsymbol{\Sigma}}_0^X(u,v)dudv$. (ii) If $\{\boldsymbol{\epsilon}_t(\cdot)\}$ follows a white noise process, then, inspired from Lam and Yao (2012) and the fact that $\boldsymbol{\Sigma}_h^X(u,v) = \mathbf{B}\boldsymbol{\Sigma}_h^f(u,v)\mathbf{B}^{\mathrm{T}}$ for $h \geqslant 1$, an autocovariance-based procedure can be developed to estimate model (A.1). To theoretically support both estimation procedures under high-dimensional settings, the main challenge is to investigate convergence properties of $\widehat{\boldsymbol{\Sigma}}_h^X - \boldsymbol{\Sigma}_h^X$ for $h = 0, 1, \ldots,$ and hence our concentration results in Theorem 1 and Proposition 1 can be applied.

Second, in the dimension reduction step of the three-step procedure, one can also perform dynamic FPCA (Hörmann, Kidziński and Hallin, 2015) based on $\{X_{tj}(\cdot)\}_{t=1}^n$ for each j. Such dimension reduction technique provides an optimal truncated approximation for functional time series, but is computationally intensive as it relies on the eigenanalysis of spectral density functions $f_{\mathbf{X},\theta}$ with sample estimators given by $\widehat{f}_{\mathbf{X},\theta} = (2\pi)^{-1}\sum_h w_H(h)\widehat{\boldsymbol{\Sigma}}_h \exp(-ih\theta)$, where $w_H(\cdot) = w(\cdot/H)$ is some appropriate weight function with H (the lag window size). To provide theoretical guarantees for relevant estimated terms under a dynamic FPCA framework similar to those in Theorems 3 and 4, our established non-asymptotic error bounds on $\widehat{\boldsymbol{\Sigma}}_h$ for $h \in \mathbb{Z}$ become applicable.

Third, within the proposed functional stability measure framework, we believe our established concentration results can be extended beyond Gaussian functional time series to accommodate linear processes with functional sub-Gaussian errors. It is also interesting to develop suitable concentration results for heavy-tailed non-Gaussian functional time series under the functional generalization of the β-mixing condition in Wong, Li and Tewari (2020).

Fourth, it is of great interest to develop new inference tools for high-dimensional functional time series and to apply these techniques to quantify deviations of autocoefficient functions in sparse VFAR models. Fifth, our analysis is based on the estimation where smoothness parameters are assumed to be known. It is interesting to develop adaptive estimation procedures that do not require the knowledge of the parameter space and automatically adjust to the smoothness properties. However, this would pose complicated challenges under the high-dimensional, functional and dependent setting we consider.

These topics are beyond the scope of the current paper and will be pursued elsewhere.

B. Examples satisfying Condition 2

It is clear that Condition 2 holds for finite-dimensional functional data. In this section, we give some illustrative examples for infinite-dimensional functional data, where the upper bounds of their functional stability measures can be easily controlled. As long as the denominator of $\mathcal{M}(f_{\mathbf{X}})$ is arbitrarily small, the numerator can also be arbitrarily small, Condition 2 in this sense still holds for a large class of infinite-dimensional functional data including, but not limited to the examples below.

Consider the functional linear process $\mathbf{X}_t(\cdot) = \sum_{l=0}^{\infty} \mathbf{A}_l(\boldsymbol{\varepsilon}_{t-l})(\cdot)$ (Fang, Guo and Qiao, 2021). Denote the polynomial $\mathcal{B}(z)(u,v) = \sum_{l=0}^{\infty} \mathbf{A}_l(u,v)z^l$ for $u, v \in \mathcal{U}$. We can derive the spectral density matrix function as

$$\mathbf{f}_{\mathbf{X},\theta}(u,v) = \frac{1}{2\pi}\int\int \mathcal{B}\left(e^{-i\theta}\right)(u,u')\boldsymbol{\Sigma}_0^{\varepsilon}(u',v')\mathcal{B}\left(e^{-i\theta}\right)^*(v,v')du'dv' \tag{B.3}$$

and the covariance matrix function as

$$\boldsymbol{\Sigma}_0(u,v) = \sum_{l=0}^{\infty}\int\int \mathbf{A}_l(u,u')\boldsymbol{\Sigma}_0^{\varepsilon}(u',v')\mathbf{A}_l^*(v,v')du'dv', \tag{B.4}$$

where $\boldsymbol{\Sigma}_0^{\varepsilon}(u,v) = \mathrm{Cov}(\boldsymbol{\varepsilon}_t(u), \boldsymbol{\varepsilon}_t(v))$ and $*$ denotes the conjugate. Note that the functional stability measure $\mathcal{M}(f_{\mathbf{X}})$ is defined in (3) based on the functional Rayleigh quotients of $\mathbf{f}_{\mathbf{X},\theta}$ in (B.3) relative to $\boldsymbol{\Sigma}_0$ in (B.4). We can express $\mathcal{M}(f_{\mathbf{X}})$ based on (B.3) and (B.4). However, unlike VAR or *vector moving average* (VMA) model, it seems difficult to give explicit expressions of $\mathcal{M}(f_{\mathbf{X}})$ for functional versions of VAR or VMA model to accommodate multivariate functional time series, since one need to derive expressions in terms of abstract functional analysis language in the direct sum of multiple Hilbert spaces rather than the compact matrix forms for multivariate scalar time series, e.g. Basu and Michailidis (2015). We first give two specific examples of one-dimensional case and derive the upper bounds on the functional stability measure, which help us to understand the usefulness of our framework.

Example 1: Consider the functional moving average model $X_t(\cdot) = \sum_{l=0}^{\infty} A_l\varepsilon_{t-l}(\cdot)$, where $A_l \in \mathbb{R}$ and $\{\varepsilon_t(\cdot)\}$ follows a white noise process. Denote $\Sigma_0^{\varepsilon}(u,v) = \mathrm{Cov}(\varepsilon_t(u), \varepsilon_t(v))$, then we can obtain that

$$f_{X,\theta}(u,v) = \frac{1}{2\pi}\Sigma_0^{\varepsilon}(u,v)\Big|\sum_{l=0}^{\infty}A_l e^{-i\theta l}\Big|^2$$

and

$$\Sigma_0(u,v) = \Big(\sum_{l=0}^{\infty}A_l^2\Big)\Sigma_0^{\varepsilon}(u,v).$$

By (3),

$$\mathcal{M}(f_X) \leqslant \frac{1}{2\pi}\frac{\sum_{l=0}^{\infty}|A_l|}{\sum_{l=0}^{\infty}A_l^2}.$$

Note $\sum_{l=0}^{\infty}|A_l| < \infty$ is a sufficient condition to guarantee the stationarity of $\{X_t(\cdot)\}$.

Example 2: Consider the functional linear process $X_t(u) = \sum_{l=0}^{\infty}\int_u A_l(u,v)\varepsilon_t(v)dv$, where $\{\varepsilon_t(\cdot)\}$ is a white noise process. Write $\Sigma_0^{\varepsilon}(u,v) = \sum_{k=1}^{\infty}\omega_k\psi_k(u)\psi_k(v)$, then

$$f_{X,\theta}(u,v) = \frac{1}{2\pi}\sum_{k=1}^{\infty}\omega_k\Big\{\sum_{l=0}^{\infty}a_{lk}(u)e^{-i\theta l}\Big\}\Big\{\sum_{l=0}^{\infty}a_{lk}(v)e^{i\theta l}\Big\}$$

and

$$\Sigma_0(u,v) = \sum_{k=1}^{\infty}\omega_k\Big\{\sum_{l=0}^{\infty}a_{lk}(u)a_{lk}(v)\Big\},$$

where $a_{lk}(u) = \int_u A_l(u,v)\psi_k(v)dv$. Suppose that $A_l(u,v)$ for each l is symmetric with respect to u,v and can be decomposed as $A_l(u,v) = \sum_{k=1}^{\infty} \beta_{lk}\psi_k(u)\psi_k(v)$ leading to $a_{lk}(u) = \beta_{lk}\psi_{lk}(u)$. By the fact that $(a+b)/(c+d) \leqslant \max\{a/c, b/d\}$ for $a,b,c,d > 0$, we can obtain that

$$\begin{aligned}\mathcal{M}(f_X) &\leqslant \sup_{\phi,\theta} \frac{\langle \phi, f_{X,\theta}(\phi)\rangle}{\langle \phi, \Sigma_0(\psi)\rangle} \\ &\leqslant \frac{1}{2\pi}\sup_{\phi,\theta} \frac{\sum_{k=1}^{\infty}\omega_k \left|\sum_{l=0}^{\infty}\beta_{lk}\langle\phi,\psi_k\rangle e^{-i\theta l}\right|^2}{\sum_{k=1}^{\infty}\omega_k\sum_{l=0}^{\infty}\beta_{lk}^2\langle\phi,\psi_k\rangle^2} \\ &\leqslant \frac{1}{2\pi}\sup_{\theta,k}\frac{\left|\sum_{l=0}^{\infty}\beta_{lk}e^{-i\theta l}\right|^2}{\sum_{l=0}^{\infty}\beta_{lk}^2} \leqslant \frac{1}{2\pi}\sup_{k}\frac{\left(\sum_{l=0}^{\infty}|\beta_{lk}|\right)^2}{\sum_{l=0}^{\infty}\beta_{lk}^2}\end{aligned}$$

where $\sum_{l=0}^{\infty}|\beta_{lk}| < \infty$ can be satisfied by a wide family of parameters.

Example 3: When $\mathbf{X}_t$ and $\mathbf{X}_s$ are independent for any $t \neq s$ and each $X_{jt}(\cdot)$ is infinite dimensional for $j = 1,\ldots,p$, then $\mathcal{M}(f_{\mathbf{X}}) = 1 < \infty$.

Example 4: It can be shown that for any $p > 1$, if $\{X_{t1}, t \in \mathbb{Z}\},\ldots,$ $\{X_{tp}, t \in \mathbb{Z}\}$ are independent and $\sup_{1\leqslant j\leqslant p}\mathcal{M}(f_{X_j}) < \infty$, then the functional stability measure of $\mathbf{X}_t = (X_{t1},\ldots,X_{tp})^{\mathrm{T}}$ is

$$\mathcal{M}(f_{\mathbf{X}}) \leqslant \sup_{1\leqslant j\leqslant p}\mathcal{M}(\mathcal{F}_{X_j}) < \infty.$$

Example 5: Consider a general case $\mathbf{Y}_t(u) = \mathbf{A}\mathbf{X}_t(u)$ with $\mathbf{A} \in \mathbb{R}^{p\times p}$ and $\mathcal{M}(f_{\mathbf{X}}) < \infty$. We can easily obtain that

$$\mathcal{M}(f_{\mathbf{Y}}) \leqslant \mathcal{M}(f_{\mathbf{X}}) < \infty,$$

which implies that linear transformation of $\mathbf{X}_t$ does not increase the functional stability measure. It is also worth noting that components of $\mathbf{Y}_t$ can be dependent in this example.

Example 6: Consider a more general scenario, $\mathbf{Y}_t(u) = \mathbf{A}\mathbf{X}_t(u) + \boldsymbol{\xi}_t(u)$, where $\mathbf{A} \in \mathbb{R}^{p\times r}$, $\mathbf{X}_t(u)$ is a r-dimensional vector of Gaussian processes and $\mathbf{X}_t(u)$, $\boldsymbol{\xi}_s(u)$ are independent for all t and s. When r is fixed, it implies that $\mathbf{Y}_t(u)$ can be expressed under a factor model structure. Note that $(a+b)/(c+d) \leqslant \max(a/c, b/d)$ for all positive real numbers a,b,c and d. Hence if $\max\{\mathcal{M}(f_{\mathbf{X}}), \mathcal{M}(f_{\boldsymbol{\xi}})\} < \infty$, then

$$\mathcal{M}(f_{\mathbf{Y}}) \leqslant \max\{\mathcal{M}(f_{\mathbf{X}}), \mathcal{M}(f_{\boldsymbol{\xi}})\} < \infty.$$

C. Proofs of main theorems

C.1. Proof of Theorem 1

(i) Define $\mathbf{Y} = \left(\langle\boldsymbol{\Phi}_1, \mathbf{X}_1\rangle_{\mathbb{H}},\ldots,\langle\boldsymbol{\Phi}_1,\mathbf{X}_n\rangle_{\mathbb{H}}\right)^{\mathrm{T}}$, then $\mathbf{Y} \sim N(\mathbf{0},\mathbf{Q})$, where $Q_{rs} = \left\langle\boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_{r-s}(\boldsymbol{\Phi}_1)\right\rangle_{\mathbb{H}}$ for $r,s = 1,\ldots,n$. Note $\left\langle\boldsymbol{\Phi}_1, \widehat{\boldsymbol{\Sigma}}_0(\boldsymbol{\Phi}_1)\right\rangle_{\mathbb{H}} = n^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Q}\mathbf{Z}$ with $\mathbf{Z} \sim N(\mathbf{0},\mathbf{I}_n)$ and $\left\langle\boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_0(\boldsymbol{\Phi}_1)\right\rangle_{\mathbb{H}} = \mathrm{E}\left(n^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{Q}\mathbf{Z}\right)$. By the Hanson-Wright inequality in Rudelson and Vershynin (2013),

$$P\left\{\left|\left\langle\boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0)(\boldsymbol{\Phi}_1)\right\rangle_{\mathbb{H}}\right| > \epsilon\right\} \leqslant 2\exp\left\{-c\min\left(\frac{n^2\epsilon^2}{\|\mathbf{Q}\|_F^2}, \frac{n\epsilon}{\|\mathbf{Q}\|}\right)\right\}$$

for some constant $c > 0$. By $\|\mathbf{Q}\|_F^2/n \leqslant \|\mathbf{Q}\|^2$ and letting $\epsilon = \eta\|\mathbf{Q}\|$, we obtain that

$$P\left\{\left|\left\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0)(\mathbf{\Phi}_1)\right\rangle_{\mathbb{H}}\right| > \eta\|\mathbf{Q}\|\right\} \leqslant 2\exp\left\{-cn\min\left(\eta^2, \eta\right)\right\} \tag{C.5}$$

for some universal constant $c > 0$.

Next we derive an upper bound on the operator norm $\|\mathbf{Q}\|$. Specifically, for any $\mathbf{w} = (w_1, \ldots, w_n)^{\mathrm{T}} \in \mathbb{R}^n$ with $||\mathbf{w}|| = 1$, define $G_{\mathbf{w}}(\theta) = \sum_{r=1}^n w_r \exp(-ir\theta)$ and its conjugate by $G_{\mathbf{w}}^*(\theta)$. Then we obtain that

$$\begin{aligned}
\mathbf{w}^{\mathrm{T}}\mathbf{Q}\mathbf{w} &= \sum_{r=1}^n \sum_{s=1}^n w_r w_s \left\langle \mathbf{\Phi}_1, \mathbf{\Sigma}_{r-s}(\mathbf{\Phi}_1)\right\rangle_{\mathbb{H}} \\
&= \sum_{r=1}^n \sum_{s=1}^n w_r w_s \int_{-\pi}^{\pi} \left\langle \mathbf{\Phi}_1, f_{\mathbf{X},\theta}(\mathbf{\Phi}_1)\right\rangle_{\mathbb{H}} \exp\{i(r-s)\theta\} d\theta \\
&= \int_{-\pi}^{\pi} \left\langle \mathbf{\Phi}_1, f_{\mathbf{X},\theta}(\mathbf{\Phi}_1)\right\rangle_{\mathbb{H}} G_{\mathbf{w}}(\theta) G_{\mathbf{w}}^*(\theta) d\theta,
\end{aligned}$$

where the second line follows from the inversion formula (2). For a fixed $\mathbf{\Phi} \in \mathbb{H}$, denote $\mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}) = 2\pi \cdot \operatorname{ess\,sup}_{\theta\in[-\pi,\pi]} \left|\langle \mathbf{\Phi}, f_{\mathbf{X},\theta}(\mathbf{\Phi})\rangle_{\mathbb{H}}\right|$. Since $\left\langle \mathbf{\Phi}_1, f_{\mathbf{X},\theta}(\mathbf{\Phi}_1)\right\rangle_{\mathbb{H}}$ is Hermitian and $\int_{-\pi}^{\pi} G_{\mathbf{w}}(\theta) G_{\mathbf{w}}^*(\theta) d\theta = 2\pi$, we have $\|\mathbf{Q}\| \leqslant \mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_1)$. Then it follows from (6) that

$$\|\mathbf{Q}\| \leqslant \mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_1) \leqslant \mathcal{M}_k(f_{\mathbf{X}}) \left\langle \mathbf{\Phi}_1, \mathbf{\Sigma}_0(\mathbf{\Phi}_1)\right\rangle_{\mathbb{H}}.$$

This result, together with (C.5) implies (8).

(ii) Note that

$$4\left\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0)(\mathbf{\Phi}_2)\right\rangle_{\mathbb{H}} \leqslant \left\langle \widetilde{\mathbf{\Phi}}_1, (\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0)(\widetilde{\mathbf{\Phi}}_1)\right\rangle_{\mathbb{H}} - \left\langle \widetilde{\mathbf{\Phi}}_2, (\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0)(\widetilde{\mathbf{\Phi}}_2)\right\rangle_{\mathbb{H}}.$$

where $\widetilde{\mathbf{\Phi}}_1 = \mathbf{\Phi}_1 + \mathbf{\Phi}_2$, $\widetilde{\mathbf{\Phi}}_2 = \mathbf{\Phi}_1 - \mathbf{\Phi}_2$ and $\mathcal{M}(f_{\mathbf{X}}, \widetilde{\mathbf{\Phi}}_i) \leqslant 2\{\mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_1) + \mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_2)\}$ for $i = 1, 2$. Combing these with results in (i) leads to

$$\begin{aligned}
&P\left[\left|\left\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0)(\mathbf{\Phi}_2)\right\rangle_{\mathbb{H}}\right| > \{\mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_1) + \mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_2)\}\eta\right] \\
\leqslant\; &\sum_{i=1}^2 P\left[\left|\left\langle \widetilde{\mathbf{\Phi}}_i, (\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0)(\widetilde{\mathbf{\Phi}}_i)\right\rangle_{\mathbb{H}}\right| > \mathcal{M}(f_{\mathbf{X}}, \widetilde{\mathbf{\Phi}}_i)\eta\right] \leqslant 4\exp\left\{-cn\min\left(\eta^2, \eta\right)\right\}
\end{aligned}$$

for some universal constant $c > 0$. This, together with, $\mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_i) \leqslant \mathcal{M}_k(f_{\mathbf{X}}) \left\langle \mathbf{\Phi}_i, \mathbf{\Sigma}_0(\mathbf{\Phi}_i)\right\rangle_{\mathbb{H}}$ for $i = 1, 2$, implies (9), which completes the proof. ▫

C.2. Proof of Theorem 2

First, we derive the concentration bound on $\|\widehat{\Sigma}_{jk}^{(0)} - \Sigma_{jk}^{(0)}\|_{\mathcal{S}}$ for each j and k. Let $\Delta_{jklm} = (\lambda_{jl}\lambda_{km})^{-1/2}\left\langle \phi_{jl}, (\widehat{\Sigma}_{jk}^{(0)} - \Sigma_{jk}^{(0)})(\phi_{km})\right\rangle$ for $j, k = 1, \ldots, p$, and $l, m = 1, \ldots, \infty$. Then we have that $\|\widehat{\Sigma}_{jk}^{(0)} - \Sigma_{jk}^{(0)}\|_{\mathcal{S}}^2 = \sum_{l,m=1}^{\infty} \lambda_{jl}\lambda_{km}\Delta_{jklm}^2$. By Jensen's inequality, we have that

$$\mathrm{E}\Big\{\|\widehat{\Sigma}_{jk}^{(0)} - \Sigma_{jk}^{(0)}\|_{\mathcal{S}}^{2q}\Big\} \leqslant \Big(\sum_{l,m=1}^{\infty} \lambda_{jl}\lambda_{km}\Big)^{q-1} \sum_{l,m=1}^{\infty} \lambda_{jl}\lambda_{km}\mathrm{E}\big|\Delta_{jklm}\big|^{2q} \leqslant \lambda_0^{2q} \sup_{l,m} \mathrm{E}\big|\Delta_{jklm}\big|^{2q}. \tag{C.6}$$

For any given (j,k,l,m), let

$$\mathbf{\Phi}_1 = (0,\ldots,0,\lambda_{jl}^{-1/2}\phi_{jl},0,\ldots,0)^{\mathrm{T}} \text{ and } \mathbf{\Phi}_2 = (0,\ldots,0,\lambda_{km}^{-1/2}\phi_{km},0,\ldots,0)^{\mathrm{T}}.$$

By the definition of Δ_{jklm} and orthonormality of $\{\phi_{jl}(\cdot)\}$ and $\{\phi_{km}(\cdot)\}$ for each $j,k = 1,\ldots,p$, we have $\Delta_{jklm} = \langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0)(\mathbf{\Phi}_2)\rangle_{\mathbb{H}}$, $\langle \mathbf{\Phi}_1, \mathbf{\Sigma}_0(\mathbf{\Phi}_1)\rangle_{\mathbb{H}} = \langle \mathbf{\Phi}_2, \mathbf{\Sigma}_0(\mathbf{\Phi}_2)\rangle_{\mathbb{H}} = 1$. Applying (9) in Theorem 1, we can obtain that

$$P\Big\{\big|\Delta_{jklm}\big| > 2\mathcal{M}_1(f_{\mathbf{X}})\eta\Big\} \leqslant 4\exp\Big\{-cn\min(\eta^2,\eta)\Big\}, \tag{C.7}$$

for $j,k = 1,\ldots,p$, $l = 1,\ldots,d_j$ and $m = 1,\ldots,d_k$. It then follows from Lemma 6 in Section D of the Supplementary Material that for each integer $q \geqslant 1$,

$$\{2\mathcal{M}_1(f_{\mathbf{X}})\}^{-2q}\,\mathrm{E}\big|\Delta_{jklm}\big|^{2q} \leqslant q!4(4c^{-1}n^{-1})^q + 4(2q)!(4c^{-1}n^{-1})^{2q}.$$

This together with (C.6) implies that

$$(2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0)^{-2q}\,\mathrm{E}\Big\{\|\widehat{\Sigma}_{jk}^{(0)} - \Sigma_{jk}^{(0)}\|_{\mathcal{S}}^{2q}\Big\} \leqslant q!4(4c^{-1}n^{-1})^q + (2q)!4(4c^{-1}n^{-1})^{2q}. \tag{C.8}$$

Finally, it follows from Lemma 6 that there exists some universal constant $\tilde{c} > 0$ such that

$$P\Big\{\|\widehat{\Sigma}_{jk}^{(0)} - \Sigma_{jk}^{(0)}\|_{\mathcal{S}} \geqslant 2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\eta\Big\} \leqslant 4\exp\Big\{-\tilde{c}n\min(\eta^2,\eta)\Big\}.$$

Using the definition of $\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\|_{\max} = \max_{1\leqslant j,k\leqslant p}\|\widehat{\Sigma}_{jk}^{(0)} - \Sigma_{jk}^{(0)}\|_{\mathcal{S}}$ and applying the union bound of probability, we obtain that

$$P\Big\{\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\|_{\max} \geqslant 2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\eta\Big\} \leqslant 4p^2\cdot\exp\Big\{-\tilde{c}n\min(\eta^2,\eta)\Big\}.$$

Let $\eta = \rho\sqrt{\log p/n} \leqslant 1$ and $\rho^2\tilde{c} > 2$, which can be achieved for sufficiently large n. We obtain that

$$P\left\{\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\|_{\max} \geqslant 2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\rho\sqrt{\frac{\log p}{n}}\right\} \leqslant 4p^{2-\tilde{c}\rho^2}.$$

The proof is complete. ▫

C.3. Proof of Theorem 3

To simplify our notation, for each $j,k=1,\dots,p$, we will denote $\Sigma_{jk}^{(0)}$ and $\widehat{\Sigma}_{jk}^{(0)}$ by Σ_{jk} and $\widehat{\Sigma}_{jk}$, respectively, in our subsequent proofs. Let $\delta_{jl}=\min_{1\leqslant k\leqslant l}\{\lambda_{jk}-\lambda_{j(k+1)}\}$ and $\widehat{\Delta}_{jk}=\widehat{\Sigma}_{jk}-\Sigma_{jk}$ for $j,k=1,\dots,p$ and $l=1,2\dots$. It follows from (4.43) and Lemma 4.3 of Bosq (2000) that

$$\sup_{l\geqslant 1}|\widehat{\lambda}_{jl}-\lambda_{jl}|\leqslant\|\widehat{\Delta}_{jj}\|_{\mathcal{S}}\ \text{ and }\ \sup_{l\geqslant 1}\delta_{jl}\|\widehat{\phi}_{jl}-\phi_{jl}\|\leqslant 2\sqrt{2}\|\widehat{\Delta}_{jj}\|_{\mathcal{S}}. \tag{C.9}$$

Moreover, we can express $\widehat{\lambda}_{jl}-\lambda_{jl}$ and $\widehat{\phi}_{jl}-\phi_{jl}$, as stated in Lemma 7 in Section D of the Supplementary Material. The proof of Theorem 3 relies on the concentration inequalities for eigenvalues and eigenvectors as stated in the following Lemmas 1 and 2, whose proofs are provided in Section D.

Lemma 1. *Suppose that Conditions 1–3 hold. Then there exists some universal constant $\tilde{c}_1>0$ such that for each $j=1,\dots,p,l=1,\dots,\infty$, and any $\eta>0$,*

$$P\left\{\left|\frac{\widehat{\lambda}_{jl}-\lambda_{jl}}{\lambda_{jl}}\right|>\mathcal{M}_1(f_{\mathbf{X}})\eta+\rho_1 l^{2\alpha+1}\mathcal{M}_1^2(f_{\mathbf{X}})\eta^2\right\}\leqslant 4\exp\Big\{-\tilde{c}_1 n\min(\eta^2,\eta)\Big\}, \tag{C.10}$$

where $\rho_1=16\sqrt{2}c_0^{-2}\alpha\lambda_0^2$.

Lemma 2. *Suppose that Conditions 1–3 hold. Then there exists some universal constant $\tilde{c}>0$ such that for each $j=1,\dots,p,l=1,\dots,\infty$, and any $\eta>0$,*

$$P\left\{\left\|\widehat{\phi}_{jl}-\phi_{jl}\right\|>4\sqrt{2}\mathcal{M}_1(f_{\mathbf{X}})\lambda_0 c_0^{-1}l^{\alpha+1}\eta\right\}\leqslant 4\exp\Big\{-\tilde{c}n\min(\eta^2,\eta)\Big\}. \tag{C.11}$$

Proof of Theorem 3. Applying the union bound of probability in (C.10), we obtain that

$$P\left\{\max_{1\leqslant j\leqslant p,1\leqslant l\leqslant M}\left|\frac{\widehat{\lambda}_{jl}-\lambda_{jl}}{\lambda_{jl}}\right|>\mathcal{M}_1(f_{\mathbf{X}})\eta+\rho_1 l^{2\alpha+1}\mathcal{M}_1^2(f_{\mathbf{X}})\eta^2\right\}\leqslant 4pM\exp\Big\{-\tilde{c}_1 n\min(\eta^2,\eta)\Big\}.$$

Let $\eta=\tilde{\rho}_2\sqrt{\log(pM)/n}\leqslant 1$ and $1+\rho_1 M^{2\alpha+1}\mathcal{M}_1(f_{\mathbf{X}})\eta\leqslant\tilde{\rho}_1$, which can be achieved for sufficiently large $n\gtrsim M^{4\alpha+2}\mathcal{M}_1^2(f_{\mathbf{X}})\log(pM)$. We obtain that

$$P\left\{\max_{1\leqslant j\leqslant p,1\leqslant l\leqslant M}\left|\frac{\widehat{\lambda}_{jl}-\lambda_{jl}}{\lambda_{jl}}\right|>\tilde{\rho}_1\tilde{\rho}_2\mathcal{M}_1(f_{\mathbf{X}})\sqrt{\frac{\log(pM)}{n}}\right\}\leqslant 4(pM)^{1-\tilde{c}_1\tilde{\rho}_2^2}. \tag{C.12}$$

Finally, letting $\eta=\tilde{\rho}_3\sqrt{\frac{\log(pM)}{n}}<1$ and following the same developments, we obtain that

$$P\left\{\max_{1\leqslant j\leqslant p,1\leqslant l\leqslant M}\left(\frac{\|\widehat{\phi}_{jl}-\phi_{jl}\|}{l^{\alpha+1}}\right)>4\sqrt{2}\lambda_0 c_0^{-1}\tilde{\rho}_3\mathcal{M}_1(f_{\mathbf{X}})\sqrt{\frac{\log(pM)}{n}}\right\}\leqslant 4(pM)^{1-\tilde{c}\tilde{\rho}_3^2}. \tag{C.13}$$

It follows from (C.12) and (C.13) that, for sufficiently large $n \gtrsim M^{4\alpha+2}\mathcal{M}_1^2(f_{\mathbf{X}})\log(pM)$ and suitable choices of constants $c_1, c_2 > 0$, (17) holds. $\square$

C.4. Proof of Theorem 4

The proof of Theorem 4 is based on the following Lemma 3.

Lemma 3. *Suppose that Conditions 1–3 hold. Then there exist some positive constants* $\rho_4, \rho_5, \tilde{c}_3$ *and* $\tilde{c}_4$ *such that*

(i) for each $j = 1, \dots, p$, $l = 1, \dots, \infty$, *and any* $\eta > 0$,

$$P\left\{\left|\frac{\widehat{\sigma}_{jjll}^{(0)} - \sigma_{jjll}^{(0)}}{\lambda_{jl}}\right| > \mathcal{M}_1(f_{\mathbf{X}})\eta + \rho_1\mathcal{M}_1^2(f_{\mathbf{X}})l^{2\alpha+1}\eta^2\right\} \leqslant 4\exp\Big\{-\tilde{c}_3 n\min(\eta^2,\eta)\Big\}; \tag{C.14}$$

(ii) for each $j, k = 1, \dots, p$, $l, m = 1, \dots, \infty$, *but* $j \neq k$ *or* $l \neq m$, *a fixed* h *and any* $\eta > 0$,

$$\begin{aligned} &P\left\{\left|\frac{\widehat{\sigma}_{jklm}^{(h)} - \sigma_{jklm}^{(h)}}{\lambda_{jl}^{1/2}\lambda_{km}^{1/2}}\right| \geqslant \rho_4\mathcal{M}_1(f_{\mathbf{X}})(l\vee m)^{\alpha+1}\eta + \rho_5\mathcal{M}_1^2(f_{\mathbf{X}})(l\vee m)^{3\alpha+2}\eta^2\right\} \\ &\leqslant \tilde{c}_4\exp\big\{-\tilde{c}_3 n\min(\eta^2,\eta)\big\} + \tilde{c}_4\exp\big\{-\tilde{c}_3\mathcal{M}_1^{-2}(f_{\mathbf{X}})n(l\vee m)^{-2(\alpha+1)}\big\}. \end{aligned} \tag{C.15}$$

Proof of Lemma 3. For the special case of (j,k,l,m) with $j = k$, provided that $\widehat{\sigma}_{jjlm}^{(0)} = \widehat{\lambda}_{jl}I(l=m)$ and $\sigma_{jjlm}^{(0)} = \lambda_{jl}I(l=m)$, (C.14) follows directly from Lemma 1.

For general cases of (j,k,l,m) with $j \neq k$, $\widehat{\sigma}_{jklm}^{(h)} = (n-h)^{-1}\sum_{t=1}^{n-h}\widehat{\xi}_{tjl}\widehat{\xi}_{(t+h)km}$ and $\sigma_{jklm}^{(h)} = E(\xi_{tjk}\xi_{(t+h)lm})$. Let $\widehat{r}_{jl} = \widehat{\phi}_{jl} - \phi_{jl}$, then $\widehat{\sigma}_{h,jklm} - \sigma_{h,jklm}$ can be decomposed as

$$\begin{aligned} \widehat{\sigma}_{jklm}^{(h)} - \sigma_{jklm}^{(h)} &= \big\langle \widehat{r}_{jl}, \widehat{\Sigma}_{jk}^{(h)}(\widehat{r}_{km})\big\rangle + \Big(\big\langle \widehat{r}_{jl}, \widehat{\Delta}_{jk}^{(h)}(\phi_{km})\big\rangle + \big\langle \phi_{jl}, \widehat{\Delta}_{jk}^{(h)}(\widehat{r}_{km})\big\rangle\Big) \\ &\quad + \Big(\big\langle \widehat{r}_{jl}, \Sigma_{jk}^{(h)}(\phi_{km})\big\rangle + \big\langle \phi_{jl}, \Sigma_{jk}^{(h)}(\widehat{r}_{km})\big\rangle\Big) + \big\langle \phi_{jl}, \widehat{\Delta}_{jk}^{(h)}(\phi_{km})\big\rangle \\ &= I_1 + I_2 + I_3 + I_4. \end{aligned}$$

For a fixed $h > 0$, let $\Omega_{jk}^{(h)} = \Big\{\|\widehat{\Delta}_{jk}^{(h)}\|_{\mathcal{S}} \leqslant \lambda_0\Big\}$ and $\widetilde{\Omega}_{jk}^{(h)} = \Big\{\|\widehat{\Delta}_{jk}^{(h)}\|_{\mathcal{S}} \leqslant 4\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\eta\Big\}$. It follows from the same developments as in the proof of (10) in Theorem 2 and Proposition 1 that there exists some universal constant $\tilde{c} > 0$ such that for any $\eta > 0$, a fixed $h \neq 0$ and each $j, k = 1, \dots, p$,

$$P\Big\{\big\|\widehat{\Sigma}_{jk}^{(h)} - \Sigma_{jk}^{(h)}\big\|_{\mathcal{S}} > 4\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\eta\Big\} \leqslant 8\exp\Big\{-\tilde{c}n\min(\eta^2,\eta)\Big\}. \tag{C.16}$$

On the event $\Omega^{(h)}_{jk} \cap \widetilde{\Omega}^{(0)}_{jj} \cap \widetilde{\Omega}^{(0)}_{kk} \cap \widetilde{\Omega}^{(h)}_{jk}$, it follows from Condition 3 with $\lambda_{jl} \geqslant c_0 \alpha^{-1} l^{-\alpha}$, (C.9), Lemma 8 in Section D that

$$\begin{aligned} \left| \frac{I_1}{\lambda_{jl}^{1/2} \lambda_{km}^{1/2}} \right| &\lesssim (lm)^{\alpha/2} \|\widehat{r}_{jl}\| \big(\|\widehat{\Delta}^{(h)}_{jk}\| + \|\Sigma^{(h)}_{jk}\|_{\mathcal{S}} \big) \|\widehat{r}_{km}\| \\ &\lesssim \mathcal{M}_1^2(f_{\mathbf{X}})(l \vee m)^{3\alpha+2} \eta^2, \end{aligned} \tag{C.17}$$

$$\begin{aligned} \left| \frac{I_2}{\lambda_{jl}^{1/2} \lambda_{km}^{1/2}} \right| &\lesssim (lm)^{\alpha/2} \|\widehat{\Delta}^{(h)}_{jk}\|_{\mathcal{S}} \Big(l^{\alpha+1} \|\widehat{\Delta}^{(0)}_{jj}\|_{\mathcal{S}} + m^{\alpha+1} \|\widehat{\Delta}^{(0)}_{kk}\|_{\mathcal{S}} \Big) \\ &\lesssim \mathcal{M}_1^2(f_{\mathbf{X}})(l \vee m)^{2\alpha+1} \eta^2. \end{aligned} \tag{C.18}$$

For the term I_4, it follows from (16) in Proposition 1 and the fact $\lambda_{jl} + \lambda_{jm} \geqslant 2\lambda_{jl}^{1/2} \lambda_{jm}^{1/2}$ that

$$P \left\{ \left| \frac{I_4}{\lambda_{jl}^{1/2} \lambda_{km}^{1/2}} \right| \geqslant 4 \mathcal{M}_1(f_{\mathbf{X}}) \lambda_0 \eta \right\} \leqslant 8 \exp \Big\{ - cn \min(\eta^2, \eta) \Big\}. \tag{C.19}$$

Finally, we consider the term I_3. By Lemma 8, we have that $\|\Sigma^{(h)}_{jk}(\phi_{km})\| \leqslant \lambda_{km}^{1/2} \lambda_0^{1/2}$ and $\|\Sigma^{(h)}_{jk}(\phi_{jl})\| \leqslant \lambda_{jl}^{1/2} \lambda_0^{1/2}$. These results together with (C.39) in Lemma 4 in Section D and Condition 3 with $\lambda_{jl} \geqslant c_0 \alpha^{-1} l^{-\alpha}$ imply that

$$\left| \frac{I_3}{\lambda_{jl}^{1/2} \lambda_{km}^{1/2}} \right| \lesssim \mathcal{M}_1(f_{\mathbf{X}})(l \vee m)^{\alpha+1} \eta + \mathcal{M}_1^2(f_{\mathbf{X}})(l \vee m)^{(5\alpha+4)/2} \eta^2 \tag{C.20}$$

holds with probability greater than $1 - 16 \exp\big\{-\tilde{c}_2 n \min(\eta^2, \eta)\big\} - 8 \exp\big\{-\tilde{c}_2 \mathcal{M}_1^{-2}(f_{\mathbf{X}}) n (l \vee m)^{-2(\alpha+1)}\big\}$, with some positive constant $\tilde{c}_2$.

Combining (C.16)–(C.20) and by Theorem 2, we obtain that there exist four positive constants ρ_4, ρ_5, $\tilde{c}_3$, $\tilde{c}_4$ such that (C.15) in Lemma 3 holds. For the case of $h = 0$, we follow the same developments as above by applying Theorems 1–2 and hence (C.15) follows with the different choice of relevant positive constants. The proof of Lemma 3 is complete. ▫

Proof of Theorem 4 . Let $\eta = \tilde{\rho}_4 \sqrt{\log(pM)/n} < 1$ and $\rho_4 + \rho_5 M^{2\alpha+1} \mathcal{M}_1(f_{\mathbf{X}}) \eta \leqslant \tilde{\rho}_5$, which can be achieved for sufficiently large $n \gtrsim M^{4\alpha+2} \mathcal{M}_1^2(f_{\mathbf{X}}) \log(pM)$. Following the similar techniques as used in the proof of (17) in Theorem 3, we can obtain (18), which completes the proof. ▫

C.5. Proof of Theorem 5

Since $\widehat{\mathbf{B}}_j \in \mathbb{R}^{pq \times q}$ is the minimizer of (23), we have

$$-\langle\langle \widehat{\mathbf{Y}}_j, \widehat{\mathbf{B}}_j \rangle\rangle + \frac{1}{2} \langle\langle \widehat{\mathbf{B}}_j, \widehat{\mathbf{\Gamma}} \widehat{\mathbf{B}}_j \rangle\rangle + \gamma_{nj} \|\widehat{\mathbf{B}}_j\|_1^{(q)} \leqslant -\langle\langle \widehat{\mathbf{Y}}_j, \mathbf{B}_j \rangle\rangle + \frac{1}{2} \langle\langle \mathbf{B}_j, \widehat{\mathbf{\Gamma}} \widehat{\mathbf{B}}_j \rangle\rangle + \gamma_{nj} \|\mathbf{B}_j\|_1^{(q)}.$$

Letting $\boldsymbol{\Delta}_j = \widehat{\mathbf{B}}_j - \mathbf{B}_j$ and S_j^c be the complement of S_j in the set $\{1, \ldots, p\}$, we have

$$\begin{aligned}
\frac{1}{2}\langle\langle \boldsymbol{\Delta}_j, \widehat{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_j \rangle\rangle &\leqslant \langle\langle \boldsymbol{\Delta}_j, \widehat{\mathbf{Y}}_j - \widehat{\boldsymbol{\Gamma}}\mathbf{B}_j \rangle\rangle + \gamma_{nj}\Big(\|\mathbf{B}_j\|_1^{(q)} - \|\mathbf{B}_{1j} + \boldsymbol{\Delta}_j\|_1^{(q)}\Big) \\
&\leqslant \langle\langle \boldsymbol{\Delta}_j, \widehat{\mathbf{Y}}_j - \widehat{\boldsymbol{\Gamma}}\mathbf{B}_j \rangle\rangle + \gamma_{nj}\Big(\|\mathbf{B}_{jS_j}\|_1^{(q)} - \|\mathbf{B}_{jS_j} + \boldsymbol{\Delta}_{jS_j}\|_1^{(q)} - \|\boldsymbol{\Delta}_{jS_j^c}\|_1^{(q)}\Big) \\
&\leqslant \langle\langle \boldsymbol{\Delta}_j, \widehat{\mathbf{Y}}_j - \widehat{\boldsymbol{\Gamma}}\mathbf{B}_j \rangle\rangle + \gamma_{nj}\Big(\|\boldsymbol{\Delta}_{jS_j}\|_1^{(q)} - \|\boldsymbol{\Delta}_{jS_j^c}\|_1^{(q)}\Big)
\end{aligned}$$

By Lemma 10 in Section D, Condition 7 and the choice of γ_{nj}, we have

$$\big|\langle\langle \boldsymbol{\Delta}_j, \widehat{\mathbf{Y}}_j - \widehat{\boldsymbol{\Gamma}}\mathbf{B}_j \rangle\rangle\big| \leqslant \|\widehat{\mathbf{Y}}_j - \widehat{\boldsymbol{\Gamma}}\mathbf{B}_j\|_{\max}^{(q)}\|\boldsymbol{\Delta}_j\|_1^{(q)} \leqslant \frac{\gamma_{nj}}{2}\big(\|\boldsymbol{\Delta}_{jS_j}\|_1^{(q)} + \|\boldsymbol{\Delta}_{jS_j^c}\|_1^{(q)}\big).$$

Combing the above two results, we have

$$0 \leqslant \frac{1}{2}\langle\langle \boldsymbol{\Delta}_j, \widehat{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_j \rangle\rangle \leqslant \frac{3\gamma_{nj}}{2}\|\boldsymbol{\Delta}_{jS_j}\|_1^{(q)} - \frac{\gamma_{nj}}{2}\|\boldsymbol{\Delta}_{jS_j^c}\|_1^{(q)},$$

which implies $\|\boldsymbol{\Delta}_{jS_j^c}\|_1^{(q)} \leqslant 3\|\boldsymbol{\Delta}_{jS_j}\|_1^{(q)}$ and therefore $\|\boldsymbol{\Delta}_j\|_1^{(q)} \leqslant 4\|\boldsymbol{\Delta}_{jS_j}\|_1^{(q)} \leqslant 4\sqrt{s_j}\|\boldsymbol{\Delta}_j\|_F$. This result together with Condition 5 and $\tau_2 \geqslant 32\tau_1 q^2 s_j$ implies that

$$\langle\langle \boldsymbol{\Delta}_j, \widehat{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_j \rangle\rangle \geqslant \tau_2\|\boldsymbol{\Delta}_j\|_F^2 - \tau_1 q^2\big\{\|\boldsymbol{\Delta}_j\|_1^{(q)}\big\}^2 \geqslant \big(\tau_2 - 16\tau_1 q^2 s_j\big)\|\boldsymbol{\Delta}_j\|_F^2 \geqslant \frac{\tau_2}{2}\|\boldsymbol{\Delta}_j\|_F^2. \quad (C.21)$$

Therefore,

$$\frac{\tau_2}{4}\|\boldsymbol{\Delta}_j\|_F^2 \leqslant \frac{3}{2}\gamma_{nj}\|\boldsymbol{\Delta}_j\|_1^{(q)} \leqslant 6\gamma_{nj}s_j^{1/2}\|\boldsymbol{\Delta}_j\|_F,$$

which implies that

$$\|\boldsymbol{\Delta}_j\|_F \leqslant \frac{24 s_j^{1/2}\gamma_{nj}}{\tau_2} \text{ and } \|\boldsymbol{\Delta}_j\|_1^{(q)} \leqslant \frac{96 s_j \gamma_{nj}}{\tau_2}, \quad (C.22)$$

as is claimed in Theorem 5.

Next we prove the upper bound on $\widehat{\mathbf{A}} - \mathbf{A}$.

For $k \in S_j$, it follows from $\boldsymbol{\Psi}_{jk} = \int\int \boldsymbol{\phi}_k(v)A_{jk}(u,v)\boldsymbol{\psi}_j(u)^{\mathrm{T}}dudv$, Condition 4 with $A_{jk}(u,v) = \boldsymbol{\phi}_k(v)^{\mathrm{T}}\mathbf{a}_{jk}\boldsymbol{\phi}_j(u) + (\sum_{l,m=1}^{\infty} - \sum_{l,m=1}^{q})a_{jklm}\phi_{jl}(u)\phi_{km}(v)$ and orthonormality of $\{\phi_{jl}(\cdot)\}_{l\geqslant 1}$ and $\{\phi_{km}(\cdot)\}_{m\geqslant 1}$ that $\|\boldsymbol{\Psi}_{jk}\|_F = ||\mathbf{a}_{jk}||_F = \big\{\sum_{l,m=1}^{q}\mu_{jk}^2(l+m)^{-2\beta-1}\big\}^{1/2} \leqslant \big\{\mu_{jk}^2\int_1^q\int_1^q(x+y)^{-2\beta-1}dxdy\big\}^{1/2} = O(\mu_{jk})$. For $k \in S_j^c$, we have $\boldsymbol{\Psi}_{jk} = \mathbf{0}$. Hence

$$\|\boldsymbol{\Psi}_j\|_1^{(q)} = \sum_{k=1}^{p}||\boldsymbol{\Psi}_{jk}||_F = O\Big(\sum_{k\in S_j}\mu_{jk}\Big) = O(s_j). \quad (C.23)$$

Observe that $\widehat{\boldsymbol{\Psi}}_j - \boldsymbol{\Psi}_j = \widehat{\mathbf{D}}^{-1}\widehat{\mathbf{B}}_j - \mathbf{D}^{-1}\mathbf{B}_j = (\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1})\mathbf{B}_j + \mathbf{D}^{-1}(\widehat{\mathbf{B}}_j - \mathbf{B}_j) + (\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1})(\widehat{\mathbf{B}}_j - \mathbf{B}_j)$. It follows from the diagonal structure of $\widehat{\mathbf{D}}^{-1}$ and $\mathbf{D}^{-1}$ that

$$\begin{aligned}
\|\widehat{\boldsymbol{\Psi}}_j - \boldsymbol{\Psi}_j\|_1^{(q)} \leqslant &\|(\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1})\|_{\max}\|\mathbf{B}_j\|_1^{(q)} + \|\mathbf{D}^{-1}\|_{\max}\|\widehat{\mathbf{B}}_j - \mathbf{B}_j\|_1^{(q)} \\
&+ \|(\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1})\|_{\max}\|\widehat{\mathbf{B}}_j - \mathbf{B}_j\|_1^{(q)}.
\end{aligned} \quad (C.24)$$

By Conditions 3, 6 and the fact $\widehat{\mathbf{D}}_k = \operatorname{diag}\big(\widehat{\lambda}_{k1}^{1/2}, \ldots, \widehat{\lambda}_{kq}^{1/2}\big)$, $\mathbf{D}_k = \operatorname{diag}\big(\lambda_{k1}^{1/2}, \ldots, \lambda_{kq}^{1/2}\big)$, we have $\|(\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1})\|_{\max} \leqslant \alpha^{1/2} c_0^{-1/2} q^{\alpha/2} C_\lambda \mathcal{M}(f_{\mathbf{X}}) \sqrt{\frac{\log(pq)}{n}}$ and $\|\mathbf{D}^{-1}\|_{\max} \leqslant \alpha^{1/2} c_0^{-1/2} q^{\alpha/2}$. By Condition 1 and (C.23), we have $||\mathbf{B}||_1^{(q)} \leqslant ||\mathbf{D}||_{\max}^{(q)} \|\mathbf{\Psi}_j\|_1^{(q)} = O(\lambda_0^{1/2} s_j)$. These results together with (C.22) implies that

$$\|\widehat{\mathbf{\Psi}}_j - \mathbf{\Psi}_j\|_1^{(q)} \leqslant \frac{96 \alpha^{1/2} q^{\alpha/2} s_j \gamma_{nj}}{c_0^{1/2} \tau_2} \Big\{1 + o(1)\Big\}, \tag{C.25}$$

where the constant comes from the second term in (C.24), since the first and third terms are of smaller orders relative to the second term.

For each $j, k = 1, \ldots, p$, note that

$$\begin{aligned} \widehat{\mathrm{A}}_{jk}(u,v) - \mathrm{A}_{jk}(u,v) &= \widehat{\boldsymbol{\phi}}_k(v)^{\mathrm{T}} \widehat{\mathbf{\Psi}}_{jk} \widehat{\boldsymbol{\phi}}_j(u) - \boldsymbol{\phi}_k(v)^{\mathrm{T}} \mathbf{\Psi}_{jk} \boldsymbol{\phi}_j(u) + \mathrm{R}_{jk}(u,v) \\ &= \widehat{\boldsymbol{\phi}}_k(v)^{\mathrm{T}} \widehat{\mathbf{\Psi}}_{jk} \Big\{\widehat{\boldsymbol{\phi}}_j(u) - \boldsymbol{\phi}_j(u)\Big\} + \Big\{\widehat{\boldsymbol{\phi}}_k(v) - \boldsymbol{\phi}_k(v)\Big\}^{\mathrm{T}} \widehat{\mathbf{\Psi}}_{jk} \boldsymbol{\phi}_j(u) \\ &\quad + \boldsymbol{\phi}_k(v)^{\mathrm{T}} (\widehat{\mathbf{\Psi}}_{jk} - \mathbf{\Psi}_{jk}) \boldsymbol{\phi}_j(u) + \mathrm{R}_{jk}(u,v), \end{aligned}$$

We bound the first three terms. By Lemma 9 in Section D, we have

$$\begin{aligned} &\Big\|\widehat{\boldsymbol{\phi}}_k(v)^{\mathrm{T}} \widehat{\mathbf{\Psi}}_{jk} \Big\{\widehat{\boldsymbol{\phi}}_j(u) - \boldsymbol{\phi}_j(u)\Big\}\Big\|_{\mathcal{S}} \leqslant q^{1/2} \max_{1 \leqslant l \leqslant q} \|\widehat{\phi}_{jl} - \phi_{jl}\| \|\widehat{\mathbf{\Psi}}_{jk}\|_F, \\ &\Big\|\Big\{\widehat{\boldsymbol{\phi}}_k(v) - \boldsymbol{\phi}_k(v)\Big\}^{\mathrm{T}} \widehat{\mathbf{\Psi}}_{jk} \boldsymbol{\phi}_j(u)\Big\|_{\mathcal{S}} \leqslant q^{1/2} \max_{1 \leqslant m \leqslant q} \|\widehat{\phi}_{km} - \phi_{km}\| \|\widehat{\mathbf{\Psi}}_{jk}\|_F, \\ &\Big\|\boldsymbol{\phi}_k(v)^{\mathrm{T}} (\widehat{\mathbf{\Psi}}_{jk} - \mathbf{\Psi}_{jk}) \boldsymbol{\phi}_j(u)\Big\|_{\mathcal{S}} = \|\widehat{\mathbf{\Psi}}_{jk} - \mathbf{\Psi}_{jk}\|_F. \end{aligned} \tag{C.26}$$

We then bound the fourth term. By $R_{jk}(u,v) = (\sum_{l,m=1}^{q} - \sum_{l,m=1}^{\infty}) a_{jklm} \phi_{jl}(u) \phi_{km}(v)$, we have

$$\begin{aligned} ||R_{jk}||_{\mathcal{S}}^2 &= O(1) \Big\| \sum_{l=q+1}^{\infty} \sum_{m=1}^{\infty} a_{jklm} \phi_{jl}(u) \phi_{km}(v) \Big\|_{\mathcal{S}}^2 \\ &= O(1) \sum_{l=q+1}^{\infty} \sum_{m=1}^{\infty} a_{jklm}^2 \leqslant O(1) \mu_{jk}^2 \sum_{l=q+1}^{\infty} \sum_{m=1}^{\infty} (l+m)^{-2\beta-1} = O(\mu_{jk}^2 q^{-2\beta+1}). \end{aligned}$$

This together with Condition 4 implies that

$$\max_{1 \leqslant j \leqslant p} \sum_{k=1}^{p} ||R_{jk}||_{\mathcal{S}} \leqslant O\big(q^{-\beta+1/2} \max_{1 \leqslant j \leqslant p} \sum_{k \in S_j} \mu_{jk}\big) = O\big(s q^{-\beta+1/2}\big). \tag{C.27}$$

It follows from (C.23), (C.25), (C.26), (C.27) and the fact $\|\widehat{\mathbf{\Psi}}_j\|_1^{(q)} \leqslant \|\widehat{\mathbf{\Psi}}_j - \mathbf{\Psi}_j\|_1^{(q)} + \|\mathbf{\Psi}_j\|_1^{(q)} = O(s_j)$ that

$$\begin{aligned} \|\widehat{\mathbf{A}} - \mathbf{A}\|_\infty &\leqslant 2q^{1/2} \max_{\substack{1 \leqslant j \leqslant p \\ 1 \leqslant l \leqslant q}} \|\widehat{\phi}_{jl} - \phi_{jl}\| \max_{1 \leqslant j \leqslant p} \|\widehat{\mathbf{\Psi}}_j\|_1^{(q)} + \max_{1 \leqslant j \leqslant p} \|\widehat{\mathbf{\Psi}}_j - \mathbf{\Psi}_j\|_1^{(q)} + ||\mathbf{R}||_\infty. \\ &\leqslant \frac{96 \alpha^{1/2} q^{\alpha/2} s \gamma_n}{c_0^{1/2} \tau_2} \Big\{1 + o(1)\Big\}, \end{aligned}$$

where the constant comes from $\max_j \|\widehat{\mathbf{\Psi}}_j - \mathbf{\Psi}_j\|_1^{(q)}$, since other terms are of smaller orders of this term. The proof is complete. ▫

D. Additional technical proofs

D.1. Proof of Proposition 1

Let $\mathbf{Y}_{1,t} = \mathbf{X}_t + \mathbf{X}_{t+h}$, $\mathbf{\Sigma}_{\mathbf{Y}_1,\ell}(u,v) = \mathrm{Cov}\{\mathbf{Y}_{1,t}(u), \mathbf{Y}_{1,(t+\ell)}(v)\}, \ell \in \mathbb{Z}, (u,v) \in \mathcal{U}^2$. Define the spectral density operator of $\mathbf{Y}_{1,t}$ by

$$f_{\mathbf{Y}_1,\theta} = \frac{1}{2\pi}\sum_{\ell=-\infty}^{\infty} \mathbf{\Sigma}_{\mathbf{Y}_1,\ell}\exp(-i\ell\theta),\ \theta \in [-\pi,\pi].$$

Then we can obtain that $f_{\mathbf{Y}_1,\theta} = \{2 + \exp(-ih\theta) + \exp(ih\theta)\} f_{\mathbf{X},\theta}$. Similarly, by letting $\mathbf{Y}_{2,t}(u) = \mathbf{X}_t(u) - \mathbf{X}_{t+h}(u)$, $\mathbf{\Sigma}_{\mathbf{Y}_2,\ell}(u,v) = \mathrm{Cov}\{\mathbf{Y}_{2,t}(u), \mathbf{Y}_{2,(t+\ell)}(v)\}, \ell \in \mathbb{Z}(u,v), \in \mathcal{U}^2$, and $f_{\mathbf{Y}_2,\theta}$ be the spectral density operator of $\mathbf{Y}_2$, $\theta \in [-\pi,\pi]$, we have $f_{\mathbf{Y}_2,\theta} = \{2 - \exp(-ih\theta) - \exp(ih\theta)\} f_{\mathbf{X},\theta}$. Note that

$$4\big\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}_h)(\mathbf{\Phi}_1)\big\rangle_{\mathbb{H}} = \big\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_{\mathbf{Y}_1,0} - \mathbf{\Sigma}_{\mathbf{Y}_1,0})(\mathbf{\Phi}_1)\big\rangle_{\mathbb{H}} - \big\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_{\mathbf{Y}_2,0} - \mathbf{\Sigma}_{\mathbf{Y}_2,0})(\mathbf{\Phi}_1)\big\rangle_{\mathbb{H}}$$

and $\mathcal{M}(f_{\mathbf{Y}_i}, \mathbf{\Phi}_1) \leqslant 4\mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_1)$ for $i = 1,2$. Combing these with results in the proof of (8) leads to

$$\begin{aligned} & P\Big[\Big|\big\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}_h)(\mathbf{\Phi}_1)\big\rangle_{\mathbb{H}}\Big| > 2\mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_1)\eta\Big] \\ \leqslant & \sum_{i=1}^{2} P\Big[\Big|\big\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_{\mathbf{Y}_i,0} - \mathbf{\Sigma}_{\mathbf{Y}_i,0})(\mathbf{\Phi}_1)\big\rangle_{\mathbb{H}}\Big| > \mathcal{M}(f_{\mathbf{Y}_i}, \mathbf{\Phi}_1)\eta\Big] \leqslant 4\exp\Big\{-cn\min\big(\eta^2,\eta\big)\Big\}, \end{aligned}$$

for some constant $c > 0$. This result, together with, $\mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_1) \leqslant \mathcal{M}_k(f_{\mathbf{X}})\big\langle \mathbf{\Phi}_1, \mathbf{\Sigma}_0(\mathbf{\Phi}_1)\big\rangle_{\mathbb{H}}$ implies (15).

Note that

$$4\big\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}_h)(\mathbf{\Phi}_2)\big\rangle_{\mathbb{H}} \leqslant \big\langle \widetilde{\mathbf{\Phi}}_1, (\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}_h)(\widetilde{\mathbf{\Phi}}_1)\big\rangle_{\mathbb{H}} - \big\langle \widetilde{\mathbf{\Phi}}_2, (\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}_h)(\widetilde{\mathbf{\Phi}}_2)\big\rangle_{\mathbb{H}},$$

where $\widetilde{\mathbf{\Phi}}_1 = \mathbf{\Phi}_1 + \mathbf{\Phi}_2$, $\widetilde{\mathbf{\Phi}}_2 = \mathbf{\Phi}_1 - \mathbf{\Phi}_2$ and $\mathcal{M}(f_{\mathbf{X}}, \widetilde{\mathbf{\Phi}}_i) \leqslant 2\{\mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_1) + \mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_2)\}$ for $i = 1,2$. Combing these with results and the proof of (15) leads to

$$\begin{aligned} & P\Big[\Big|\big\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}_h)(\mathbf{\Phi}_2)\big\rangle_{\mathbb{H}}\Big| > 2\{\mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_1) + \mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_2)\}\eta\Big] \\ \leqslant & \sum_{i=1}^{2} P\Big[\Big|\big\langle \widetilde{\mathbf{\Phi}}_i, (\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}_h)(\widetilde{\mathbf{\Phi}}_i)\big\rangle_{\mathbb{H}}\Big| > 2\mathcal{M}(f_{\mathbf{X}}, \widetilde{\mathbf{\Phi}}_i)\eta\Big] \leqslant 8\exp\Big\{-cn\min\big(\eta^2,\eta\big)\Big\} \end{aligned}$$

for some constant $c > 0$. This, together with, $\mathcal{M}(f_{\mathbf{X}}, \mathbf{\Phi}_i) \leqslant \mathcal{M}_k(f_{\mathbf{X}})\big\langle \mathbf{\Phi}_i, \mathbf{\Sigma}_0(\mathbf{\Phi}_i)\big\rangle_{\mathbb{H}}$ for $i = 1,2$, implies (16), which completes the proof. ▫

D.2. Proof of Proposition 2

It is easy to see that $\boldsymbol{\theta}^T\widehat{\boldsymbol{\Gamma}}\boldsymbol{\theta} = \boldsymbol{\theta}^T\boldsymbol{\Gamma}\boldsymbol{\theta} + \boldsymbol{\theta}^T(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})\boldsymbol{\theta}$. Hence we have

$$\boldsymbol{\theta}^T\widehat{\boldsymbol{\Gamma}}\boldsymbol{\theta} \geqslant \boldsymbol{\theta}^T\boldsymbol{\Gamma}\boldsymbol{\theta} - \|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_{\max}\|\boldsymbol{\theta}\|_1^2.$$

By Condition 8, $\lambda_{\min}(\boldsymbol{\Gamma}) \geqslant \underline{\mu}$, where $\lambda_{\min}(\boldsymbol{\Gamma})$ denotes the minimum eigenvalue of $\boldsymbol{\Gamma}$. Together with Lemma 5 in Section D.9, this proposition follows. ▫

D.3. Proof of Proposition 3

Note that on the event $\{|\widehat{\lambda}_{jl} - \lambda_{jl}| \leqslant 2^{-1}\lambda_{jl}\}$, we have $\widehat{\lambda}_{jl} \geqslant \lambda_{jl}/2$, $\widehat{\lambda}_{jl}^{-1/2} \leqslant \sqrt{2}\lambda_{jl}^{-1/2}$ and $|\widehat{\lambda}_{jl}^{-1/2} - \lambda_{jl}^{-1/2}| \leqslant \frac{\widehat{\lambda}_{jl}^{-1}|\widehat{\lambda}_{jl}-\lambda_{jl}|\lambda_{jl}^{-1}}{\widehat{\lambda}_{jl}^{-1/2}+\lambda_{jl}^{-1/2}} \leqslant 2\lambda_{jl}^{-3/2}|\widehat{\lambda}_{jl} - \lambda_{jl}|$, which implies that $\left|\frac{\widehat{\lambda}_{jl}^{-1/2}-\lambda_{jl}^{-1/2}}{\lambda_{jl}^{-1/2}}\right| \leqslant 2\left|\frac{\widehat{\lambda}_{jl}-\lambda_{jl}}{\lambda_{jl}}\right|$. Then it follows from Theorem 3 that there exist positive constants C_λ, C_ϕ, c_5 and c_6 such that the first and second deviation bounds in (25) respectively hold with probability greater than $1 - c_5(pq)^{-c_6}$. The proof is complete. ▫

D.4. Proof of Proposition 4

Notice that

$$\begin{aligned}\widehat{\mathbf{Y}}_j - \widehat{\boldsymbol{\Gamma}}\mathbf{B}_j \;=\; & \Big\{(n-1)^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{V}}_j - (n-1)^{-1}\mathbf{D}^{-1}\mathrm{E}(\mathbf{Z}^{\mathrm{T}}\mathbf{V}_j)\Big\} \\ & +(n-1)^{-1}\mathbf{D}^{-1}\mathrm{E}\Big\{\mathbf{Z}^{\mathrm{T}}(\mathbf{V}_j - \mathbf{Z}\mathbf{D}^{-1}\mathbf{B}_j)\Big\} - \big(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\big)\mathbf{B}_j.\end{aligned} \tag{C.28}$$

First, we show the deviation bounds of $\widehat{\mathbf{D}}^{-1}(n-1)^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{V}}_j - \mathbf{D}^{-1}\mathrm{E}((n-1)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{V}_j)$. We decompose this term as $\widehat{\mathbf{D}}^{-1}\Big\{(n-1)^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{V}}_j - \mathrm{E}((n-1)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{V}_j)\Big\} + (\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1})\mathrm{E}((n-1)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{V}_j)$. It follows from Theorem 4 that there exists positive constants C_1^*, c_5 and c_6 that

$$\sup_{j,k}\Big\|\mathbf{D}_k^{-1}\Big\{(n-1)^{-1}\widehat{\mathbf{Z}}_k^{\mathrm{T}}\widehat{\mathbf{V}}_j - \mathrm{E}\big((n-1)^{-1}\mathbf{Z}_k^{\mathrm{T}}\mathbf{V}_j\big)\Big\}\mathbf{D}_j^{-1}\Big\|_{\max} \leqslant C_1^*\mathcal{M}_1(f_{\mathbf{X}})q^{\alpha+1}\sqrt{\frac{\log(pq)}{n}}, \tag{C.29}$$

with probability greater than $1 - c_5(pq)^{-c_6}$. Note that $\widehat{\mathbf{D}}_k = \mathrm{diag}(\widehat{\lambda}_{k1}^{1/2}, \ldots, \widehat{\lambda}_{kq}^{1/2})$ and $\mathbf{D}_k = \mathrm{diag}(\lambda_{k1}^{1/2}, \ldots, \lambda_{kq}^{1/2})$, it follows from Proposition 3 that there exists positive constant C_2^*, such that

$$\Big\|\Big(\widehat{\mathbf{D}}^{-1} - \mathbf{D}^{-1}\Big)\mathbf{D}\Big\|_{\max} \leqslant C_2^*\mathcal{M}_1(f_{\mathbf{X}})\sqrt{\frac{\log(pq)}{n}}, \tag{C.30}$$

with probability great than $1-c_5(pq)^{-c_6}$. By Condition 1, we have $\max_j ||\mathbf{D}_j||_F \leqslant \lambda_0^{1/2}$ and $\|\mathbf{D}^{-1}\mathrm{E}\big((n-1)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{V}_j\big)\|_{\max}^{(q)} \leqslant q^{1/2}\|\mathbf{D}^{-1}\mathrm{E}\big((n-1)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{V}_j\big)\mathbf{D}_j^{-1}\|_{\max}||\mathbf{D}_j||_F = O(q^{1/2})$, where the fact that, for $q\times q$ matrix $\mathbf{A}$ and a diagonal matrix $\mathbf{B}$, $\|\mathbf{AB}\|_F \leqslant q^{1/2}\|\mathbf{A}\|_{\max}\|\mathbf{B}\|_F$, is used. These results together with (C.29) and (C.30) imply that there exists C_3^*

$$\left\|\widehat{\mathbf{D}}^{-1}(n-1)^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{V}}_j - \mathbf{D}^{-1}\mathrm{E}\big((n-1)^{-1}\mathbf{Z}^{\mathrm{T}}\mathbf{V}_j\big)\right\|_{\max}^{(q)} \leqslant C_3^*\mathcal{M}_1(f_{\mathbf{X}})q^{\alpha+3/2}\sqrt{\frac{\log(pq)}{n}} \tag{C.31}$$

Second, consider the bias term $(n-1)^{-1}\mathbf{D}^{-1}\mathrm{E}\{\mathbf{Z}^{\mathrm{T}}(\mathbf{V}_j-\mathbf{Z}\mathbf{D}^{-1}\mathbf{B}_j)\}$. By Section F.1, $\mathbf{R}_j$ is a $(n-1)\times q$ matrix whose row vectors are formed by $\{\mathbf{r}_{tj}, t=2,\ldots,n\}$ with $\mathbf{r}_{tj}=(r_{tj1},\ldots,r_{tjq})^{\mathrm{T}}$ and $r_{tjl}=\sum_{k=1}^p\sum_{m=q+1}^{\infty}\langle\phi_{jl},\langle A_{jk},\phi_{km}\rangle\rangle\xi_{(t-1)km}$ for $l=1,\ldots,q$. It follows from Conditions 1, 4 and similar arguments in deriving (C.27) and (C.31) that there exists some positive constant C_4^* such that

$$\begin{aligned}&\big\|(n-1)^{-1}\mathbf{D}^{-1}\mathrm{E}\{\mathbf{Z}^{\mathrm{T}}(\mathbf{V}_j-\mathbf{Z}\mathbf{D}^{-1}\mathbf{B}_j)\}\big\|_{\max}^{(q)}\\ &\leqslant q^{1/2}\big\|(n-1)^{-1}\mathbf{D}^{-1}\mathrm{E}(\mathbf{Z}^{\mathrm{T}}\mathbf{R}_j)\widetilde{\mathbf{D}}^{-1}\big\|_{\max}||\widetilde{\mathbf{D}}||_F \leqslant C_4^* s_j q^{-\beta+1},\end{aligned} \tag{C.32}$$

where $\widetilde{\mathbf{D}}=\lambda_0\mathbf{I}_q$.

Third, it follows from Lemma 5 in Section D.9, Lemma 15 in the Supplementary Material of Qiao et al. (2020) and $||\mathbf{B}||_1^{(q)} = O\big(\lambda_0^{1/2}s_j\big)$ that there exist some positive constants C_5^* such that

$$\left\|(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma})\mathbf{B}_j\right\|_{\max}^{(q)} \leqslant \left\|\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right\|_{\max}^{(q)}||\mathbf{B}_j||_1^{(q)} \leqslant \mathcal{M}_1(f_{\mathbf{X}})s_jq^{\alpha+2}\sqrt{\frac{\log(pq)}{n}}, \tag{C.33}$$

with probability great than $1-c_5(pq)^{-c_6}$.

Combing results in (C.28), (C.31), (C.32) and (C.33) implies that there exist positive constants C_E, c_4 and c_5 such that

$$||\widehat{\mathbf{Y}}_j-\widehat{\boldsymbol{\Gamma}}\mathbf{B}_j||_{\max}^{(q)} \leqslant C_E\mathcal{M}_1(f_{\mathbf{X}})s_j\Big\{q^{\alpha+2}\sqrt{\frac{\log(pq)}{n}}+q^{-\beta+1}\Big\},\ j=1,\ldots,p,$$

with probability greater than $1-c_5(pq)^{-c_6}$. The proof is complete. □

D.5. Proposition 5 and its proof

Proposition 5. *Suppose that Conditions 1 and 2 hold. Then there exists some universal constant $\tilde{c}>0$ such that for any $\eta>0$*

$$P\left\{\left\|\widehat{\boldsymbol{\Sigma}}_0-\boldsymbol{\Sigma}_0\right\|_F > 2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\eta\right\} \leqslant \frac{p^2}{\eta^2 n}\big(16\tilde{c}^{-1}+128\tilde{c}^{-2}n^{-1}\big). \tag{C.34}$$

In particular, if the sample size n satisfies the bound $n > 128(\tilde{\rho}^2\tilde{c}^2 - 16\tilde{c})^{-1}$, where $\tilde{\rho}$ is some positive constant with $\tilde{\rho} > 4\tilde{c}^{-1/2}$, then with probability greater than $1-\tilde{\rho}^{-2}\big(16\tilde{c}^{-1}+128\tilde{c}^{-2}n^{-1}\big)$, the estimate $\widehat{\boldsymbol{\Sigma}}_0$ satisfies the bound

$$\big\|\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0\big\|_F \leqslant 2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\tilde{\rho}\sqrt{\frac{p^2}{n}}. \tag{C.35}$$

Proof. It follows from the definition of $\|\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0\|_F^2 = \sum_{j,k=1}^p \|\widehat{\Sigma}_{jk}^{(0)} - \Sigma_{jk}^{(0)}\|_{\mathcal{S}}^2$, Chebyshev's inequality and (C.8) with $q=1$ that for any $\eta > 0$,

$$\begin{aligned} P\Big\{\|\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0\|_F > 2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\eta\Big\} &\leqslant \frac{1}{\left(2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\right)^2\eta^2}\sum_{j,k=1}^p \mathrm{E}\|\widehat{\Sigma}_{jk}^{(0)} - \Sigma_{jk}^{(0)}\|_{\mathcal{S}}^2 \\ &\leqslant \frac{p^2}{\eta^2}(16\tilde{c}^{-1}n^{-1} + 128\tilde{c}^{-2}n^{-2}) \\ &= \frac{p^2}{\eta^2 n}(16\tilde{c}^{-1} + 128\tilde{c}^{-2}n^{-1}). \end{aligned}$$

By letting $\eta = \tilde{\rho}\sqrt{p^2/n}$ with $\rho > 0$, we have that

$$P\left\{\|\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0\|_F > 2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\tilde{\rho}\sqrt{\frac{p^2}{n}}\right\} \leqslant \tilde{\rho}^{-2}(16\tilde{c}^{-1} + 128\tilde{c}^{-2}n^{-1}).$$

The proof is complete. ▫

D.6. Proof of Lemma 1

By Lemma 7, we obtain that

$$\frac{\widehat{\lambda}_{jl} - \lambda_{jl}}{\lambda_{jl}} = \frac{\big\langle \phi_{jl}, \widehat{\Delta}_{jj}(\phi_{jl})\big\rangle}{\lambda_{jl}} + \frac{R_{jl}}{\lambda_{jl}},\ \ j = 1, \dots, p, l = 1, \dots, L. \tag{C.36}$$

Note that $\lambda_{jl} = \big\langle \phi_{jl}, \Sigma_{jj}(\phi_{jl})\big\rangle$. It follows from (8) in Theorem 1 that for any $\eta > 0$,

$$P\left\{\left|\frac{\big\langle \phi_{jl}, \widehat{\Delta}_{jj}(\phi_{jl})\big\rangle}{\lambda_{jl}}\right| > \mathcal{M}_1(f_{\mathbf{X}})\eta\right\} \leqslant 2\exp\Big\{-cn\min(\eta^2, \eta)\Big\}. \tag{C.37}$$

We next turn to the term $\big|R_{jl}/\lambda_{jl}\big|$. By (C.9), Lemma 7 and Condition 3 with $\delta_{jl} \geqslant c_0 l^{-\alpha-1}$ and $\lambda_{jl} \geqslant c_0\alpha^{-1}l^{-\alpha}$, we have

$$\left|\frac{R_{jl}}{\lambda_{jl}}\right| \leqslant 4\sqrt{2}c_0^{-2}\alpha l^{2\alpha+1}\|\widehat{\Delta}_{jj}\|_{\mathcal{S}}^2.$$

It then follows from (10) in Theorem 2 that there exists some constant $\tilde{c} > 0$ such that for any $\eta > 0$

$$P\left\{\left|\frac{R_{jl}}{\lambda_{jl}}\right| > 4\sqrt{2}c_0^{-2}\alpha l^{2\alpha+1}\{2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\eta\}^2\right\} \leqslant 4\exp\Big\{-\tilde{c}n\min(\eta^2,\eta)\Big\}. \tag{C.38}$$

Let $\tilde{c}_1 = \min(c,\tilde{c})$. It follows from $\rho_1 = 16\sqrt{2}c_0^{-2}\alpha\lambda_0^2$, (C.36), (C.37) and (C.38) that (C.10) in Lemma 1 holds, which completes our proof. □

D.7. Proof of Lemma 2

It follows from (C.9), Condition 3 with $\delta_{jl} \geqslant c_0 l^{-\alpha-1}$ and (10) in Theorem 2 that there exists some universal constant $\tilde{c}$ such that for any $\eta > 0$, (C.11) holds.□

D.8. Lemma 4 and its proof

Lemma 4. *Suppose that Conditions 1–3 hold. Then there exists some universal constant $\tilde{c}_2 > 0$ such that for each $j = 1,\ldots,p, l = 1,\ldots,d_j$, any given function $g \in \mathbb{H}$ and $\eta > 0$,*

$$\begin{aligned}&P\left\{\left|\big\langle\widehat{\phi}_{jl}-\phi_{jl},g\big\rangle\right| \geqslant \rho_2\|g^{-jl}\|_\lambda\mathcal{M}_1(f_{\mathbf{X}})\lambda_{jl}^{1/2}l^{\alpha+1}\eta + \rho_3\|g\|\mathcal{M}_1^2(f_{\mathbf{X}})l^{2(\alpha+1)}\eta^2\right\}\\&\leqslant 8\exp\Big\{-\tilde{c}_2 n\min(\eta^2,\eta)\Big\} + 4\exp\Big\{-\tilde{c}_2\mathcal{M}_1^{-2}(f_{\mathbf{X}})nl^{-2(\alpha+1)}\Big\},\end{aligned} \tag{C.39}$$

where $g(\cdot) = \sum_{m=1}^\infty g_{jm}\phi_{jm}(\cdot)$, $\|g^{-jl}\|_\lambda = \big(\sum_{m:m\neq l}\lambda_{jm}g_{jm}^2\big)^{1/2}$, $\rho_2 = 2c_0^{-1}$ and $\rho_3 = 4(6+2\sqrt{2})c_0^{-2}\lambda_0^2$ with $c_0 \leqslant 4\mathcal{M}_1(f_{\mathbf{X}})\lambda_0 l^{\alpha+1}$.

Proof. It follows from the expansion $g(\cdot) = \sum_{m=1}^\infty g_{jm}\phi_{jm}(\cdot)$ and (C.43) in Lemma 7 in Section D.11 that

$$\begin{aligned}\big\langle\widehat{\phi}_{jl}-\phi_{jl},g\big\rangle &= \sum_{m:m\neq l}(\widehat{\lambda}_{jl}-\lambda_{jm})^{-1}g_{jm}\big\langle\widehat{\phi}_{jl},\langle\widehat{\Delta}_{jj},\phi_{jm}\rangle\big\rangle + g_{jl}\big\langle\widehat{\phi}_{jl}-\phi_{jl},\phi_{jl}\big\rangle\\&= \sum_{m:m\neq l}\Big\{(\widehat{\lambda}_{jl}-\lambda_{jm})^{-1}-(\lambda_{jl}-\lambda_{jm})^{-1}\Big\}g_{jm}\big\langle\widehat{\phi}_{jl},\langle\widehat{\Delta}_{jj},\phi_{jm}\rangle\big\rangle\\&\quad+\sum_{m:m\neq l}(\lambda_{jl}-\lambda_{jm})^{-1}g_{jm}\big\langle\widehat{\phi}_{jl}-\phi_{jl},\langle\widehat{\Delta}_{jj},\phi_{jm}\rangle\big\rangle\\&\quad+\sum_{m:m\neq l}(\lambda_{jl}-\lambda_{jm})^{-1}g_{jm}\big\langle\phi_{jl},\langle\widehat{\Delta}_{jj},\phi_{jm}\rangle\big\rangle + g_{jl}\big\langle\widehat{\phi}_{jl}-\phi_{jl},\phi_{jl}\big\rangle\\&= I_1+I_2+I_3+I_4.\end{aligned}$$

Let $\Omega_{d_j} = \{2\|\widehat{\Delta}_{jj}\|_{\mathcal{S}} \leqslant \delta_{jd_j}\}$. It follows from Condition 3 and (10) in Theorem 2 with the choice of $\eta = \{4\mathcal{M}_1(f_{\mathbf{X}})\lambda_0 l^{\alpha+1}\}^{-1}c_0 \leqslant 1$ that

$$P\big(\Omega_{d_j}^C\big) \leqslant P\big(\|\widehat{\Delta}_{jj}\|_{\mathcal{S}} \geqslant 2^{-1}c_0 l^{-\alpha-1}\big) \leqslant 4\exp\Big\{-16^{-1}\tilde{c}\mathcal{M}_1^{-2}(f_{\mathbf{X}})\lambda_0^{-2}c_0^2 l^{-2(\alpha+1)}n\Big\}. \tag{C.40}$$

On the event Ω_{d_j}, we can see that $\sup_{l\leqslant d_j}|\widehat{\lambda}_{jl}-\lambda_{jl}| \leqslant \lambda_{jd_j}/2$, which implies that $2^{-1}\lambda_{jl} \leqslant \widehat{\lambda}_{jl} \leqslant 2\lambda_{jl}$. Moreover, $|\widehat{\lambda}_{jl}-\lambda_{jl}| \leqslant 2^{-1}|\lambda_{jl}-\lambda_{jm}|$ for $1 \leqslant l \neq m \leqslant d_j$ and hence $|\widehat{\lambda}_{jl}-\lambda_{jm}| \geqslant 2^{-1}|\lambda_{jl}-\lambda_{jm}|$ for $j = 1,\ldots,p$. By Condition 3, $|\lambda_{jl}-\lambda_{jm}| \geqslant c_0 l^{-\alpha-1}$ for $1 \leqslant m \neq l \leqslant d_j$. Using the above results, we have

$$\begin{aligned}
|I_1|^2 &\leqslant \big(\widehat{\lambda}_{jl}-\lambda_{jl}\big)^2 \sum_{m:m\neq l}(\widehat{\lambda}_{jl}-\lambda_{jm})^{-2}(\lambda_{jl}-\lambda_{jm})^{-2}g_{jm}^2\|\widehat{\Delta}_{jj}\|_{\mathcal{S}}^2\\
&\leqslant 4\big(\widehat{\lambda}_{jl}-\lambda_{jl}\big)^2\|\widehat{\Delta}_{jj}\|_{\mathcal{S}}^2 \sum_{m:m\neq l}(\lambda_{jl}-\lambda_{jm})^{-4}g_{jm}^2\\
&\leqslant 4c_0^{-4}\|g^{-jl}\|^2 l^{4(\alpha+1)}\big(\widehat{\lambda}_{jl}-\lambda_{jl}\big)^2\|\widehat{\Delta}_{jj}\|_{\mathcal{S}}^2,
\end{aligned}$$

where $||g^{-jl}|| = (\sum_{m:m\neq l} g_{jm}^2)^{1/2}$. This together with (C.9) implies that, on the event Ω_{d_j},

$$|I_1| \leqslant 2c_0^{-2}\|g^{-jl}\| l^{2(\alpha+1)}\big\|\widehat{\Delta}_{jj}\big\|_{\mathcal{S}}^2.$$

Similarly, we can show that

$$|I_2| \leqslant c_0^{-1}\|g^{-jl}\| l^{\alpha+1}\|\widehat{\phi}_{jl}-\phi_{jl}\|\|\widehat{\Delta}_{jj}\|_{\mathcal{S}} \leqslant 2\sqrt{2}c_0^{-2}\|g^{-jl}\| l^{2(\alpha+1)}\|\widehat{\Delta}_{jj}\|_{\mathcal{S}}^2.$$

Moreover, by the result $||\widehat{\phi}_{jl}-\phi_{jl}||^2 = \langle\widehat{\phi}_{jl}-\phi_{jl}, -2\phi_{jl}\rangle$ and (C.9) we have

$$|I_4| = 2^{-1}|g_{jl}|\|\widehat{\phi}_{jl}-\phi_{jl}\|^2 \leqslant 4c_0^{-2}|g_{jl}| l^{2(\alpha+1)}\|\widehat{\Delta}_{jj}\|_{\mathcal{S}}^2.$$

Combing the above upper bound results, we have

$$|I_1|+|I_2|+|I_4| \leqslant (6+2\sqrt{2})c_0^{-2}\|g\| l^{2(\alpha+1)}\|\widehat{\Delta}_{jj}\|_{\mathcal{S}}^2.$$

Let $\widetilde{\lambda}_g = \sum_{m:m\neq l}\lambda_{jm}(\lambda_{jl}-\lambda_{jm})^{-2}g_{jm}^2 \leqslant c_0^{-2}l^{2(\alpha+1)}\|g^{-jl}\|_{\lambda}^2$. Then it follows from (9) in Theorem 1 that

$$P\Big\{\Big|\lambda_{jl}^{-1/2}\widetilde{\lambda}_g^{-1/2}I_3\Big| \geqslant 2\mathcal{M}_1(f_{\mathbf{X}})\eta\Big\} \leqslant 4\exp\Big\{-cn\min(\eta^2,\eta)\Big\}. \tag{C.41}$$

Define $\Omega_{1,\eta} = \Big\{\|\widehat{\Delta}_{jj}\|_{\mathcal{S}} \leqslant 2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\eta\Big\}$ and $\Omega_{2,\eta} = \Big\{\big|I_3\big| \leqslant 2c_0^{-1}\lambda_{jl}^{1/2}\|g^{-jl}\|_{\lambda}\mathcal{M}_1(f_{\mathbf{X}})l^{\alpha+1}\eta\Big\}$. Let $\rho_2 = 2c_0^{-1}$ and $\rho_3 = 4(6+2\sqrt{2})c_0^{-2}\lambda_0^2$. Under the event $\Omega_{d_j}\cap\Omega_{1,\eta}\cap\Omega_{2,\eta}$, we obtain that

$$\Big|\big\langle\widehat{\phi}_{jl}-\phi_{jl}, g\big\rangle\Big| \leqslant \rho_2\|g^{-jl}\|_{\lambda}\mathcal{M}_1(f_{\mathbf{X}})\lambda_{jl}^{1/2}l^{\alpha+1}\eta + \rho_3\|g\|\mathcal{M}_1^2(f_{\mathbf{X}})l^{2(\alpha+1)}\eta^2.$$

Let $\tilde{c}_2 = \min(16^{-1}\lambda_0^{-2}c_0^2\tilde{c}, \tilde{c}_1)$. It follows from (10) in Theorem 2 and (C.41) that

$$P\big(\Omega_{1,\eta}^C \cup \Omega_{2,\eta}^C\big) \leqslant 8\exp\Big\{-\tilde{c}_2 n\min(\eta^2,\eta)\Big\}.$$

This together with (C.40) completes the proof of (C.39). ▫

D.9. Lemma 5 and its proof

Lemma 5. *Suppose that Conditions 1–3 hold. Then there exist some positive constants* C_Γ, c_5 *and* c_6 *such that*

$$\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_{\max} \leqslant C_\Gamma \mathcal{M}_1(f_{\mathbf{X}}) q^{\alpha+1} \sqrt{\frac{\log(pq)}{n}}$$

with probability greater than $1 - c_5(pq)^{-c_6}$.

Proof. Note that

$$\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_{\max} = \max_{1\leqslant j,k\leqslant p, 1\leqslant l,m\leqslant q} \left| \widehat{\lambda}_{jl}^{-1/2} \widehat{\lambda}_{km}^{-1/2} \widehat{\sigma}_{jklm} - \lambda_{jl}^{-1/2} \lambda_{km}^{-1/2} \sigma_{jklm} \right|.$$

Let $\widehat{s}_{jklm} = \frac{\widehat{\lambda}_{jl}\widehat{\lambda}_{km}}{\lambda_{jl}\lambda_{km}}$ for each (j,k,l,m). Then we have

$$\widehat{\lambda}_{jl}^{-1/2} \widehat{\lambda}_{km}^{-1/2} \widehat{\sigma}_{jklm} - \lambda_{jl}^{-1/2} \lambda_{km}^{-1/2} \sigma_{jklm} = \widehat{s}_{jklm}^{-1/2} \left(\frac{\widehat{\sigma}_{jklm} - \sigma_{jklm}}{\lambda_{jl}^{1/2} \lambda_{km}^{1/2}} \right) + \left(\widehat{s}_{jklm}^{-1/2} - 1 \right) \frac{\sigma_{jklm}}{\lambda_{jl}^{1/2} \lambda_{km}^{1/2}}.$$

Let $\Omega_\lambda = \left\{ \sup_{1\leqslant j\leqslant p, 1\leqslant l\leqslant q} \left| \frac{\widehat{\lambda}_{jl} - \lambda_{jl}}{\lambda_{jl}} \right| \leqslant 1/5 \right\}$. Observe that

$$\widehat{s}_{jklm} - 1 = \left(\frac{\widehat{\lambda}_{jl} - \lambda_{jl}}{\lambda_{jl}} + 1 \right) \left(\frac{\widehat{\lambda}_{km} - \lambda_{km}}{\lambda_{km}} \right) + \frac{\widehat{\lambda}_{jl} - \lambda_{jl}}{\lambda_{jl}}.$$

Then under the event Ω_λ, we have $|\widehat{s}_{jklm} - 1| \leqslant 1/2$, and thus $\widehat{s}_{jklm}^{-1/2} \leqslant \sqrt{2}$. Moreover, provided that fact that $|(1+x)^{-1/2} - 1| \leqslant x$ if $|x| \leqslant 1/2$, we have

$$\left| \widehat{s}_{jklm}^{-1/2} - 1 \right| \leqslant \frac{6}{5} \left(\left| \frac{\widehat{\lambda}_{km} - \lambda_{km}}{\lambda_{km}} \right| + \left| \frac{\widehat{\lambda}_{jl} - \lambda_{jl}}{\lambda_{jl}} \right| \right).$$

Under the event Ω_λ, the above results together with the fact of $\sigma_{jklm} \leqslant \lambda_{jl}^{1/2} \lambda_{km}^{1/2}$ imply that

$$\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_{\max} \leqslant \sqrt{2} \max_{1\leqslant j,k\leqslant p, 1\leqslant l,m\leqslant q} \left| \frac{\widehat{\sigma}_{jklm} - \sigma_{jklm}}{\lambda_{jl}^{1/2} \lambda_{km}^{1/2}} \right| + \frac{12}{5} \max_{1\leqslant j\leqslant p, 1\leqslant l\leqslant q} \left| \frac{\widehat{\lambda}_{jl} - \lambda_{jl}}{\lambda_{jl}} \right|.$$

Then it follows from Theorems 3 and 4 that there exist some positive constants C_Γ, c_5 and c_6 such that

$$\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\|_{\max} \leqslant C_\Gamma \mathcal{M}_1(f_{\mathbf{X}}) q^{\alpha+1} \sqrt{\frac{\log(pq)}{n}}$$

with probability greater than $1 - c_5(pq)^{-c_6}$. The proof is complete. □

D.10. Lemma 6 and its proof

The following lemma shows how to derive the tail probability through moment conditions.

Lemma 6. *Let X be a random variable. If for some constants $c_1, c_2 > 0$*

$$P(|X| > t) \leqslant c_1 \exp\{-c_2^{-1} \min(t^2, t)\} \quad \text{for any } t > 0,$$

then for any integer $q \geqslant 1$,

$$E(X^{2q}) \leqslant q! c_1 (4c_2)^q + (2q)! c_1 (4c_2)^{2q}.$$

Conversely, if for some positive constants a_1, a_2, $E(X^{2q}) \leqslant q! a_1 a_2^q + (2q)! a_1 a_2^{2q}$, $q \geqslant 1$, then by letting $c_2^* = 8 \max\{4(a_2 + a_2^2), a_2\}$ and $c_1^* = a_1$, we have that

$$P(|X| > t) \leqslant c_1^* \exp\{-c_2^{*-1} \min(t^2, t)\} \quad \text{for any } t > 0.$$

Proof. This lemma can be proved in a similar way to Theorem 2.3 of Boucheron, Lugosi and Massart (2014) and hence the proof is omitted here. In the proof, the following two inequalities are used, i.e. for any $c > 0$ and $t > 0$,

$$\frac{1}{2} \min(t^2, t) \leqslant \frac{t^2}{1+t} \leqslant \min(t^2, t),$$

and

$$\sqrt{\frac{ct}{2}} + \frac{ct}{2} \leqslant \frac{c(t + \sqrt{t^2 + 4t/c})}{2} \leqslant \sqrt{ct} + ct. \quad \square$$

D.11. Lemma 7 and its proof

Lemma 7. *For each $j = 1, \ldots, p$ and $l = 1, \ldots,$ the term of $\widehat{\lambda}_{jl} - \lambda_{jl}$ can be expressed as*

$$\widehat{\lambda}_{jl} - \lambda_{jl} = \big\langle \phi_{jl}, \widehat{\Delta}_{jj}(\phi_{jl}) \big\rangle + R_{jl}, \tag{C.42}$$

where $|R_{jl}| \leqslant 2\|\widehat{\phi}_{jl} - \phi_{jl}\| \|\widehat{\Delta}_{jj}\|_{\mathcal{S}}$. Furthermore, if $\inf_{m:m \neq l} |\widehat{\lambda}_{jl} - \lambda_{jm}| > 0$, then

$$\widehat{\phi}_{jl} - \phi_{jl} = \sum_{m:m \neq l} (\widehat{\lambda}_{jl} - \lambda_{jm})^{-1} \phi_{jm} \big\langle \widehat{\phi}_{jl}, \widehat{\Delta}_{jj}(\phi_{jm}) \big\rangle + \phi_{jl} \big\langle \widehat{\phi}_{jl} - \phi_{jl}, \phi_{jl} \big\rangle. \tag{C.43}$$

Proof. This lemma follows directly from Lemma 5.1 of Hall and Horowitz (2007) and hence the proof is omitted here. $\square$

D.12. Lemma 8 and its proof

Lemma 8. *For a $p \times p$ lag-h autocovariance function, $\mathbf{\Sigma}_h = \big(\Sigma_{jk}^{(h)}\big)_{1 \leqslant j,k \leqslant p}$, $h = 0, 1, \ldots,$ we have*

$$\|\Sigma_{jk}^{(h)}\|_{\mathcal{S}} \leqslant \lambda_0 \quad \text{and} \quad \|\Sigma_{jk}^{(h)}(\phi_{km})\| \leqslant \lambda_{km}^{1/2}\lambda_0^{1/2} \quad \text{for } m \geqslant 1.$$

Proof. By the expansion $\Sigma_{jk}^{(h)}(u,v) = \sum_{l,m=1}^{\infty} E\{\xi_{tjl}\xi_{(t+h)km}\}\phi_{jl}(u)\phi_{km}(v)$, the orthonormality of $\{\phi_{jl}(\cdot)\}$ for $\{\phi_{km}(\cdot)\}$ for each (j,k) and the Cauchy-Schwarz inequality, we have

$$\|\Sigma_{jk}^{(h)}\|_{\mathcal{S}} \leqslant \Big[\sum_{l,m=1}^{\infty} E(\xi_{tjl}^2)E\{\xi_{(t+h)km}^2\}\Big]^{1/2} \leqslant \Big\{\sum_{l=1}^{\infty}\lambda_{jl}\sum_{m=1}^{\infty}\lambda_{km}\Big\}^{1/2} \leqslant \lambda_0.$$

Moreover, applying similar techniques, we have

$$\begin{aligned}
\|\Sigma_{jk}^{(h)}(\phi_{km})\|^2 &= \int_{\mathcal{U}}\Big(\sum_{l=1}^{\infty}E(\xi_{tjl}\xi_{tkm})\phi_{jl}(u)\Big)^2 du \\
&\leqslant \sum_{l=1}^{\infty}E(\xi_{tjl}^2)E(\xi_{tkm}^2) = \int_{\mathcal{U}}\Sigma_{jj}(u,u)du\lambda_{km} \leqslant \lambda_0\lambda_{km},
\end{aligned}$$

which completes the proof. ▫.

D.13. Lemma 9 and its proof

Lemma 9. *For each $j,k = 1,\ldots,p$, let $\{\phi_{jl}(\cdot)\}_{1 \leqslant l \leqslant q}$ and $\{\widehat{\phi}_{jl}(\cdot)\}_{1 \leqslant l \leqslant q}$ correspond to true and estimated eigenfunctions, respectively, and $\widehat{\psi}_{jklm}$ be the estimate of ψ_{jklm} for $l,m = 1,\ldots,q$. Then we have*

$$\begin{aligned}
\Big\|\sum_{l=1}^{q}\sum_{m=1}^{q}\widehat{\phi}_{km}(\cdot)\widehat{\psi}_{jklm}\Big\{\widehat{\phi}_{jl}(\cdot) - \phi_{jl}(\cdot)\Big\}\Big\|_{\mathcal{S}}^2 &\leqslant \sum_{l=1}^{q}\|\widehat{\phi}_{jl} - \phi_{jl}\|^2\sum_{l=1}^{q}\sum_{m=1}^{q}\widehat{\psi}_{jklm}^2, \\
\Big\|\sum_{l=1}^{q}\sum_{m=1}^{q}\phi_{km}(\cdot)(\widehat{\psi}_{jklm} - \psi_{jklm})\phi_{jl}(\cdot)\Big\|_{\mathcal{S}}^2 &= \sum_{l=1}^{q}\sum_{m=1}^{q}(\widehat{\psi}_{jklm} - \psi_{jklm})^2.
\end{aligned}$$

Proof. We prove the first result

$$\begin{aligned}
&\Big\|\sum_{l=1}^{q}\sum_{m=1}^{q}\widehat{\phi}_{km}\widehat{\psi}_{jklm}(\widehat{\boldsymbol{\phi}}_{jl} - \boldsymbol{\phi}_{jl})\Big\|_{\mathcal{S}}^2 \\
&= \sum_{m=1}^{q}\Big\|\sum_{l=1}^{q}\widehat{\psi}_{jklm}(\widehat{\boldsymbol{\phi}}_{jl} - \boldsymbol{\phi}_{jl})\Big\|^2 \leqslant \sum_{m=1}^{q}\sum_{l=1}^{q}\widehat{\psi}_{jklm}^2\sum_{l=1}^{q}\|\widehat{\boldsymbol{\phi}}_{jl} - \boldsymbol{\phi}_{jl}\|^2,
\end{aligned}$$

where the first equality from the orthonormality of $\{\widehat{\phi}_{km}(\cdot)\}_{1\leqslant m\leqslant q}$ and the second inequality comes from Cauchy-Schwarz inequality. By the orthonormality of of $\{\phi_{km}(\cdot)\}_{1\leqslant m\leqslant q}$ and $\{\phi_{jl}(\cdot)\}_{1\leqslant l\leqslant q}$, we can prove the second result

$$\begin{aligned} & \left\| \sum_{l=1}^{q} \sum_{m=1}^{q} \phi_{km}(\widehat{\psi}_{jklm} - \psi_{jklm})\boldsymbol{\phi}_{jl} \right\|_{\mathcal{S}}^2 \\ = & \sum_{m=1}^{q} \left\| \sum_{l=1}^{q} (\widehat{\psi}_{jklm} - \psi_{jklm})\boldsymbol{\phi}_{jl} \right\|^2 = \sum_{m=1}^{q} \sum_{l=1}^{q} (\widehat{\psi}_{jklm} - \psi_{jklm})^2, \end{aligned}$$

which completes the proof. ▫

D.14. Lemma 10 and its proof

Lemma 10. *Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{pq\times q}$ with j-th blocks given by $\mathbf{A}_j, \mathbf{B}_j \in \mathbb{R}^{q\times q}$, respectively. We have*

$$\langle\langle \mathbf{A}, \mathbf{B} \rangle\rangle \leqslant ||\mathbf{B}||_{\max}^{(q)} ||\mathbf{A}||_1^{(q)}. \tag{C.44}$$

Proof. By the definition and Cauchy-Schwarz inequality

$$\begin{aligned} \langle\langle \mathbf{A}, \mathbf{B} \rangle\rangle &= \sum_{j=1}^{p} \langle\langle \mathbf{A}_j, \mathbf{B}_j \rangle\rangle \\ &\leqslant \sum_{j=1}^{p} \langle\langle \mathbf{A}_j, \mathbf{A}_j \rangle\rangle^{1/2} \langle\langle \mathbf{B}_j, \mathbf{B}_j \rangle\rangle^{1/2} \\ &\leqslant \max_j ||\mathbf{B}_j||_F \sum_{j=1}^{p} ||\mathbf{A}_j||_F = ||\mathbf{B}||_{\max}^{(q)} ||\mathbf{A}||_1^{(q)}, \end{aligned}$$

which completes the proof. ▫

E. An illustrative example

In the following, for any $\mathbf{A} = (A_{jk})_{1\leqslant j,k\leqslant p}, \mathbf{B} = (B_{jk})_{1\leqslant j,k\leqslant p}$ with their (j,k)-th components $A_{jk}, B_{jk} \in \mathbb{S}$ and $\mathbf{x} \in \mathbb{H}$, write $\mathbf{AB}$, $\mathbf{Ax}$ and $\mathbf{x}^{\mathrm{T}}\mathbf{A}$ for

$$\int_{\mathcal{U}} \mathbf{A}(u,v')\mathbf{B}(v',v)dv' \ , \ \int_{\mathcal{U}} \mathbf{A}(u,v)\mathbf{x}(v)dv \text{ and } \int_{\mathcal{U}} \mathbf{x}(u)^{\mathrm{T}}\mathbf{A}(u,v)du,$$

respectively. For a $p\times p$ matrix, $\mathbf{C}$, we denote its maximum eigenvalue, spectral radius and operator norm by $\lambda_{\max}(\mathbf{C})$, $\rho(\mathbf{C}) = |\lambda_{\max}(\mathbf{C})|$ and $\|\mathbf{C}\| = \sqrt{\lambda_{\max}(\mathbf{C}^{\mathrm{T}}\mathbf{C})}$, respectively.

Let $p = 2$, $\mathbf{x}_t = \big(x_{t1}, x_{t2}\big)^{\mathrm{T}}$, $\boldsymbol{\psi} = \mathrm{diag}\big(\psi_1, \psi_2\big)$, $\mathbf{C} = \begin{pmatrix} a & b \\ 0 & a \end{pmatrix}$ and $\mathbf{e}_t = \big(e_{t1}, e_{t2}\big)^{\mathrm{T}}$, then the VFAR(1) model in (4) and (5) can be rewritten as

$$\boldsymbol{\psi}(u)\mathbf{x}_t = \int_{\mathcal{U}} \boldsymbol{\psi}(u)\mathbf{C}\boldsymbol{\psi}(v)\boldsymbol{\psi}(v)\mathbf{x}_{t-1}dv + \boldsymbol{\psi}(u)\mathbf{e}_t,$$

which leads to a VAR(1) model

$$\mathbf{x}_t = \mathbf{C}\mathbf{x}_{t-1} + \mathbf{e}_t. \tag{D.45}$$

Provided that $\mathbf{A}(u,v) = \boldsymbol{\psi}(u)\mathbf{C}\boldsymbol{\psi}(v)$ and $||\mathbf{C}|| = \sqrt{\lambda_{\max}(\mathbf{C}^{\mathrm{T}}\mathbf{C})} = \lambda_1$ with $\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{y} = \lambda_1^2\mathbf{y}$ for $||\mathbf{y}|| = 1$, it is easy to see that

$$\mathbf{A}^{\mathrm{T}}\mathbf{A} = \int \mathbf{A}^{\mathrm{T}}(u,v')\mathbf{A}(v',v)dv' = \int \boldsymbol{\psi}(u)\mathbf{C}^{\mathrm{T}}\boldsymbol{\psi}(v')\boldsymbol{\psi}(v')\mathbf{C}\boldsymbol{\psi}(v)dv' = \boldsymbol{\psi}(u)\mathbf{C}^{\mathrm{T}}\mathbf{C}\boldsymbol{\psi}(v)$$

and

$$\int (\mathbf{A}^{\mathrm{T}}\mathbf{A})(u,v)(\boldsymbol{\psi}(v)\mathbf{y})dv = \int \boldsymbol{\psi}(u)\mathbf{C}^{\mathrm{T}}\mathbf{C}\boldsymbol{\psi}(v)\boldsymbol{\psi}(v)\mathbf{y}dv = \boldsymbol{\psi}(u)\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{y} = \lambda_1^2\boldsymbol{\psi}(u)\mathbf{y}.$$

Hence $||\mathbf{A}||_{\mathcal{L}} = \sqrt{\lambda_{\max}(\mathbf{A}^{\mathrm{T}}\mathbf{A})} = ||\mathbf{C}|| = \lambda_1$. The left side of Figure **??** plots $||\mathbf{A}||_{\mathcal{L}}$ vs b for different values of $a \in (0,1)$.

Let $(\omega_j, \mathbf{v}_j), j = 1, 2$, be the eigen-pairs of $\mathbf{C}$ satisfying $\mathbf{C}\mathbf{v}_j = \omega_j\mathbf{v}_j$. Then

$$\int_{\mathcal{U}} \boldsymbol{\psi}(u)\mathbf{C}\boldsymbol{\psi}(v)\boldsymbol{\psi}(v)\mathbf{v}_j dv = \int_{\mathcal{U}} \mathbf{A}(u,v)\boldsymbol{\psi}(v)\mathbf{v}_j dv = \omega_j\boldsymbol{\psi}(u)\mathbf{v}_j.$$

Hence $\mathbf{A}$ and $\mathbf{C}$ share the same eigenvalues, which are $\omega_1 = \omega_2 = a$. When $\rho(\mathbf{A}) = \rho(\mathbf{C}) = |a| < 1$, (4) and (D.45) correspond to stationary VFAR(1) and VAR(1) models, respectively.

For the VFAR(1) model in (4),the spectral density function and the covariance function of $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ are

$$f_{\mathbf{X},\theta} = \frac{1}{2\pi}\left(\boldsymbol{\Sigma}_0 + \sum_{h=1}^{\infty}\Big\{\boldsymbol{\Sigma}_0(\mathbf{A}^{\mathrm{T}})^h\exp(-ih\theta) + \mathbf{A}^h\boldsymbol{\Sigma}_0\exp(ih\theta)\Big\}\right) \tag{D.46}$$

and

$$\boldsymbol{\Sigma}_0 = \sigma^2\sum_{h=1}^{\infty}\mathbf{A}^h(\mathbf{A}^h)^{\mathrm{T}}, \tag{D.47}$$

respectively. For the VAR(1) model in (D.45), the spectral density matrix and the covariance matrix of $\{\mathbf{x}_t\}_{t\in\mathbb{Z}}$ are

$$f_{\mathbf{x},\boldsymbol{\theta}} = \frac{1}{2\pi}\left(\mathbf{S}_0 + \sum_{h=1}^{\infty}\Big\{\mathbf{S}_0(\mathbf{C}^{\mathrm{T}})^h\exp(-ih\theta) + \mathbf{C}^h\mathbf{S}_0\exp(ih\theta)\Big\}\right) \tag{D.48}$$

and

$$\mathbf{S}_0 = \sigma^2 \sum_{h=1}^{\infty} \mathbf{C}^h (\mathbf{C}^h)^{\mathrm{T}}, \tag{D.49}$$

respectively. Noting that $\mathbf{A}^h \boldsymbol{\Sigma}_0 = \int_{\mathcal{U}} \boldsymbol{\psi}(u)\mathbf{C}^h \boldsymbol{\psi}(v')\boldsymbol{\psi}(v')\mathbf{S}_0\boldsymbol{\psi}(v)dv' = \boldsymbol{\psi}(u)\mathbf{C}^h\mathbf{S}_0\boldsymbol{\psi}(v)$ and applying similar techniques, we can obtain that $f_{\mathbf{X},\theta} = \boldsymbol{\psi}(u) f_{\mathbf{x},\theta}\boldsymbol{\psi}(v)$ and $\boldsymbol{\Sigma}_0 = \boldsymbol{\psi}(u)\mathbf{S}_0\boldsymbol{\psi}(v)$. The functional stability measure of $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ under (4) is

$$2\pi \cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\Phi}\in\mathbb{H}_0} \frac{\boldsymbol{\Phi}^{\mathrm{T}} f_{\mathbf{X},\theta}\boldsymbol{\Phi}}{\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Sigma}_0\boldsymbol{\Phi}} = 2\pi \cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\Phi}\in\mathbb{H}_0} \frac{(\boldsymbol{\psi}\boldsymbol{\Phi})^{\mathrm{T}} f_{\mathbf{x},\theta}\boldsymbol{\psi}\boldsymbol{\Phi}}{(\boldsymbol{\psi}\boldsymbol{\Phi})^{\mathrm{T}}\mathbf{S}_0\boldsymbol{\psi}\boldsymbol{\Phi}},$$

where $\boldsymbol{\psi}\boldsymbol{\Phi} \in \mathbb{R}^2$ with $(\boldsymbol{\psi}\boldsymbol{\Phi})_j = \langle \boldsymbol{\phi}_j, \boldsymbol{\Phi}\rangle$, $j = 1, 2$. Hence the functional stability measure of $\{\mathbf{X}_t\}_{t\in\mathbb{Z}}$ under (4) is the same as that of $\{\mathbf{x}_t\}_{t\in\mathbb{Z}}$ under (D.45), i.e. the essential supremum of the maximal eigenvalue of $2\pi\mathbf{S}_0^{-1/2} f_{\mathbf{x},\theta}\mathbf{S}_0^{-1/2}$ over $\theta \in [-\pi, \pi]$. Some calculations yield $f_{\mathbf{x},\theta}$ and $\mathbf{S}_0$ as follows.

By (D.48), we have

$$\begin{aligned} f_{\mathbf{x},\boldsymbol{\theta}} &= \frac{1}{2\pi}\Bigg[\mathbf{S}_0 + \mathbf{S}_0 \begin{pmatrix} \sum_{h=1}^{\infty} a^h \exp(-ih\theta) & 0 \\ \sum_{h=1}^{\infty} h a^{h-1} \exp(-ih\theta) b & \sum_{h=1}^{\infty} a^h \exp(-ih\theta) \end{pmatrix} \\ &\quad + \begin{pmatrix} \sum_{h=1}^{\infty} a^h \exp(ih\theta) & \sum_{h=1}^{\infty} h a^{h-1} \exp(ih\theta) b \\ 0 & \sum_{h=1}^{\infty} a^h \exp(ih\theta) \end{pmatrix} \mathbf{S}_0\Bigg] \\ &= \frac{1}{2\pi}\Bigg[\mathbf{S}_0 + \mathbf{S}_0 \begin{pmatrix} \frac{\alpha\exp(-i\theta)}{1-a\exp(-i\theta)} & 0 \\ \frac{b\exp(-i\theta)}{(1-a\exp(-i\theta))^2} & \frac{a\exp(-i\theta)}{1-a\exp(-i\theta)} \end{pmatrix} + \begin{pmatrix} \frac{a\exp(i\theta)}{1-a\exp(i\theta)} & \frac{b\exp(i\theta)}{(1-a\exp(i\theta))^2} \\ 0 & \frac{a\exp(i\theta)}{1-a\exp(i\theta)} \end{pmatrix} \mathbf{S}_0\Bigg]. \end{aligned}$$

By (D.49), we have

$$\begin{aligned} \mathbf{S}_0 &= \begin{pmatrix} \sum_{h=0}^{\infty} a^{2h} + \sum_{h=0}^{\infty} h^2 a^{2h-2} b^2 & \sum_{h=0}^{\infty} h a^{2h-1} b \\ \sum_{h=0}^{\infty} h a^{2h-1} b & \sum_{h=0}^{\infty} a^{2h} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{1-a^2} + \frac{(a^2+1)b^2}{(1-a^2)^3} & \frac{ab}{(1-a^2)^2} \\ \frac{ab}{(1-a^2)^2} & \frac{1}{1-a^2} \end{pmatrix}. \end{aligned}$$

F. Derivations for VFAR models

F.1. Matrix representation of a VFAR(L) model in (20)

Note that the VFAR(L) model in (19) can be equivalently represented as

$$X_{tj}(u) = \sum_{h=1}^{L}\sum_{k=1}^{p} \langle A_{jk}^{(h)}(u,\cdot), X_{(t-h)k}(\cdot)\rangle + \varepsilon_{tj}(u), \quad t = L+1,\ldots,n, j = 1\ldots,p. \tag{E.50}$$

It then follows from the Karhunen-Loéve expansion that (E.50) can be rewritten as

$$\sum_{l=1}^{\infty} \xi_{tjl}\phi_{jl}(u) = \sum_{h=1}^{L}\sum_{k=1}^{p}\sum_{m=1}^{\infty} \langle A_{jk}^{(h)}(u,\cdot), \phi_{km}(\cdot)\rangle \xi_{(t-h)km} + \varepsilon_{tj}(u).$$

This, together with orthonormality of $\{\phi_{jm}(\cdot)\}_{m\geqslant 1}$, implies that

$$\xi_{tjl} = \sum_{h=1}^{L}\sum_{k=1}^{p}\sum_{m=1}^{q_k}\langle \phi_{jl}, A_{jk}^{(h)}(\phi_{km})\rangle\, \xi_{(t-h)km} + r_{tjl} + \epsilon_{tjl},$$

where $r_{tjl} = \sum_{h=1}^{L}\sum_{k=1}^{p}\sum_{m=q_k+1}^{\infty}\langle\phi_{jl}, A_{jk}^{(h)}(\phi_{km})\rangle\xi_{(t-h)km}$ and $\epsilon_{tjl} = \langle\phi_{jl}, \varepsilon_{tj}\rangle$ for $l = 1,\ldots,q_j$, represent the approximation and random errors, respectively. Let $\mathbf{r}_{tj} = (r_{tj1},\ldots,r_{tjq_j})^{\mathrm{T}}$ and $\boldsymbol{\epsilon}_{tj} = (\epsilon_{tj1},\ldots,\epsilon_{tjq_j})^{\mathrm{T}}$. Let $\mathbf{R}_j, \mathbf{E}_j$ be $(n-L)\times q_j$ matrices whose row vectors are formed by $\{\mathbf{r}_{tj}, t = L+1,\ldots,n\}$ and $\{\boldsymbol{\epsilon}_{tj}, t = L+1,\ldots,n\}$ respectively. Then (E.50) can be represented in the matrix form of (20).

F.2. VFAR(1) representation of a VFAR(L) model

We can represent a p-dimensional VFAR(L) model in (19) as a pL-dimensional VFAR(1) model in the form of

$$\widetilde{\mathbf{X}}_t(u) = \int_{\mathcal{U}} \widetilde{\mathbf{A}}_1(u,v)\widetilde{\mathbf{X}}_{t-1}(v)dv + \widetilde{\boldsymbol{\varepsilon}}_{t-1}(u),\ \ u\in\mathcal{U}, \tag{E.51}$$

where $\widetilde{\mathbf{X}}_t = \begin{pmatrix}\mathbf{X}_t\\ \mathbf{X}_{t-1}\\ \vdots\\ \mathbf{X}_{t-L+1}\end{pmatrix}$, $\widetilde{\mathbf{A}}_1 = \begin{pmatrix}\mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{L-1} & \mathbf{A}_L\\ \mathbf{I}_p & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_p & \cdots & \mathbf{0} & \mathbf{0}\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_p & \mathbf{0}\end{pmatrix}$, $\widetilde{\boldsymbol{\varepsilon}}_t = \begin{pmatrix}\boldsymbol{\varepsilon}_t\\ \boldsymbol{\varepsilon}_{t-1}\\ \vdots\\ \boldsymbol{\varepsilon}_{t-L+1}\end{pmatrix}$
and $\mathbf{I}_p$ denotes the identity operator. In the non-functional setting, a similar VAR(1) representation of a VAR(L) model can be found in Basu and Michailidis (2015).

F.3. VFAR(1) representation of the simulation example

Noting that $\boldsymbol{\theta}_t = \mathbf{B}\boldsymbol{\theta}_{t-1} + \boldsymbol{\eta}_t$, we have $\boldsymbol{\theta}_{tj} = \sum_{k=1}^{p}\mathbf{B}_{jk}\boldsymbol{\theta}_{(t-1)k} + \boldsymbol{\eta}_{tj}$ for $j = 1,\ldots,p$. Multiplying both sides by $\mathbf{s}(u)^{\mathrm{T}}$ and applying $\int_{\mathcal{U}}\mathbf{s}(v)\mathbf{s}(v)^{\mathrm{T}}dv = \mathbf{I}_5$, we obtain that $\mathbf{s}(u)^{\mathrm{T}}\boldsymbol{\theta}_{tj} = \int_{\mathcal{U}}\sum_{k=1}^{p}\mathbf{s}(u)^{\mathrm{T}}\mathbf{B}_{jk}\mathbf{s}(v)\mathbf{s}(v)^{\mathrm{T}}\boldsymbol{\theta}_{(t-1)k}dv + \mathbf{s}(u)^{\mathrm{T}}\boldsymbol{\eta}_{tj}$. Letting $A_{jk}(u,v) = \mathbf{s}(u)^{\mathrm{T}}\mathbf{B}_{jk}\mathbf{s}(v)$, $X_{tj}(u) = \mathbf{s}(u)^{\mathrm{T}}\boldsymbol{\theta}_{tj}$ and $\varepsilon_{tj}(u) = \mathbf{s}(u)^{\mathrm{T}}\boldsymbol{\eta}_{tj}$, we have $X_{tj}(u) = \sum_{k=1}^{p}\langle A_{jk}(u,\cdot)X_{(t-1)k}(\cdot)\rangle + \varepsilon_{tj}(u)$.

G. Algorithms in fitting VFAR models

G.1. Selection of tuning parameters

To fit the proposed sparse VFAR model, we need choose values for three tuning parameters, q_j (the number of selected principal components for $j = 1,\ldots,p$), η_j (the smoothing parameter when performing regularized FPCA, as described in Section G.2

below) and γ_{nj} (the regularization parameter in (21) to control the block sparsity level in $\{\widehat{\mathbf{\Psi}}_{jk}^{(h)} : h = 1, \ldots, L, k = 1, \ldots, p\}$).

We adopt a K-fold cross-validated method to choose (q_j, η_j) for each j. Specifically, let W_{tjs} be observed values of $X_{tj}(u_s)$ at $u_1, \ldots, u_T$. We randomly divide the set $\{1, \ldots, n\}$ into K groups, $\mathcal{D}_1, \ldots, \mathcal{D}_K$ of approximately equal size, with the first group treated as a validation set. Implementing regularized FPCA on the remaining $K-1$ groups, we obtain estimated mean function $\widehat{\mu}_{jl}^{(-1)}(u)$, FPC scores $\widehat{\xi}_{tjl,\eta_j}^{(-1)}$ and eigenfunctions $\widehat{\phi}_{jl}^{(-1)}(u; \eta_j)$ for $l = 1, \ldots, q_j$. The predicted curve for the t-th sample in group one can be computed by $\widehat{W}_{tjs}^{(1)} = \widehat{\mu}_{jl}^{(-1)}(u_s) + \sum_{l=1}^{q_j} \widehat{\xi}_{tjl,\eta_j}^{(-1)} \widehat{\phi}_{jl}^{(-1)}(u_s; \eta_j)$. This procedure is repeated K times. Finally, we choose q_j and η_j as the values that minimize the mean cross-validated error,

$$\mathrm{CV}(q_j, \eta_j) = (KT)^{-1} \sum_{k=1}^{K} \sum_{s=1}^{T} \sum_{t \in \mathcal{D}_k} (W_{tjs} - \widehat{W}_{tjs}^{(k)})^2.$$

The optimal γ_{nj}'s are selected by minimizing AICs or BICs in (29), where details can be found in Section 3.4.

G.2. Regularized FPCA

In this section, we drop subscripts j for simplicity of notation. Suppose we observe $\mathbf{X}(\cdot) = (X_1(\cdot), \ldots, X_n(\cdot))^{\mathrm{T}}$ on $\mathcal{U}$, our goal is to find the first q regularized principal component functions $\{\phi_l(\cdot), l = 1, \ldots, q\}$. We obtain the l-th leading principal component $\phi_l(\cdot)$ through a smoothing approach, which maximizes the following penalized sample variance [(9.1) in Ramsay and Silverman (2005)]

$$\mathrm{PEN}_\eta(\phi_l) = \frac{\mathrm{Var}(\langle \phi_l, X_i \rangle)}{||\phi_l||^2 + \eta ||\phi_l''||^2}, \tag{F.52}$$

subject to $||\phi_l|| = 1$ and $\langle \phi_l, \phi_{l'} \rangle + \eta \langle \phi_l'', \phi_{l'}'' \rangle = 0, l' = 1, \ldots, l-1$, where $\eta \geqslant 0$ is a smoothing parameter to control the roughness of $\phi_l(\cdot)$.

Suppose that $\mathbf{X}(u) = \boldsymbol{\delta}^{\mathrm{T}} \mathbf{b}(u)$ and $\phi_l(u) = \boldsymbol{\zeta}_l^{\mathrm{T}} \mathbf{b}(u)$ where $\mathbf{b}(\cdot)$ is a G-dimensional B-spline basis function, $\boldsymbol{\delta} \in \mathbb{R}^{n \times G}$ and $\boldsymbol{\zeta}_l \in \mathbb{R}^G$ are the basis coefficients for $\mathbf{X}(\cdot)$ and $\phi_l(\cdot)$, respectively. Let $\mathbf{J} = \int \mathbf{b}(u)\mathbf{b}(u)^{\mathrm{T}} du$, $\mathbf{U} = \mathbf{J}\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\delta}\mathbf{J}$ and $\mathbf{Q} = \int \mathbf{b}''(u)\mathbf{b}''(u)^{\mathrm{T}} du$, (F.52) is equivalent to maximizing

$$\mathrm{PEN}_\eta(\phi_l) = \frac{\boldsymbol{\zeta}_l^{\mathrm{T}} \mathbf{U} \boldsymbol{\zeta}_l}{\boldsymbol{\zeta}_l^{\mathrm{T}} (\mathbf{J} + \eta \mathbf{Q}) \boldsymbol{\zeta}_l}, \tag{F.53}$$

subject to $\boldsymbol{\zeta}_l^{\mathrm{T}} \mathbf{J} \boldsymbol{\zeta}_l = 1$ and $\boldsymbol{\zeta}_l^{\mathrm{T}} (\mathbf{J} + \eta \mathbf{Q}) \boldsymbol{\zeta}_{l'} = 0, l' = 1, \ldots, l-1$. By singular value decomposition (SVD), we obtain eigen-pairs, $(\mathbf{S}_1, \mathbf{P}_1)$ and $(\mathbf{S}_2, \mathbf{P}_2)$ such that $\mathbf{J} + \eta \mathbf{Q} = \mathbf{P}_1 \mathbf{S}_1^{-2} \mathbf{P}_1^{\mathrm{T}}$ and $\mathbf{S}_1 \mathbf{P}_1^{\mathrm{T}} \mathbf{U} \mathbf{P}_1 \mathbf{S}_1 = \mathbf{P}_2 \mathbf{S}_2^{-2} \mathbf{P}_2^{\mathrm{T}}$. Then (F.53) becomes $\mathrm{PEN}_\eta(\phi_l) = \frac{\mathbf{x}_l^{\mathrm{T}} \mathbf{P}_2^{\mathrm{T}} \mathbf{S}_2 \mathbf{P}_2 \mathbf{x}_l}{\mathbf{x}_l^{\mathrm{T}} \mathbf{x}_l}$, where $\mathbf{x}_l = \mathbf{S}_1^{-1} \mathbf{P}_1^{\mathrm{T}} \boldsymbol{\zeta}_l$. This suggests us to perform SVD on $\mathbf{P}_2^{\mathrm{T}} \mathbf{S}_2 \mathbf{P}_2$, where we can obtain $\widehat{\mathbf{x}}_l$, $\widehat{\boldsymbol{\zeta}}_l = \mathbf{P}_1 \mathbf{S}_1 \widehat{\mathbf{x}}_l$ and $\widehat{\phi}_l(u) = \widehat{\boldsymbol{\zeta}}_l^{\mathrm{T}} \mathbf{b}(u) / \big(\widehat{\boldsymbol{\zeta}}_l^{\mathrm{T}} \mathbf{J} \widehat{\boldsymbol{\zeta}}_l\big)^{1/2}, l = 1, \ldots, q$. In practice, we can set G to a pre-specified large enough value, and implement the cross-validation procedure as described in Section G.1 to select q and η.

G.3. Block FISTA algorithm to solve (21)

The optimization problem in (21) can be reformulated as follows.

$$\min_{\mathbf{X}\in\mathbb{R}^{r\times q_j}} g(\mathbf{X}), \quad g(\mathbf{X}) = f(\mathbf{X}) + \gamma_{nj}\sum_{k=1}^{pL}\|\mathbf{X}_k\|_F, \tag{F.54}$$

where $f(\mathbf{X}) = 2^{-1}\text{trace}\{(\mathbf{Y}-\mathbf{B}\mathbf{X})^{\mathrm{T}}(\mathbf{Y}-\mathbf{B}\mathbf{X})\}$, $r = \sum_{h=1}^{L}\sum_{k=1}^{p} q_k$, $\mathbf{Y}\in\mathbb{R}^{(n-L)\times q_j}$, $\mathbf{B}\in\mathbb{R}^{(n-L)\times r}$, and $\mathbf{X} = (\mathbf{X}_1^{\mathrm{T}},\ldots,\mathbf{X}_{pL}^{\mathrm{T}})^{\mathrm{T}}\in\mathbb{R}^{r\times q_j}$ with $\mathbf{X}_k\in\mathbb{R}^{q_k\times q_j}$ for $k=1,\ldots,p$. (F.54) is a convex problem including the smooth part for $\mathbf{X}_k$, i.e. $f(\mathbf{X})$ and the non-smooth part for $\mathbf{X}_k$, i.e.$\gamma_{nj}\sum_{k=1}^{pL}\|\mathbf{X}_k\|_F$. To solve the minimization problem in (F.54), we adopt a block version of fast iterative shrinkage-thresholding algorithm (FISTA) (Beck and Teboulle, 2009) combined with a restarting technique (O'Donoghue and Candès, 2015), namely block FISTA.

The basic idea behind our proposed block FISTA is summarized as follows. Let $\nabla f(\mathbf{X})$ be the gradient of $f(\mathbf{X})$ at $\mathbf{X}$. We start with an initial value $\mathbf{X}^{(0)}$. At the $(m+1)$-th iteration we first try to solve a regularized sub-problem

$$\min_{\mathbf{X}\in\mathbb{R}^{r\times q_j}} \text{trace}\left\{(\nabla f(\mathbf{X}^{(m)})^{\mathrm{T}}(\mathbf{X}-\mathbf{X}^{(m)})\right\} + (2C)^{-1}\left\|\mathbf{X}-\mathbf{X}^{(m)}\right\|_F^2 + \gamma_{nj}\sum_{k=1}^{pL}\|\mathbf{X}_k\|_F, \tag{F.55}$$

where $\mathbf{X}^{(m)}$ is the m-th iterate and $C>0$ is a small constant controlling the stepsize at $(m+1)$-th step. The second term in (F.55) can be interpreted as a quadratic regularization, which restricts the updated iterate not to be very far from $\mathbf{X}^{(m)}$. The analytical solution to (F.55) takes the form of

$$\widetilde{\mathbf{X}}^{(m+1)} = \left(\widetilde{\mathbf{X}}_k^{(m+1)}\right) \text{ with } \widetilde{\mathbf{X}}_k^{(m+1)} = \left(1-\gamma_{nj}C\|\mathbf{Z}_k^{(m)}\|_F^{-1}\right)_+ \mathbf{Z}_k^{(m)}, k=1,\ldots,pL, \tag{F.56}$$

where $\mathbf{Z}^{(m)} = \mathbf{X}^{(m)} - C\nabla f(\mathbf{X}^{(m)}) = \left((\mathbf{Z}_1^{(m)})^{\mathrm{T}},\ldots,(\mathbf{Z}_{pL}^{(m)})^{\mathrm{T}}\right)^{\mathrm{T}}$ and $x_+ = \max(0,x)$. (See also (3.a) and (3.b) of Algorithm 1).

We then take block FISTA (Beck and Teboulle, 2009) by adding an extrapolation step in the algorithm (see also (3.c) and (3.d) of Algorithm 1):

$$\mathbf{X}^{(m+1)} = \widetilde{\mathbf{X}}^{(m+1)} + \omega^{(m+1)}(\widetilde{\mathbf{X}}^{(m+1)} - \widetilde{\mathbf{X}}^{(m)}),$$

where the weight $\omega^{(m+1)}$ is specified in Algorithm 1. Finally, at the end of each iteration, we evaluate the generalized gradient at $\mathbf{X}^{(m+1)}$ by computing the sign of

$$\text{trace}\left\{(\mathbf{X}^{(m)} - \widetilde{\mathbf{X}}^{(m+1)})^{\mathrm{T}}(\widetilde{\mathbf{X}}^{(m+1)} - \widetilde{\mathbf{X}}^{(m)})\right\},$$

which can be thought of a proxy of trace$\left\{(\nabla g(\mathbf{X}^{(m)}))^{\mathrm{T}}(\widetilde{\mathbf{X}}^{(m+1)} - \widetilde{\mathbf{X}}^{(m)})\right\}$. For a positive sign, i.e. the objective function is increasing at $\widetilde{\mathbf{X}}^{(m+1)}$, we then restart our accelerated

Algorithm 1 Block FISTA for solving (F.54)

1. Input: $C = 0.9\big(\lambda_{\max}(\mathbf{B}^{\mathrm{T}}\mathbf{B})\big)^{-1}$, $\theta_0 = 1$, $\mathbf{X}^{(0)} = (\mathbf{X}_1^{(0)\mathrm{T}}, \dots, \mathbf{X}_{pL}^{(0)\mathrm{T}})^{\mathrm{T}} = \mathbf{0}$, $\mathbf{Z}^{(0)} = (\mathbf{Z}_1^{(0)\mathrm{T}}, \dots, \mathbf{Z}_{pL}^{(0)\mathrm{T}})^{\mathrm{T}} = \mathbf{0}$, $\widetilde{\mathbf{X}}^{(0)} = (\widetilde{\mathbf{X}}_1^{(0)\mathrm{T}}, \dots, \widetilde{\mathbf{X}}_{pL}^{(0)\mathrm{T}})^{\mathrm{T}} = \mathbf{0}$.
2. For $m = 0, 1, \dots$ do

   (3.a) $\mathbf{Z}^{(m)} = \mathbf{X}^{(m)} - C\nabla f(\mathbf{X}^{(m)})$,

   (3.b) $\widetilde{\mathbf{X}}_k^{(m+1)} = \Big(1 - \gamma_{nj} C \|\mathbf{Z}_k^{(m)}\|_F^{-1}\Big)_+ \mathbf{Z}_k^{(m)}, k = 1, \dots, pL$,

   (3.c) $\theta_{m+1} = \big(1 + \sqrt{1 + 4\theta_m^2}\big)/2$,

   (3.d) $\mathbf{X}^{(m+1)} = \widetilde{\mathbf{X}}^{(m+1)} + \frac{\theta_m - 1}{\theta_{m+1}}\big(\widetilde{\mathbf{X}}^{(m+1)} - \widetilde{\mathbf{X}}^{(m)}\big)$,

   (3.e) If trace $\Big\{(\mathbf{X}^{(m)} - \widetilde{\mathbf{X}}^{(m+1)})^{\mathrm{T}}(\widetilde{\mathbf{X}}^{(m+1)} - \widetilde{\mathbf{X}}^{(m)})\Big\} > 0$, set

   $$\mathbf{X}^{(m+1)} = \mathbf{X}^{(m)}, \theta_{m+1} = 1.$$

   end do until convergence.
3. Output: the final estimator $\mathbf{X}^{(m+1)}$.

algorithm by setting $\mathbf{X}^{(m+1)} = \mathbf{X}^{(m)}$ and $\omega^{(m+1)} = \omega^{(1)}$ (O'Donoghue and Candès, 2015). This step can guarantee that the objective function g decreases over each iteration. We iterative the above steps until convergence. We summarize the restarting-based block FISTA in Algorithm 1. In practice, one issue is how to choose the stepsize parameter C. In general, the proposed scheme is guaranteed to converge when $C < \big(\lambda_{\max}(\mathbf{B}^{\mathrm{T}}\mathbf{B})\big)^{-1}$. Here we choose $C = 0.9\big(\lambda_{\max}(\mathbf{B}^{\mathrm{T}}\mathbf{B})\big)^{-1}$, which turns to work well in empirical studies. Alternatively, C can be selected through a line search and one simple backtracking rule.

H. Additional empirical results

H.1. Simulation studies

Figures 1 and 2 of the Supplementary Material plot the median best ROC curves (we rank ROC curves by the corresponding AUROCs) over the 100 stimulation runs in Models (i) and (ii), respectively. Again we see that ℓ_1/ℓ_2-LS$_2$, which explains the partial curve information, although performing better than ℓ_1-LS$_1$ is substantially outperformed by ℓ_1/ℓ_2-LS$_\mathrm{a}$ in terms of model selection consistency.

H.2. Real data analysis

Table 1 of the Supplementary Material provides tickers, company names and classified sectors of 98 stocks under our study. Figure 3 of the Supplementary Material plots the sparsity patterns in $\widehat{\mathbf{A}}$ (estimated transition function) for either 18 or 36 stocks. For a

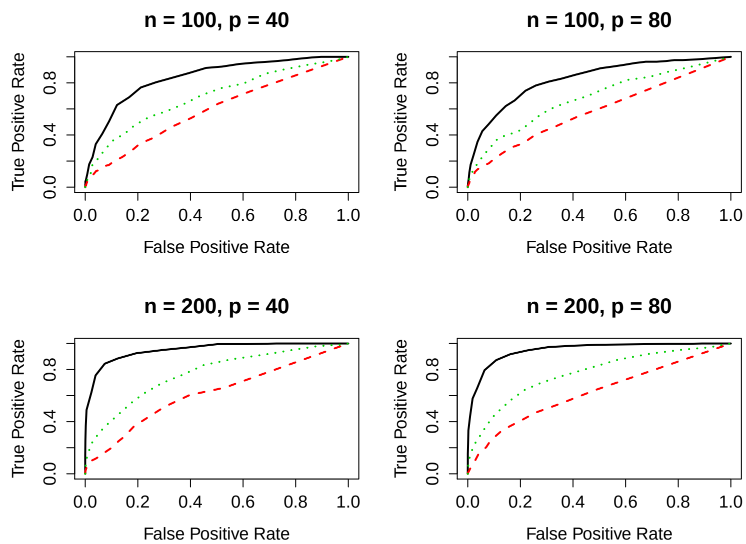

Figure 1. Comparisons of median estimated ROC curves over 100 simulation runs. ℓ_1/ℓ_2-LS$_{\mathrm{a}}$ (black solid), ℓ_1/ℓ_2-LS$_2$ (green dotted) and ℓ_1-LS$_1$ (red dashed) for Model (i).

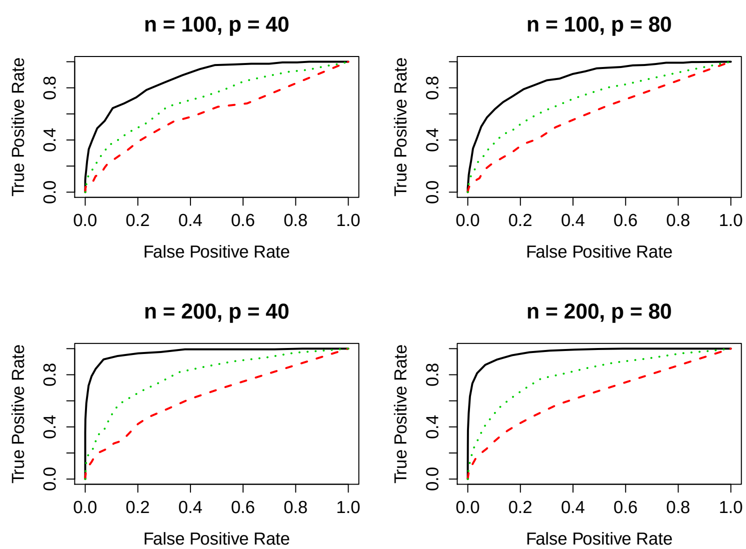

Figure 2. Comparisons of median estimated ROC curves over 100 simulation runs. ℓ_1/ℓ_2-LS$_{\mathrm{a}}$ (black solid), ℓ_1/ℓ_2-LS$_2$ (green dotted) and ℓ_1-LS$_1$ (red dashed) for Model (ii).

large collection of $p = 98$ S&P100 stocks, to better visualize a large causal network, we set the row-wise sparsity to 3/98. We then plot a large and sparse directed graph in Figure 4 of the Supplementary Material. We observe that companies, e.g. Allergan, Halliburton, Target Corp., have relatively higher causal impacts on all 98 stocks in terms of their CIDR curves.

Table 1. List of S&P 100 stocks under study.

Ticker	Company name	Sector	Ticker	Company name	Sector
AAPL	APPLE INC	Information Technology	JPM	JPMORGAN CHASE & CO	Financials
ABBV	ABBVIE INC	Health Care	KHC	KRAFT HEINZ	Consumer Staples
ABT	ABBOTT LABORATORIES	Health Care	KMI	KINDER MORGAN INC	Energy
ACN	ACCENTURE PLC CLASS A	Information Technology	KO	COCA-COLA	Consumer Staples
AGN	ALLERGAN	Health Care	LLY	ELI LILLY	Health Care
AIG	AMERICAN INTERNATIONAL GROUP INC	Financials	LMT	LOCKHEED MARTIN CORP	Industrials
ALL	ALLSTATE CORP	Financials	LOW	LOWES COMPANIES INC	Consumer Discretionary
AMGN	AMGEN INC	Health Care	MA	MASTERCARD INC CLASS A	Information Technology
AMZN	AMAZON COM INC	Consumer Discretionary	MCD	MCDONALDS CORP	Consumer Discretionary
AXP	AMERICAN EXPRESS	Financials	MDLZ	MONDELEZ INTERNATIONAL INC CLASS A	Consumer Staples
BA	BOEING	Industrials	MDT	MEDTRONIC PLC	Health Care
BAC	BANK OF AMERICA CORP	Financials	MET	METLIFE INC	Financials
BIIB	BIOGEN INC INC	Health Care	MMM	3M	Industrials
BK	BANK OF NEW YORK MELLON CORP	Financials	MO	ALTRIA GROUP INC	Consumer Staples
BLK	BLACKROCK INC	Financials	MON	MONSANTO	Materials
BMY	BRISTOL MYERS SQUIBB	Health Care	MRK	MERCK & CO INC	Health Care
C	CITIGROUP INC	Financials	MS	MORGAN STANLEY	Financials
CAT	CATERPILLAR INC	Industrials	MSFT	MICROSOFT CORP	Information Technology
CELG	CELGENE CORP	Health Care	NEE	NEXTERA ENERGY INC	Utilities
CHTR	CHARTER COMMUNICATIONS INC CLASS A	Consumer Discretionary	NKE	NIKE INC CLASS B	Consumer Discretionary
CL	COLGATE-PALMOLIVE	Consumer Staples	ORCL	ORACLE CORP	Information Technology
COF	CAPITAL ONE FINANCIAL CORP	Financials	OXY	OCCIDENTAL PETROLEUM CORP	Energy
COP	CONOCOPHILLIPS	Energy	PCLN	THE PRICELINE GROUP INC	Consumer Discretionary
COST	COSTCO WHOLESALE CORP	Consumer Staples	PEP	PEPSICO INC	Consumer Staples
CSCO	CISCO SYSTEMS INC	Information Technology	PFE	PFIZER INC	Health Care
CVS	CVS HEALTH CORP	Consumer Staples	PG	PROCTER & GAMBLE	Consumer Staples
CVX	CHEVRON CORP	Energy	PM	PHILIP MORRIS INTERNATIONAL INC	Consumer Staples
DHR	DANAHER CORP	Health Care	PYPL	PAYPAL HOLDINGS INC	Information Technology
DIS	WALT DISNEY	Consumer Discretionary	QCOM	QUALCOMM INC	Information Technology
DUK	DUKE ENERGY CORP	Utilities	RTN	RAYTHEON	Industrials
EMR	EMERSON ELECTRIC	Industrials	SBUX	STARBUCKS CORP	Consumer Discretionary
EXC	EXELON CORP	Utilities	SLB	SCHLUMBERGER NV	Energy
F	F MOTOR	Consumer Discretionary	SO	SOUTHERN	Utilities
FB	FACEBOOK CLASS A INC	Information Technology	SPG	SIMON PROPERTY GROUP REIT INC	Real Estate
FDX	FEDEX CORP	Industrials	T	AT&T INC	Telecommunications
FOX	TWENTY-FIRST CENTURY FOX INC CLASS	Consumer Discretionary	TGT	TARGET CORP	Consumer Discretionary
FOX	TWENTY FIRST CENTURY FOX INC CLASS	Consumer Discretionary	TWX	TIME WARNER INC	Consumer Discretionary
GD	GENERAL DYNAMICS CORP	Industrials	TXN	TEXAS INSTRUMENT INC	Information Technology
GE	GENERAL ELECTRIC	Industrials	UNH	UNITEDHEALTH GROUP INC	Health Care
GILD	GILEAD SCIENCES INC	Health Care	UNP	UNION PACIFIC CORP	Industrials
GM	GENERAL MOTORS	Consumer Discretionary	UPS	UNITED PARCEL SERVICE INC CLASS B	Industrials
GOOG	ALPHABET INC CLASS C	Information Technology	USB	US BANCORP	Financials
GS	GOLDMAN SACHS GROUP INC	Financials	UTX	UNITED TECHNOLOGIES CORP	Industrials
HAL	HALLIBURTON	Energy	V	VISA INC CLASS A	Information Technology
HD	HOME DEPOT INC	Consumer Discretionary	VZ	VERIZON COMMUNICATIONS INC	Telecommunications
HON	HONEYWELL INTERNATIONAL INC	Industrials	WBA	WALGREEN BOOTS ALLIANCE INC	Consumer Staples
IBM	INTERNATIONAL BUSINESS MACHINES CO	Information Technology	WFC	WELLS FARGO	Financials
INTC	INTEL CORPORATION CORP	Information Technology	WMT	WALMART STORES INC	Consumer Staples
JNJ	JOHNSON & JOHNSON	Health Care	XOM	EXXON MOBIL CORP	Energy

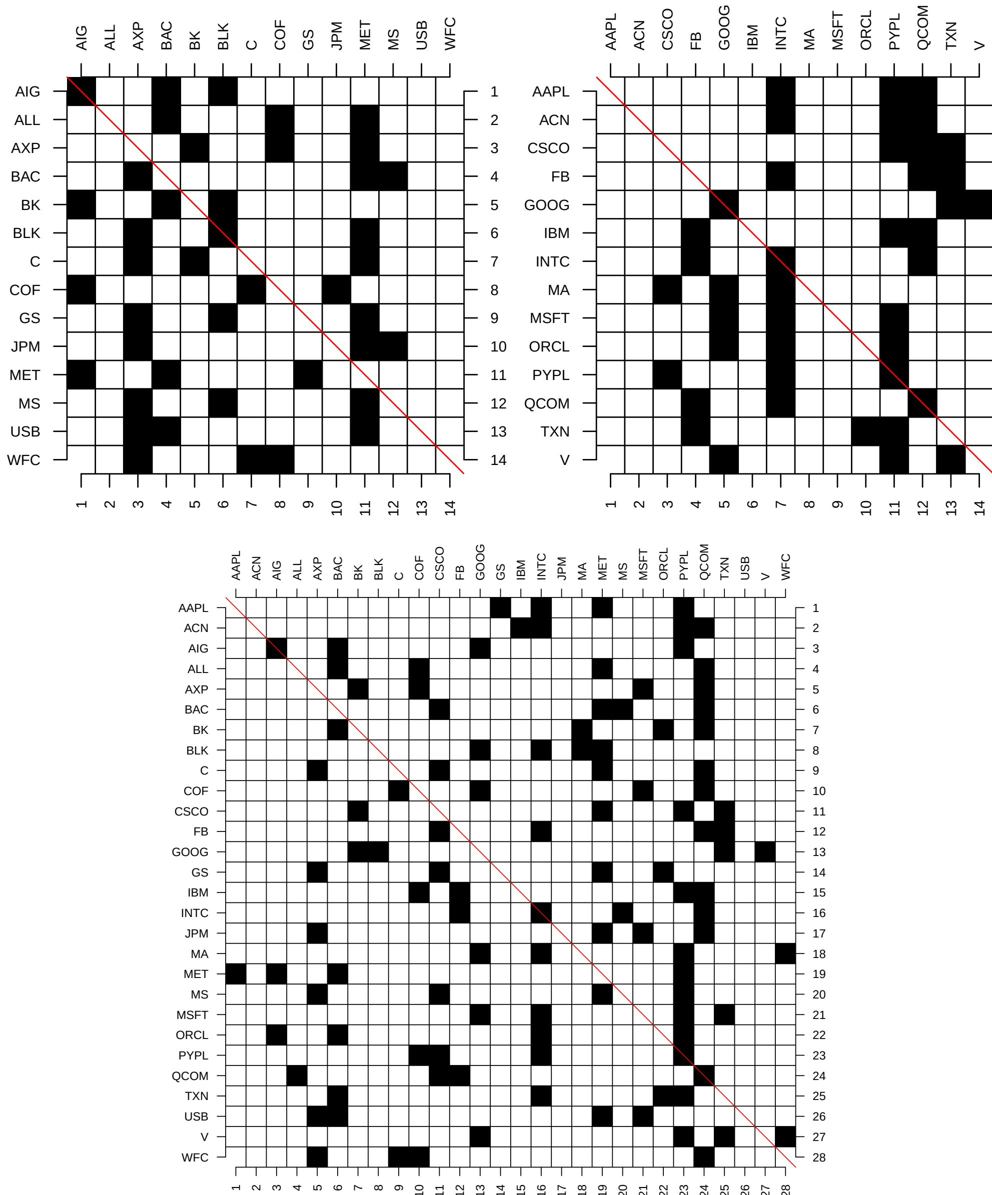

Figure 3. *Topleft, topright and bottom graphs plot the sparsity structures in* $\widehat{\mathbf{A}}$ *for stocks in only the financial sector, only the IT sector and both financial and IT sectors, respectively. Black and white correspond to non-zero and zero functional components in* $\widehat{\mathbf{A}}$, *respectively.*

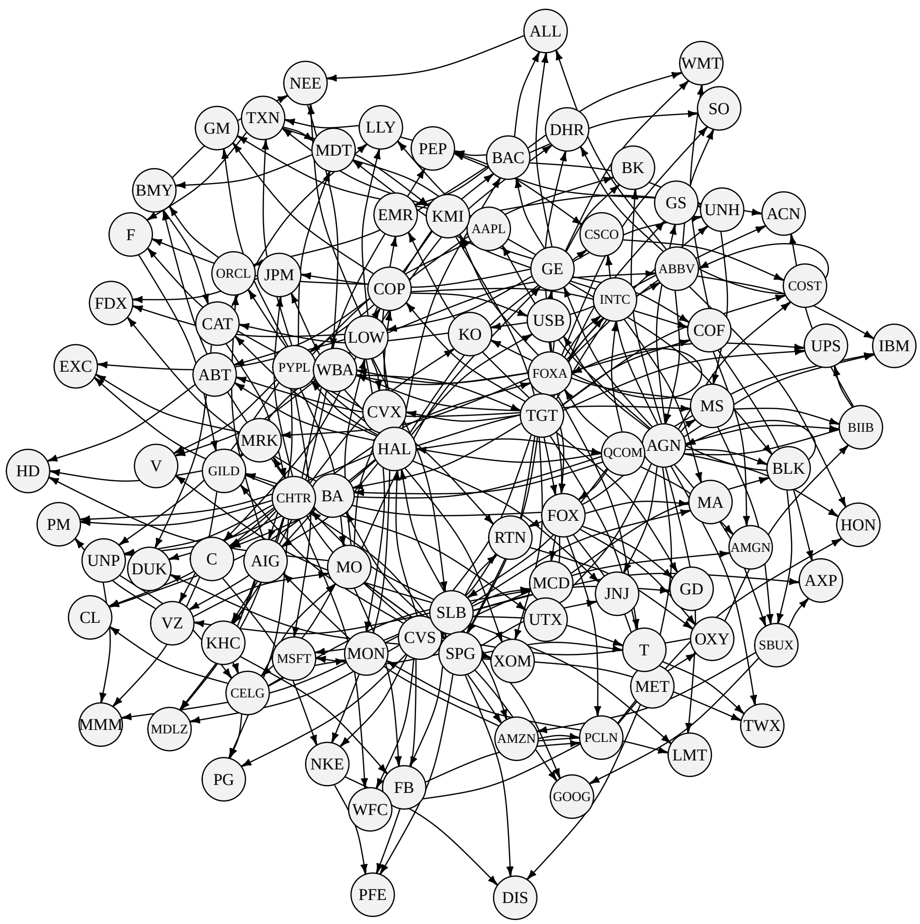

Figure 4. *The directed graph with indegree=3 for $p = 98$ stocks.*

References

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70** 191–221. MR1926259

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models. *Ann. Statist.* **43** 1535–1567. MR3357870

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2** 183–202. MR2486527

Bosq, D. (2000). *Linear processes in function spaces*. *Lecture Notes in Statistics* **149**. Springer-Verlag, New York Theory and applications. MR1783138

Boucheron, S., Lugosi, G. and Massart, P. (2014). *Concentation inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 603–680. MR3091653

Fang, Q., Guo, S. and Qiao, X. (2021). Finite sample theory for high-dimensional functional/scalar time series with applications. *arXiv:2004.07781v2*.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91. MR2332269

Hörmann, S., Kidziński, L. u. and Hallin, M. (2015). Dynamic functional principal components. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 319–348. MR3310529

Lam, C. and Yao, Q. (2012). Factor modeling for high-dimensional time series: inference for the number of factors. *Ann. Statist.* **40** 694–726. MR2933663

O'Donoghue, B. and Candès, E. (2015). Adaptive restart for accelerated gradient schemes. *Found. Comput. Math.* **15** 715–732. MR3348171

Qiao, X., Qian, C., James, G. M. and Guo, S. (2020). Doubly functional graphical models in high dimensions. *Biometrika* **107** 415–431. MR4108937

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*, second ed. *Springer Series in Statistics*. Springer, New York. MR2168993

Rudelson, M. and Vershynin, R. (2013). Hanson-Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** no. 82, 9. MR3125258

Wong, K. C., Li, Z. and Tewari, A. (2020). Lasso guarantees for β-mixing heavy-tailed time series. *Ann. Statist.* **48** 1124–1142. MR4102690