Factor-guided estimation of large covariance matrix function with conditional functional sparsity

Dong Li[1], Xinghao Qiao[2], and Zihan Wang[1]

[1] *Center for Statistical Science, Tsinghua University, Beijing 100084, China*
[2] *Department of Statistics, London School of Economics, London WC2A 2AE, U.K.*

Abstract

This paper addresses the fundamental task of estimating covariance matrix functions for high-dimensional functional data/functional time series. We consider two functional factor structures encompassing either functional factors with scalar loadings or scalar factors with functional loadings, and postulate functional sparsity on the covariance of idiosyncratic errors after taking out the common unobserved factors. To facilitate estimation, we rely on the spiked matrix model and its functional generalization, and derive some novel asymptotic identifiability results, based on which we develop DIGIT and FPOET estimators under two functional factor models, respectively. Both estimators involve performing associated eigenanalysis to estimate the covariance of common components, followed by adaptive functional thresholding applied to the residual covariance. We also develop functional information criteria for the purpose of model selection. The convergence rates of estimated factors, loadings, and conditional sparse covariance matrix functions under various functional matrix norms, are respectively established for DIGIT and FPOET estimators. Numerical studies including extensive simulations and two real data applications on mortality rates and functional portfolio allocation are conducted to examine the finite-sample performance of the proposed methodology.

Keywords: Adaptive functional thresholding; Asymptotic identifiablity; Eigenanalysis; Functional factor model; High-dimensional functional data/functional time series; Model selection.

1 Introduction

With advancements in data collection technology, multivariate functional data/functional time series are emerging in a wide range of scientific and economic applications. Examples include different types of brain imaging data in neuroscience, intraday return trajectories for a collection of stocks, age-specific mortality rates across different countries, and daily energy consumption curves from thousands of households, among others. Such data can be represented as $\mathbf{y}_t(\cdot) = \{y_{t1}(\cdot), \ldots, y_{tp}(\cdot)\}^{\mathrm{T}}$ defined on a compact interval $\mathcal{U}$, with the marginal- and cross-covariance operators induced from the associated kernel functions. These operators together form the operator-valued covariance matrix, which is also referred to as the following covariance matrix function for notational simplicity:

$$\boldsymbol{\Sigma}_y = \{\Sigma_{y,jk}(\cdot,\cdot)\}_{p\times p}, \quad \Sigma_{y,jk}(u,v) = \mathrm{Cov}\{y_{tj}(u), y_{tk}(v)\}, \quad (u,v) \in \mathcal{U}^2,$$

and we observe stationary $\mathbf{y}_t(\cdot)$ for $t = 1, \ldots, n$.

The estimation of covariance matrix function and its inverse is of paramount importance in multivariate functional data/functional time series analysis. An estimator of $\boldsymbol{\Sigma}_y$ is not only of interest in its own right but also essential for subsequent analyses, such as dimension reduction and modeling of $\{\mathbf{y}_t(\cdot)\}$. Examples include multivariate functional principal components analysis (MFPCA) (Happ and Greven, 2018), functional risk management to account for intraday uncertainties, functional graphical model estimation (Qiao et al., 2019), multivariate functional linear regression (Chiou et al., 2016) and functional linear discriminant analysis (Xue et al., 2023). See Section 4 for details of these applications. In increasingly available high-dimensional settings where the dimension p diverges with, or is larger than, the number of independent or serially dependent observations n, the sample covariance matrix function $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}$ performs poorly and some regularization is needed. Fang et al. (2023) pioneered this effort by assuming approximate functional sparsity in $\boldsymbol{\Sigma}_y$, where the Hilbert–Schmidt norms of some $\Sigma_{y,jk}$'s are assumed zero or close to zero. Then they applied adaptive functional thresholding to the entries of $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}$ to achieve a consistent estimator of $\boldsymbol{\Sigma}_y$.

Such functional sparsity assumption, however, is restrictive or even unrealistic for many datasets, particularly in finance and economics, where variables often exhibit high correlations. E.g., in the stock market, the co-movement of intraday return curves (Horváth et al., 2014) is typically influenced by a small number of common market factors, leading to highly correlated functional variables. To alleviate the direct imposition of sparsity assumption, we employ the functional factor model (FFM) framework for $\mathbf{y}_t(\cdot)$, which decomposes it into two uncorrelated components, one common $\boldsymbol{\chi}_t(\cdot)$ driven by low-dimensional latent factors and one idiosyncratic $\boldsymbol{\varepsilon}_t(\cdot)$. We consider two types of FFM. The first type, explored in Guo et al. (2022), admits the representation with functional factors and scalar loadings:

$$\mathbf{y}_t(\cdot) = \boldsymbol{\chi}_t(\cdot) + \boldsymbol{\varepsilon}_t(\cdot) = \mathbf{B}\mathbf{f}_t(\cdot) + \boldsymbol{\varepsilon}_t(\cdot), \quad t = 1, \ldots, n, \tag{1}$$

where $\mathbf{f}_t(\cdot)$ is a r-vector of stationary latent functional factors, $\mathbf{B}$ is a $p \times r$ matrix of factor loadings and $\boldsymbol{\varepsilon}_t(\cdot)$ is a p-vector of idiosyncratic errors. The second type, introduced by Hallin et al. (2023), involves scalar factors and functional loadings:

$$\mathbf{y}_t(\cdot) = \boldsymbol{\chi}_t(\cdot) + \boldsymbol{\varepsilon}_t(\cdot) = \mathbf{Q}(\cdot)\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t(\cdot), \quad t = 1, \ldots, n, \tag{2}$$

where $\boldsymbol{\gamma}_t$ is a r-vector of stationary latent factors and $\mathbf{Q}(\cdot)$ is a $p \times r$ matrix of functional factor loadings. We refer to $\boldsymbol{\Sigma}_f$, $\boldsymbol{\Sigma}_\chi$ and $\boldsymbol{\Sigma}_\varepsilon$ as the covariance matrix functions of $\mathbf{f}_t$, $\boldsymbol{\chi}_t$ and $\boldsymbol{\varepsilon}_t$, respectively.

Within the FFM framework, our goal is to estimate the covariance matrix function $\boldsymbol{\Sigma}_y = \boldsymbol{\Sigma}_\chi + \boldsymbol{\Sigma}_\varepsilon$. Inspired by Fan et al. (2013), we impose the approximately functional sparsity assumption on $\boldsymbol{\Sigma}_\varepsilon$ instead of $\boldsymbol{\Sigma}_y$ directly giving rise to the conditional functional sparsity structure in models (1) and (2). To effectively separate $\boldsymbol{\chi}_t(\cdot)$ from $\boldsymbol{\varepsilon}_t(\cdot)$, we rely on the spiked matrix model (Wang and Fan, 2017) and its functional generalization, i.e. a large nonnegative definite matrix or operator-valued matrix $\boldsymbol{\Lambda} = \mathbf{L} + \mathbf{S}$, where $\mathbf{L}$ is low rank and its nonzero eigenvalues grow fast as p diverges, whereas all the eigenvalues of $\mathbf{S}$ are bounded or grow much slower. The spikeness pattern ensures that the large signals are concentrated on $\mathbf{L}$, which facilitates our estimation procedure. Specifically, for model (2), with the decomposition

$$\underbrace{\boldsymbol{\Sigma}_y(\cdot,*)}_{\boldsymbol{\Lambda}} = \underbrace{\mathbf{Q}(\cdot)\mathrm{Cov}(\boldsymbol{\gamma}_t)\mathbf{Q}(*)^{\mathrm{T}}}_{\mathbf{L}} + \underbrace{\boldsymbol{\Sigma}_\varepsilon(\cdot,*)}_{\mathbf{S}}, \tag{3}$$

we perform MFPCA based on $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}$, then estimate $\boldsymbol{\Sigma}_\chi$ using the leading r functional principal components and finally propose a novel adaptive functional thresholding procedure to estimate the sparse $\boldsymbol{\Sigma}_\varepsilon$. This results in a Functional Principal Orthogonal complEment Thresholding (FPOET) estimator, extending the POET methodology for large covariance matrix estimation (Fan et al., 2013; 2018; Wang et al., 2021) to the functional domain. Alternatively, for model (1), considering the violation of nonnegative definiteness in $\boldsymbol{\Sigma}_y(u,v)$ for $u \neq v$, we utilize the nonnegative definite doubly integrated Gram covariance:

$$\underbrace{\int\int \boldsymbol{\Sigma}_y(u,v)\boldsymbol{\Sigma}_y(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v}_{\boldsymbol{\Lambda}} = \underbrace{\mathbf{B}\Big\{\int\int \boldsymbol{\Sigma}_f(u,v)\mathbf{B}^{\mathrm{T}}\mathbf{B}\boldsymbol{\Sigma}_f(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\Big\}\mathbf{B}^{\mathrm{T}}}_{\mathbf{L}} + \underbrace{\text{remaining terms}}_{\mathbf{S}}, \tag{4}$$

which is shown to be identified asymptotically as $p \to \infty$. We propose to carry out eigenanalysis of the sample version of $\boldsymbol{\Lambda}$ in (4) combined with least squares to estimate $\mathbf{B}$, $\mathbf{f}_t(\cdot)$ and hence $\boldsymbol{\Sigma}_\chi$, and then employ the same thresholding method to estimate $\boldsymbol{\Sigma}_\varepsilon$. This yields an Eigenanalysis of Doubly Integrated Gram covarIance and Thresholding (DIGIT) estimator.

The new contribution of this paper can be summarized in four key aspects. First, though our model (1) shares the same form as the one in Guo et al. (2022) and aligns with the direction of static factor models in Bai and Ng (2002) and Fan et al. (2013), substantial advances have been made in our methodology and theory: (i) We allow weak serial correlations in idiosyncratic components $\boldsymbol{\varepsilon}_t(\cdot)$ rather than assuming the white noise. (ii) Unlike the autocovariance-based method (Guo et al., 2022) for serially dependent data, we leverage the covariance information to propose a more efficient estimation procedure that encompasses independent observations as a special case. (iii) More importantly, under the pervasiveness assumption, we establish novel asymptotic identifiability in (4), where the first r eigenvalues of $\mathbf{L}$ grow at rate $O(p^2)$, whereas all the eigenvalues of $\mathbf{S}$ diverge at a rate slower than $O(p^2)$.

Second, for model (2), we extend the standard asymptotically identified covariance decomposition in Bai and Ng (2002) to the functional domain, under the functional counterpart of the pervasiveness assumption. Based on these findings, we provide mathematical insights when the functional factor analysis for models (1) and (2) and the proposed eigenanalysis of the respective $\mathbf{\Lambda}$'s in (3) and (4) are approximately the same for high-dimensional functional data/functional time series.

Third, we develop a new adaptive functional thresholding approach to estimate sparse $\mathbf{\Sigma}_\varepsilon$. Compared to the competitor in Fang et al. (2023), our approach requires weaker assumptions while achieving similar finite-sample performance. Fourth, with the aid of such thresholding technique in conjunction with our estimation of FFMs (1) and (2), we propose two factor-guided covariance matrix function estimators, DIGIT and FPOET, respectively. We derive the associated convergence rates of estimators for $\mathbf{\Sigma}_\varepsilon$, $\mathbf{\Sigma}_y$ and its inverse under various functional matrix norms. Additionally, we introduce fully functional information criteria to select the more suitable model between FFMs (1) and (2).

The rest of the paper is organized as follows. Section 2 presents the corresponding procedures for estimating $\mathbf{\Sigma}_y$ under two FFMs as well as the information criteria used for model selection. Section 3 provides the asymptotic theory for involved estimated quantities. Section 4 discusses a couple of applications of the proposed estimation. We assess the finite-sample performance of our proposal through extensive simulations in Section 5 and two real data applications in Section 6.

Throughout the paper, for any matrix $\mathbf{M} = (M_{ij})_{p\times q}$, we denote its matrices ℓ_1 norm, ℓ_∞ norm, operator norm, Frobenius norm and elementwise ℓ_∞ norm by $\|\mathbf{M}\|_1 = \max_j \sum_i |M_{ij}|$, $\|\mathbf{M}\|_\infty = \max_i \sum_j |M_{ij}|$, $\|\mathbf{M}\| = \lambda_{\max}^{1/2}(\mathbf{M}^{\mathrm{T}}\mathbf{M})$, $\|\mathbf{M}\|_{\mathrm{F}} = (\sum_{i,j} M_{ij}^2)^{1/2}$ and $\|\mathbf{M}\|_{\max} = \max_{i,j} |M_{ij}|$, respectively. Let $\mathbb{H} = L_2(\mathcal{U})$ be the Hilbert space of squared integrable functions defined on the compact set $\mathcal{U}$. We denote its p-fold Cartesian product by $\mathbb{H}^p = \mathbb{H} \times \cdots \times \mathbb{H}$ and tensor product by $\mathbb{S} = \mathbb{H} \otimes \mathbb{H}$. For $\mathbf{f} = (f_1, \ldots, f_p)^{\mathrm{T}}, \mathbf{g} = (g_1, \ldots, g_p)^{\mathrm{T}} \in \mathbb{H}^p$, we denote the inner product by $\langle \mathbf{f}, \mathbf{g} \rangle = \int_{\mathcal{U}} \mathbf{f}(u)^{\mathrm{T}}\mathbf{g}(u)\mathrm{d}u$ with induced norm $\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$. For an integral matrix operator $\mathbf{K} : \mathbb{H}^p \to \mathbb{H}^q$ induced from the kernel matrix function $\mathbf{K} = \{K_{ij}(\cdot,\cdot)\}_{q\times p}$ with each $K_{ij} \in \mathbb{S}$, $\mathbf{K}(\mathbf{f})(\cdot) = \int_{\mathcal{U}} \mathbf{K}(\cdot,u)\mathbf{f}(u)\mathrm{d}u \in \mathbb{H}^q$ for any given $\mathbf{f} \in \mathbb{H}^p$. For notational economy, we will use $\mathbf{K}$ to denote both the kernel function and the operator. We define the functional version of matrix ℓ_1 norm by $\|\mathbf{K}\|_{\mathcal{S},1} = \max_j \sum_i \|K_{ij}\|_{\mathcal{S}}$, where, for each $K_{ij} \in \mathbb{S}$, we denote its Hilbert–Schmidt norm by $\|K_{ij}\|_{\mathcal{S}} = \{\int\int K_{ij}(u,v)^2 \mathrm{d}u\mathrm{d}v\}^{1/2}$ and trace norm by $\|K_{ii}\|_{\mathcal{N}} = \int K_{ii}(u,u)\mathrm{d}u$ for $i = j$. Similarly, we define $\|\mathbf{K}\|_{\mathcal{S},\infty} = \max_i \sum_j \|K_{ij}\|_{\mathcal{S}}$, $\|\mathbf{K}\|_{\mathcal{S},\mathrm{F}} = \{\sum_{i,j} \|K_{ij}\|_{\mathcal{S}}^2\}^{1/2}$ and $\|\mathbf{K}\|_{\mathcal{S},\max} = \max_{i,j} \|K_{ij}\|_{\mathcal{S}}$ as the functional versions of matrix ℓ_∞, Frobenius and elementwise ℓ_∞ norms, respectively. We define the operator norm by $\|\mathbf{K}\|_{\mathcal{L}} = \sup_{\mathbf{x}\in\mathbb{H}^p, \|\mathbf{x}\|\leqslant 1} \|\mathbf{K}(\mathbf{x})\|$. For a positive integer m, write $[m] = \{1,\ldots,m\}$ and denote by $\mathbf{I}_m$ the identity matrix of size $m \times m$. For $x, y \in \mathbb{R}$, we use $x \wedge y = \min(x,y)$. For two positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ or $a_n = O(b_n)$ or $b_n \gtrsim a_n$ if there exists a positive constant c such that $a_n/b_n \leqslant c$, and $a_n = o(b_n)$ if $a_n/b_n \to 0$. We write $a_n \asymp b_n$ if and only if $a_n \lesssim b_n$ and $a_n \gtrsim b_n$ hold simultaneously.

2 Methodology

2.1 FFM with functional factors

Suppose that $\mathbf{y}_t(\cdot)$ admits FFM representation (1), where r common functional factors in $\mathbf{f}_t(\cdot) = \{f_{t1}(\cdot),\ldots,f_{tr}(\cdot)\}^{\mathrm{T}}$ are uncorrelated with the idiosyncratic errors $\boldsymbol{\varepsilon}_t(\cdot) = \{\varepsilon_{t1}(\cdot),\ldots,\varepsilon_{tp}(\cdot)\}^{\mathrm{T}}$ and r is assumed to be fixed. Then we have

$$\boldsymbol{\Sigma}_y(u,v) = \mathbf{B}\boldsymbol{\Sigma}_f(u,v)\mathbf{B}^{\mathrm{T}} + \boldsymbol{\Sigma}_\varepsilon(u,v), \quad (u,v) \in \mathcal{U}^2, \tag{5}$$

which is not nonnegative definite for some u, v. To ensure nonnegative definiteness and accumulate covariance information as much as possible, we propose to perform an eigenanalysis of doubly integrated Gram covariance:

$$\boldsymbol{\Omega} = \int\int \boldsymbol{\Sigma}_y(u,v)\boldsymbol{\Sigma}_y(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v \equiv \boldsymbol{\Omega}_{\mathcal{L}} + \boldsymbol{\Omega}_{\mathcal{R}}, \tag{6}$$

where $\boldsymbol{\Omega}_{\mathcal{L}} = \mathbf{B}\{\iint \boldsymbol{\Sigma}_f(u,v)\mathbf{B}^{\mathrm{T}}\mathbf{B}\boldsymbol{\Sigma}_f(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\}\mathbf{B}^{\mathrm{T}}$ and $\boldsymbol{\Omega}_{\mathcal{R}} = \iint \boldsymbol{\Sigma}_\varepsilon(u,v)\boldsymbol{\Sigma}_\varepsilon(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v + \iint \mathbf{B}\boldsymbol{\Sigma}_f(u,v)\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v + \iint \boldsymbol{\Sigma}_\varepsilon(u,v)\mathbf{B}\boldsymbol{\Sigma}_f(u,v)^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathrm{d}u\mathrm{d}v$. To make the decomposition (6) identifiable, we impose the following condition.

Assumption 1. *$p^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I}_r$ and $\iint \boldsymbol{\Sigma}_f(u,v)\boldsymbol{\Sigma}_f(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v = \mathrm{diag}(\theta_1,\ldots,\theta_r)$, where there exist some constants $\overline{\theta} > \underline{\theta} > 0$ such that $\overline{\theta} > \theta_1 > \theta_2 > \cdots > \theta_r > \underline{\theta}$.*

Remark 1. *Model (1) exhibits an identifiable issue as it remains unchanged if $\{\mathbf{B},\mathbf{f}_t(\cdot)\}$ is replaced by $\{\mathbf{B}\mathbf{U},\mathbf{U}^{-1}\mathbf{f}_t(\cdot)\}$ for any invertible matrix $\mathbf{U}$. Bai and Ng (2002) assumed two types of normalization for the scalar factor model: one is $p^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I}_r$ and the other is $Cov(\mathbf{f}_t) = \mathbf{I}_p$. We adopt the first type for model (1) to simplify the calculation of the low rank matrix $\boldsymbol{\Omega}_{\mathcal{L}}$ in (6). However, this constraint alone is insufficient to identify $\mathbf{B}$, but the space spanned by the columns of $\mathbf{B} = (\mathbf{b}_1,\ldots,\mathbf{b}_r)$. Hence, we introduce an additional constraint based on the diagonalization of $\iint \boldsymbol{\Sigma}_f(u,v)\boldsymbol{\Sigma}_f(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v$, which is ensured by the fact that any nonnegative-definite matrix can be orthogonally diagonalized. Under Assumption 1, we can express $\boldsymbol{\Omega}_{\mathcal{L}} = \sum_{i=1}^r p\theta_i\mathbf{b}_i\mathbf{b}_i^{\mathrm{T}}$, implying that $\|\boldsymbol{\Omega}_{\mathcal{L}}\| \asymp \|\boldsymbol{\Omega}_{\mathcal{L}}\|_{\min} \asymp p^2$*

We now elucidate why performing eigenanalysis of $\boldsymbol{\Omega}$ can be employed for functional factor analysis under model (1). Write $\widetilde{\mathbf{B}} = p^{-1/2}\mathbf{B} = (\widetilde{\mathbf{b}}_1,\cdots,\widetilde{\mathbf{b}}_r)$, which satisfies $\widetilde{\mathbf{B}}^{\mathrm{T}}\widetilde{\mathbf{B}} = \mathbf{I}_r$. Under Assumption 1, it holds that $\boldsymbol{\Omega}_{\mathcal{L}} = p^2\sum_{i=1}^r \theta_i\widetilde{\mathbf{b}}_i\widetilde{\mathbf{b}}_i^{\mathrm{T}}$, whose eigenvalue/eigenvector pairs are $\{(p^2\theta_i,\widetilde{\mathbf{b}}_i)\}_{i\in[r]}$. Let $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ be the ordered eigenvalues of $\boldsymbol{\Omega}$ and $\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_p$ be the corresponding eigenvectors. We then have the following proposition.

Proposition 1. *Suppose that Assumption 1 and $\|\boldsymbol{\Omega}_{\mathcal{R}}\| = o(p^2)$ hold. Then we have*
(i) $|\lambda_j - p^2\theta_j| \leqslant \|\boldsymbol{\Omega}_{\mathcal{R}}\|$ for $j\in[r]$ and $|\lambda_j| \leqslant \|\boldsymbol{\Omega}_{\mathcal{R}}\|$ for $j\in[p]\backslash[r]$;
(ii) $\|\boldsymbol{\xi}_j - \widetilde{\mathbf{b}}_j\| = O(p^{-2}\|\boldsymbol{\Omega}_{\mathcal{R}}\|)$ for $j\in[r]$.

Proposition 1 indicates that we can distinguish the leading eigenvalues $\{\lambda_j\}_{j\in[r]}$ from the remaining eigenvalues, and ensure the approximate equivalence between eigenvectors $\{\boldsymbol{\xi}_j\}_{j\in[r]}$ and the normalized factor loading columns $\{\widetilde{\mathbf{b}}_j\}_{j\in[r]}$, provided that $\|\boldsymbol{\Omega}_{\mathcal{R}}\| = o(p^2)$. Towards this, we impose an approximately functional sparsity condition on $\mathbf{\Sigma}_\varepsilon$ measured through

$$s_p = \max_{i\in[p]} \sum_{j=1}^{p} \|\sigma_i\|_{\mathcal{N}}^{(1-q)/2} \|\sigma_j\|_{\mathcal{N}}^{(1-q)/2} \|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}}^{q}, \quad \text{for some } q \in [0,1), \tag{7}$$

where $\sigma_i(u) = \Sigma_{\varepsilon,ii}(u,u)$ for $u \in \mathcal{U}$ and $i \in [p]$. Specially, when $q = 0$ and $\{\|\sigma_i\|_{\mathcal{N}}\}$ are bounded, s_p can be simplified to the exact functional sparsity, i.e., $\max_i \sum_j I(\|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}} \neq 0)$.

Remark 2. *Our proposed measure of functional sparsity in (7) for non-functional data degenerates to the measure of sparsity adopted in Cai and Liu (2011). It is worth mentioning that Fang et al. (2023) introduced a different measure of functional sparsity as*

$$\tilde{s}_p = \max_{i\in[p]} \sum_{j=1}^{p} \|\sigma_i\|_{\infty}^{(1-q)/2} \|\sigma_j\|_{\infty}^{(1-q)/2} \|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}}^{q},$$

where $\|\sigma_i\|_\infty = \sup_{u\in\mathcal{U}} \sigma_i(u) \geqslant \|\sigma_i\|_{\mathcal{N}}$. *As a result, we will use* s_p *instead of* $\tilde{s}_p$.
(ii) With bounded $\{\|\sigma_i\|_{\mathcal{N}}\}$, *we can easily obtain* $\|\mathbf{\Sigma}_\varepsilon\|_{\mathcal{S},1} = \|\mathbf{\Sigma}_\varepsilon\|_{\mathcal{S},\infty} = O(s_p)$, *which, along with Lemmas A6, B10 of the Supplementary Material and Assumption 1, yields that*

$$\|\mathbf{\Omega}_{\mathcal{R}}\| \leqslant \|\mathbf{\Sigma}_\varepsilon\|_{\mathcal{S},\infty}\|\mathbf{\Sigma}_\varepsilon\|_{\mathcal{S},1} + 2\left(\|\mathbf{B}\mathbf{\Sigma}_f\mathbf{B}^{\mathrm{T}}\|_{\mathcal{S},\infty}\|\mathbf{B}\mathbf{\Sigma}_f\mathbf{B}^{\mathrm{T}}\|_{\mathcal{S},1}\right)^{1/2}\left(\|\mathbf{\Sigma}_\varepsilon\|_{\mathcal{S},1}\|\mathbf{\Sigma}_\varepsilon\|_{\mathcal{S},\infty}\right)^{1/2} = O(s_p^2 + ps_p).$$

Hence, when $s_p = o(p)$, *Proposition 1 implies that functional factor analysis under model* (1) *and eigenanalysis of* $\mathbf{\Omega}$ *are approximately the same for high-dimensional functional data.*

To estimate model (1), we assume the number of functional factors (i.e., r) is known, and will introduce a data-driven approach to determine it in Section 2.3. Without loss of generality, we assume that $\mathbf{y}_t(\cdot)$ has been centered to have mean zero. The sample covariance matrix function of $\mathbf{\Sigma}_y(\cdot,\cdot)$ is given by $\widehat{\mathbf{\Sigma}}_y^{\mathcal{S}}(u,v) = n^{-1}\sum_{t=1}^{n} \mathbf{y}_t(u)\mathbf{y}_t(v)^{\mathrm{T}}$. Performing eigen-decomposition on the sample version of $\mathbf{\Omega}$,

$$\widehat{\mathbf{\Omega}} = \int\int \widehat{\mathbf{\Sigma}}_y^{\mathcal{S}}(u,v)\widehat{\mathbf{\Sigma}}_y^{\mathcal{S}}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v, \tag{8}$$

leads to estimated eigenvalues $\hat{\lambda}_1,\ldots,\hat{\lambda}_p$ and their associated eigenvectors $\widehat{\boldsymbol{\xi}}_1,\ldots,\widehat{\boldsymbol{\xi}}_p$. Then the estimated factor loading matrix is $\widehat{\mathbf{B}} = \sqrt{p}(\widehat{\boldsymbol{\xi}}_1,\ldots,\widehat{\boldsymbol{\xi}}_r) = (\widehat{\mathbf{b}}_1,\ldots,\widehat{\mathbf{b}}_r)$.

To estimate functional factors $\{\mathbf{f}_t(\cdot)\}_{t\in[n]}$, we minimize the least squares criterion

$$\sum_{t=1}^{n}\|\mathbf{y}_t-\widehat{\mathbf{B}}\mathbf{f}_t\|^2=\sum_{t=1}^{n}\int_{\mathcal{U}}\{\mathbf{y}_t(u)-\widehat{\mathbf{B}}\mathbf{f}_t(u)\}^{\mathrm{T}}\{\mathbf{y}_t(u)-\widehat{\mathbf{B}}\mathbf{f}_t(u)\}\mathrm{d}u \tag{9}$$

with respect to $\mathbf{f}_1(\cdot),\ldots,\mathbf{f}_n(\cdot)$. By setting the functional derivatives to zero, we obtain the least squares estimator $\widehat{\mathbf{f}}_t(\cdot)=p^{-1}\widehat{\mathbf{B}}^{\mathrm{T}}\mathbf{y}_t(\cdot)$ and the estimated idiosyncratic errors are given by $\widehat{\boldsymbol{\varepsilon}}_t(\cdot)=(\mathbf{I}_p-p^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{B}}^{\mathrm{T}})\mathbf{y}_t(\cdot)$. Hence, we can obtain sample covariance matrix functions of estimated common factors and estimated idiosyncratic errors as $\widehat{\boldsymbol{\Sigma}}_f(u,v)=n^{-1}\sum_{t=1}^{n}\widehat{\mathbf{f}}_t(u)\widehat{\mathbf{f}}_t(v)^{\mathrm{T}}$ and $\widehat{\boldsymbol{\Sigma}}_\varepsilon(u,v)=\{\widehat{\Sigma}_{\varepsilon,ij}(u,v)\}_{p\times p}=\sum_{t=1}^{n}n^{-1}\widehat{\boldsymbol{\varepsilon}}_t(u)\widehat{\boldsymbol{\varepsilon}}_t(v)^{\mathrm{T}}$, respectively.

Since $\boldsymbol{\Sigma}_\varepsilon$ is assumed to be functional sparse, we introduce an adaptive functional thresholding (AFT) estimator of $\boldsymbol{\Sigma}_\varepsilon$. To this end, we define the functional variance factors $\Theta_{ij}(u,v)=\mathrm{Var}\{\varepsilon_{ti}(u)\varepsilon_{tj}(v)\}$ for $i,j\in[p]$, whose estimators are

$$\widehat{\Theta}_{ij}(u,v)=\frac{1}{n}\sum_{t=1}^{n}\big\{\widehat{\varepsilon}_{ti}(u)\widehat{\varepsilon}_{tj}(v)-\widehat{\Sigma}_{\varepsilon,ij}(u,v)\big\}^2,$$

with $\widehat{\varepsilon}_{ti}(\cdot)=y_{ti}(\cdot)-\breve{\mathbf{b}}_i^{\mathrm{T}}\widehat{\mathbf{f}}_t(\cdot)$ and $\breve{\mathbf{b}}_i$ being the i-th row vector of $\widehat{\mathbf{B}}$. We develop an AFT procedure on $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ using entry-dependent functional thresholds that automatically adapt to the variability of $\widehat{\Sigma}_{\varepsilon,ij}$'s. Specifically, the AFT estimator is defined as $\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}}=\{\widehat{\Sigma}^{\mathcal{A}}_{\varepsilon,ij}(\cdot,\cdot)\}_{p\times p}$

$$\widehat{\Sigma}^{\mathcal{A}}_{\varepsilon,ij}=\big\|\widehat{\Theta}_{ij}^{1/2}\big\|_{\mathcal{S}}\times s_\lambda\Big(\widehat{\Sigma}_{\varepsilon,ij}/\big\|\widehat{\Theta}_{ij}^{1/2}\big\|_{\mathcal{S}}\Big)\ \text{with}\ \lambda=\dot{C}\Big(\sqrt{\frac{\log p}{n}}+\frac{1}{\sqrt{p}}\Big), \tag{10}$$

where $\dot{C}>0$ is a pre-specified constant that can be selected via multifold cross-validation and the order $\sqrt{\log p/n}+1/\sqrt{p}$ is related to the convergence rate of $\widehat{\Sigma}_{\varepsilon,ij}/\big\|\widehat{\Theta}_{ij}^{1/2}\big\|_{\mathcal{S}}$ under functional elementwise ℓ_∞ norm. Here s_λ is a functional thresholding operator with regularization parameter $\lambda\geqslant 0$ (Fang et al., 2023) and belongs to the class $s_\lambda:\mathbb{S}\to\mathbb{S}$ satisfying: (i) $\|s_\lambda(Z)\|_{\mathcal{S}}\leqslant c\|Y\|_{\mathcal{S}}$ for all $Z,Y\in\mathbb{S}$ that satisfy $\|Z-Y\|_{\mathcal{S}}\leqslant\lambda$ and some $c>0$; (ii) $\|s_\lambda(Z)\|_{\mathcal{S}}=0$ for $\|Z\|_{\mathcal{S}}\leqslant\lambda$; (iii) $\|s_\lambda(Z)-Z\|_{\mathcal{S}}\leqslant\lambda$ for all $Z\in\mathbb{S}$. This class includes functional versions of commonly adopted thresholding functions, such as hard thresholding, soft thresholding, smoothed clipped absolute deviation (Fan and Li, 2001), and the adaptive lasso (Zou, 2006).

Remark 3. *By comparison, Fang et al. (2023) introduced an alternative AFT estimator*

$$\widetilde{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}}=(\widetilde{\Sigma}^{\mathcal{A}}_{\varepsilon,ij})_{p\times p}\quad\textit{with}\quad\widetilde{\Sigma}^{\mathcal{A}}_{\varepsilon,ij}=\widehat{\Theta}_{ij}^{1/2}\times s_\lambda\Big(\widehat{\Sigma}_{\varepsilon,ij}/\widehat{\Theta}_{ij}^{1/2}\Big), \tag{11}$$

which uses a single threshold level to functionally threshold standardized entries $\widehat{\Sigma}_{\varepsilon,ij}/\widehat{\Theta}_{ij}^{1/2}$ *across all* (i,j)*, resulting in entry-dependent functional thresholds for* $\widehat{\Sigma}_{\varepsilon,ij}$*. Since* $\widetilde{\Sigma}_{\varepsilon,jk}^{\mathcal{A}}$ *requires stronger assumptions (see Remark 2 above and the remark for Assumption 5 below), we adopt the AFT estimator* $\widehat{\Sigma}_{\varepsilon,jk}^{\mathcal{A}}$ *leading to comparable empirical performance (see Section F of the Supplementary Material).*

Finally, we obtain an Eigenanalysis of Doubly Integrated Gram covarIance and Thresholding (DIGIT) estimator of $\mathbf{\Sigma}_y$ as

$$\widehat{\mathbf{\Sigma}}_y^{\mathcal{D}}(u,v) = \widehat{\mathbf{B}}\widehat{\mathbf{\Sigma}}_f(u,v)\widehat{\mathbf{B}}^{\mathrm{T}} + \widehat{\mathbf{\Sigma}}_{\varepsilon}^{\mathcal{A}}(u,v), \quad (u,v) \in \mathcal{U}^2. \tag{12}$$

2.2 FFM with functional loadings

The structure of FFM is not unique. We could also assume $\mathbf{y}_t(\cdot)$ satisfies FFM (2) with scalar factors and functional loadings $\mathbf{Q}(\cdot) = \{\mathbf{q}_1(\cdot), \dots, \mathbf{q}_p(\cdot)\}^{\mathrm{T}}$ with each $\mathbf{q}_i(\cdot) \in \mathbb{H}^r$, where r common scalar factors $\boldsymbol{\gamma}_t = (\gamma_{t1}, \dots, \gamma_{tr})^{\mathrm{T}}$ are uncorrelated with the idiosyncratic errors in $\boldsymbol{\varepsilon}_t(\cdot)$ and r is assumed to be fixed. Then we have the covariance decomposition

$$\mathbf{\Sigma}_y(u,v) = \mathbf{Q}(u)\mathbf{\Sigma}_\gamma\mathbf{Q}(v)^{\mathrm{T}} + \mathbf{\Sigma}_\varepsilon(u,v), \quad (u,v) \in \mathcal{U}^2. \tag{13}$$

By Mercer's theorem (Carmeli et al., 2006), which serves as the foundation of MFPCA (Happ and Greven, 2018), there exists an orthonormal basis consisting of eigenfunctions $\{\boldsymbol{\varphi}_i(\cdot)\}_{i=1}^{\infty}$ of $\mathbf{\Sigma}_y$ and the associated eigenvalues $\tau_1 \geqslant \tau_2 \geqslant \cdots \geqslant 0$ such that

$$\mathbf{\Sigma}_y(u,v) = \sum_{i=1}^{\infty} \tau_i \boldsymbol{\varphi}_i(u)\boldsymbol{\varphi}_i(v)^{\mathrm{T}}, \quad (u,v) \in \mathcal{U}^2. \tag{14}$$

We now provide mathematical insights into why MFPCA can be applied for functional factor analysis under model (2). To ensure the identifiability of the decomposition in (13), we impose a normalization-type condition similar to Assumption 1.

Assumption 1'. $\mathbf{\Sigma}_\gamma = \mathbf{I}_r$ *and* $p^{-1}\int \mathbf{Q}(u)^{\mathrm{T}}\mathbf{Q}(u)\mathrm{d}u = \mathrm{diag}(\vartheta_1, \dots, \vartheta_r)$*, where there exist some constants* $\overline{\vartheta} > \underline{\vartheta} > 0$ *such that* $\overline{\vartheta} > \vartheta_1 > \vartheta_2 > \cdots > \vartheta_r > \underline{\vartheta}$.

Suppose Assumption 1′ holds, and let $\widetilde{\mathbf{q}}_1(\cdot), \ldots, \widetilde{\mathbf{q}}_r(\cdot)$ be the normalized columns of $\mathbf{Q}(\cdot)$ such that $\|\widetilde{\mathbf{q}}_j\| = 1$ for $j \in [r]$. By Lemma A9 of the Supplementary Material, $\{\widetilde{\mathbf{q}}_j(\cdot)\}_{j\in[r]}$ are the orthonormal eigenfunctions of the kernel function $\mathbf{Q}(\cdot)\mathbf{Q}(\cdot)^{\mathrm{T}}$ with corresponding eigenvalues $\{p\vartheta_j\}_{j=1}^r$ and the rest 0. We then give the following proposition.

Proposition 2. *Suppose that Assumption 1′ and $\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}} = o(p)$ hold. Then we have*

(i) $|\tau_j - p\vartheta_j| \leqslant \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}}$ for $j \in [r]$ and $|\tau_j| \leqslant \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}}$ for $j \in [p]\backslash[r]$;

(ii) $\|\boldsymbol{\varphi}_j - \widetilde{\mathbf{q}}_j\| = O(p^{-1}\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}})$ for $j \in [r]$.

Proposition 2 implies that, if we can prove $\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}} = o(p)$, then we can distinguish the principle eigenvalues $\{\tau_j\}_{j\in[r]}$ from the remaining eigenvalues. Additionally, the first r eigenfunctions $\{\boldsymbol{\varphi}_j(\cdot)\}_{j\in[r]}$ are approximately the same as the normalized columns of $\{\widetilde{\mathbf{q}}_j(\cdot)\}_{j\in[r]}$. To establish this, we impose the same functional sparsity condition on $\boldsymbol{\Sigma}_\varepsilon$ as measured by s_p in (7). Applying Lemma A7(iii) of the Supplementary Material, we have $\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}} \leqslant \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},1}^{1/2}\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\infty}^{1/2} = O(s_p)$. Hence, when $s_p = o(p)$, MFPCA is approximately equivalent to functional factor analysis under model (2) for high-dimensional functional data.

We now present the estimation procedure assuming that r is known, and we will develop a ratio-based approach to identify r in Section 2.3. Let $\hat{\tau}_1 \geqslant \hat{\tau}_2 \cdots \geqslant 0$ be the eigenvalues of the sample covariance $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}$ and $\{\widehat{\boldsymbol{\varphi}}_j(\cdot)\}_{j=1}^\infty$ be their corresponding eigenfunctions. Then $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}$ has the spectral decomposition

$$\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v) = \sum_{j=1}^r \hat{\tau}_j \widehat{\boldsymbol{\varphi}}_j(u)\widehat{\boldsymbol{\varphi}}_j(v)^{\mathrm{T}} + \widehat{\mathbf{R}}(u,v),$$

where $\widehat{\mathbf{R}}(u,v) = \sum_{j=r+1}^\infty \hat{\tau}_j \widehat{\boldsymbol{\varphi}}_j(u)\widehat{\boldsymbol{\varphi}}_j(v)^{\mathrm{T}}$ is the functional principal orthogonal complement. Applying AFT as introduced in Section 2.1 to $\widehat{\mathbf{R}}$ yields the estimator $\widehat{\mathbf{R}}^{\mathcal{A}}$. Finally, we obtain a Functional Principal Orthogonal complEment Thresholding (FPOET) estimator as

$$\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{F}}(u,v) = \sum_{j=1}^r \hat{\tau}_j \widehat{\boldsymbol{\varphi}}_j(u)\widehat{\boldsymbol{\varphi}}_j(v)^{\mathrm{T}} + \widehat{\mathbf{R}}^{\mathcal{A}}(u,v). \tag{15}$$

It is noteworthy that, with $\boldsymbol{\Sigma}_y$ satisfying decompositions (5) and (13) under FFMs (1) and (2), respectively, both DIGIT and FPOET methods embrace the fundamental concept of a "low-rank plus sparse" representation generalized to the functional setting. Consequently, the common estimation steps involve applying PCA or MFPCA to estimate the factor loadings, and applying AFT to estimate sparse $\boldsymbol{\Sigma}_\varepsilon$. Essentially, these two methods are closely related, allowing the proposed estimators to exhibit empirical robustness even in cases of model misspecification (See details in Section 5). See also Section E.2 of the Supplementary Material for a discussion about the relationship between two FFMs.

We next present an equivalent representation of FPOET estimator (15) from a least squares perspective. We consider solving a constraint least squares minimization problem:

$$\{\widehat{\mathbf{Q}}(\cdot), \widehat{\boldsymbol{\Gamma}}\} = \arg\min_{\mathbf{Q}(\cdot),\boldsymbol{\Gamma}} \int \|\mathbf{Y}(u) - \mathbf{Q}(u)\boldsymbol{\Gamma}^{\mathrm{T}}\|_{\mathrm{F}}^2 \mathrm{d}u = \arg\min_{\mathbf{Q}(\cdot),\boldsymbol{\gamma}_1,\ldots,\boldsymbol{\gamma}_n} \sum_{t=1}^{n} \|\mathbf{y}_t - \mathbf{Q}\boldsymbol{\gamma}_t\|^2, \tag{16}$$

subject to the normalization constraint corresponding to Assumption 1', i.e.,

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}} = \mathbf{I}_r \quad \text{and} \quad \frac{1}{p}\int \mathbf{Q}(u)^{\mathrm{T}}\mathbf{Q}(u)\mathrm{d}u \text{ is diagonal,}$$

where $\mathbf{Y}(\cdot) = \{\mathbf{y}_1(\cdot), \ldots, \mathbf{y}_n(\cdot)\}$ and $\boldsymbol{\Gamma}^{\mathrm{T}} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n)$. Given each $\boldsymbol{\Gamma}$, setting the functional derivative of the objective in (16) w.r.t. $\mathbf{Q}(\cdot)$ to zero, we obtain the constrained least squares estimator $\widetilde{\mathbf{Q}}(\cdot) = n^{-1}\mathbf{Y}(\cdot)\boldsymbol{\Gamma}$. Plugging it into (16), the objective as a function of $\boldsymbol{\Gamma}$ becomes $\int \|\mathbf{Y}(u) - n^{-1}\mathbf{Y}(u)\boldsymbol{\Gamma}\boldsymbol{\Gamma}(u)^{\mathrm{T}}\|_{\mathrm{F}}^2\mathrm{d}u = \int \mathrm{tr}\big\{(\mathbf{I}_n - n^{-1}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}})\mathbf{Y}(u)^{\mathrm{T}}\mathbf{Y}(u)\big\}\mathrm{d}u$, the minimizer of which is equivalent to the maximizer of $\mathrm{tr}\big[\boldsymbol{\Gamma}^{\mathrm{T}}\big\{\int \mathbf{Y}(u)^{\mathrm{T}}\mathbf{Y}(u)\mathrm{d}u\big\}\boldsymbol{\Gamma}\big]$. This implies that the columns of $n^{-1/2}\widehat{\boldsymbol{\Gamma}}$ are the eigenvectors corresponding to the r largest eigenvalues of $\int \mathbf{Y}(u)^{\mathrm{T}}\mathbf{Y}(u)\mathrm{d}u \in \mathbb{R}^{n\times n}$, and then $\widehat{\mathbf{Q}}(\cdot) = n^{-1}\mathbf{Y}(\cdot)\widehat{\boldsymbol{\Gamma}}$.

Let $\widetilde{\boldsymbol{\varepsilon}}_t(\cdot) = \mathbf{y}_t(\cdot) - \widehat{\mathbf{Q}}(\cdot)\widehat{\boldsymbol{\gamma}}_t$ and $\widetilde{\boldsymbol{\Sigma}}_\varepsilon(u,v) = n^{-1}\sum_{t=1}^{n}\widetilde{\boldsymbol{\varepsilon}}_t(u)\widetilde{\boldsymbol{\varepsilon}}_t(v)^{\mathrm{T}}$. Applying our proposed AFT in (10) to $\widetilde{\boldsymbol{\Sigma}}_\varepsilon$ yields the estimator $\widetilde{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}}$. Analogous to the decomposition (13) under Assumption 1', we propose the following substitution estimator

$$\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{L}}(u,v) = \widehat{\mathbf{Q}}(u)\widehat{\mathbf{Q}}(v)^{\mathrm{T}} + \widetilde{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}}(u,v). \tag{17}$$

The following proposition reveals the equivalence between the FPOET estimator (15) and the constrained least squares estimator (17).

Proposition 3. *Suppose the same regularization parameters are used when applying AFT to $\widehat{\mathbf{R}}$ and $\widetilde{\boldsymbol{\Sigma}}_{\varepsilon}$. Then we have $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{F}} = \widehat{\boldsymbol{\Sigma}}_y^{\mathcal{L}}$ and $\widehat{\mathbf{R}}^{\mathcal{A}} = \widetilde{\boldsymbol{\Sigma}}_{\varepsilon}^{\mathcal{A}}$.*

Remark 4. *(i) While our FFM (2) shares the same form as the model studied in Tavakoli et al. (2023), which focused on the estimation of scalar factors and functional loadings from a least squares viewpoint, the main purpose of this paper lies in the estimation of large covariance matrix function. Consequently, we also propose a least-squares-based estimator of $\boldsymbol{\Sigma}_y$, which turns out to be equivalent to our FPOET estimator by Proposition 3.*
(ii) Using a similar procedure, we can also develop an alternative estimator for $\boldsymbol{\Sigma}_y$ under FFM (1) from a least squares perspective. However, this estimator is distinct from the DIGIT estimator (12) and leads to declined estimation efficiency. See detailed discussion in Section E.1 of the Supplementary Material.

2.3 Determining the number of factors

We have developed the estimation procedures for FFMs (1) and (2), assuming the known number of functional or scalar factors (i.e. r). In this section, we take the frequently-used ratio-based approach (Lam and Yao, 2012; Wang et al., 2021) to determine the value of r.

Under model (1), we let $\hat{\lambda}_1 \geqslant \cdots \geqslant \hat{\lambda}_p$ be the ordered eigenvalues of $\widehat{\boldsymbol{\Omega}}$ in (8), and propose to estimate r by

$$\hat{r}^{\mathcal{D}} = \arg\min_{r \in [c_r p]} \hat{\lambda}_{r+1}/\hat{\lambda}_r, \tag{18}$$

where we typically take $c_r = 0.75$ to circumvent the fluctuations caused by extremely small values. In practical implementation, we set $\hat{\lambda}_i/p^2$ to be 0 if its value is smaller than a pre-specified small threshold ϵ_0 (e.g., 0.01), and treat the ratio 0/0 as 1. Hence, $\hat{\lambda}_{i+1}/\hat{\lambda}_i = (\hat{\lambda}_{i+1}/p^2)/(\hat{\lambda}_i/p^2) = 0/0 = 1$ if neither $\hat{\lambda}_{i+1}/p^2$ nor $\hat{\lambda}_i/p^2$ exceeds ϵ_0 as $p \to \infty$.

For model (2), we employ a similar eigenvalue-ratio estimator given by:

$$\hat{r}^{\mathcal{F}} = \arg\min_{r \in [r_0]} \hat{\tau}_{r+1}/\hat{\tau}_r, \tag{19}$$

where $\{\hat{\tau}_i\}_{i=1}^{\infty}$ represents the ordered eigenvalues of the sample covariance $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(\cdot,\cdot)$. Similar to the previous case, we set $\hat{\tau}_i/p$ as 0 if its value is smaller than ϵ_0 and $0/0 = 1$.

2.4 Model selection criterion

A natural question that arises is which of the two candidate FFMs (1) and (2) is more appropriate for modeling $\{\mathbf{y}_t(\cdot)\}$. This section develops fully functional information criteria based on observed data for model selection.

When r functional factors are estimated under FFM (1), motivated from the least squares criterion (9), we define the mean squared residuals as

$$V^{\mathcal{D}}(r) = (pn)^{-1} \sum_{t=1}^{n} \|\mathbf{y}_t - p^{-1}\widehat{\mathbf{B}}_r\widehat{\mathbf{B}}_r^{\mathrm{T}}\mathbf{y}_t\|^2,$$

where $\widehat{\mathbf{B}}_r$ is the estimated factor loading matrix by DIGIT. Analogously, when r scalar factors are estimated under FFM (2), it follows from the objective function in (16) that the corresponding mean squared residuals is

$$V^{\mathcal{F}}(r) = (pn)^{-1} \sum_{t=1}^{n} \|\mathbf{y}_t - n^{-1}\mathbf{Y}\widehat{\boldsymbol{\Gamma}}_r\widehat{\boldsymbol{\gamma}}_{t,r}\|^2,$$

where $\widehat{\boldsymbol{\Gamma}}_r^{\mathrm{T}} = (\widehat{\boldsymbol{\gamma}}_{1,r}, \ldots, \widehat{\boldsymbol{\gamma}}_{n,r})$ is formed by estimated factors using FPOET.

For any given r, we propose the following information criteria:

$$\begin{aligned} \mathrm{PC}^{\mathcal{D}}(r) &= V^{\mathcal{D}}(r) + rg(p,n), \quad \mathrm{IC}^{\mathcal{D}}(r) = \log V^{\mathcal{D}}(r) + rg(p,n), \\ \mathrm{PC}^{\mathcal{F}}(r) &= V^{\mathcal{F}}(r) + rg(p,n), \quad \mathrm{IC}^{\mathcal{F}}(r) = \log V^{\mathcal{F}}(r) + rg(p,n), \end{aligned} \tag{20}$$

where $g(p,n)$ is a penalty function of (p,n). While there is much existing literature (c.f. Bai and Ng, 2002; Fan et al., 2013) that has adopted this type of criterion for identifying the number of factors in scalar factor models, we propose fully functional criteria for selecting the more appropriate FFM. Following Bai and Ng (2002), we suggest three examples of penalty functions, referred to as PC_1, PC_2, PC_3 and IC_1, IC_2, IC_3, respectively, in the penalized loss functions (20),

$$(i)\ g(p,n) = \frac{p+n}{pn}\log\left(\frac{pn}{p+n}\right),\ (ii)\ g(p,n) = \frac{p+n}{pn}\log(p \wedge n),\ (iii)\ g(p,n) = \frac{\log(p\wedge n)}{p\wedge n}.$$

For model selection, we define the differences in the corresponding information criteria between the two FFMs as $\Delta\mathrm{PC}_i = \mathrm{PC}_i^{\mathcal{D}}(\hat{r}^{\mathcal{D}}) - \mathrm{PC}_i^{\mathcal{F}}(\hat{r}^{\mathcal{F}})$ and $\Delta\mathrm{IC}_i = \mathrm{IC}_i^{\mathcal{D}}(\hat{r}^{\mathcal{D}}) - \mathrm{IC}_i^{\mathcal{F}}(\hat{r}^{\mathcal{F}})$ for $i = 1, 2, 3$. The negative (or positive) values of $\Delta\mathrm{PC}_i$'s and $\Delta\mathrm{IC}_i$'s indicate that FFM (1) (or FFM (2)) is more suitable based on the observed data $\{\mathbf{y}_t(\cdot)\}_{t\in[n]}$.

3 Theory

3.1 Assumptions

The assumptions for models (1) and (2) exhibit a close one-to-one correspondence. For clarity, we will present them separately in a pairwise fashion.

Assumption 2. *For model (1), $\{\mathbf{f}_t(\cdot)\}_{t\geqslant 1}$ and $\{\boldsymbol{\varepsilon}_t(\cdot)\}_{t\geqslant 1}$ are weakly stationary and $\mathbb{E}\{\varepsilon_{ti}(u)\} = \mathbb{E}\{\varepsilon_{ti}(u)f_{tj}(v)\} = 0$ for all $i \in [p], j \in [r]$ and $(u, v) \in \mathcal{U}^2$.*

Assumption 2'. *For model (2), $\{\boldsymbol{\gamma}_t\}_{t\geqslant 1}$ and $\{\boldsymbol{\varepsilon}_t(\cdot)\}_{t\geqslant 1}$ are weakly stationary and $\mathbb{E}\{\varepsilon_{ti}(u)\} = \mathbb{E}\{\varepsilon_{ti}(u)\gamma_{tj}\} = 0$ for all $i \in [p], j \in [r]$ and $u \in \mathcal{U}$.*

Assumption 3. *For model (1), there exists some constant $C > 0$ such that, for all $j \in [r]$, $t \in [n]$, (i) $\|\mathbf{b}_j\|_{\max} < C$, (ii) $\mathbb{E}\|p^{-1/2}\boldsymbol{\varepsilon}_t\|^4 < C$, (iii) $\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}} < C$, (iv) $\max_{i\in[p]}\|\Sigma_{\varepsilon,ii}\|_{\mathcal{N}} < C$.*

Assumption 3'. *For model (2), there exists some constant $C' > 0$ such that, for all $i \in [p]$, $t, t' \in [n]$: (i) $\|\mathbf{q}_i\| < C'$, (ii) $\mathbb{E}\|p^{-1/2}\sum_{i=1}^p \int \mathbf{q}_i(u)\varepsilon_{ti}(u)\mathrm{d}u\|^4 < C'$ and $\mathbb{E}\{p^{-1/2}[\langle\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t'}\rangle - \mathbb{E}\langle\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t'}\rangle]\}^4 < C'$, (iii) $\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}} < C'$, (iv) $\max_{i\in[p]}\|\Sigma_{\varepsilon,ii}\|_{\mathcal{N}} < C'$.*

Assumption 3(i) or 3'(i) requires the functional or scalar factors to be pervasive in the sense they influence a large fraction of the functional outcomes. Such pervasiveness-type assumption is commonly imposed in the literature (Bai, 2003; Fan et al., 2013). Assumption 3(ii) involves a standard moment constraint. Assumption 3'(ii) is needed to estimate scalar factors and functional loadings consistently. Assumption 3(iii) and 3'(iii) generalize the standard conditions for scalar factor models (Fan et al., 2018; Wang et al., 2021) to the functional domain. Assumptions 3(iv) and 3'(iv) are for technical convenience. However we can relax them by allowing $\max_i \|\Sigma_{\varepsilon,ii}\|_{\mathcal{N}}$ to grow at some slow rate as p increases.

We use the functional stability measure (Guo and Qiao, 2023) to characterize the serial dependence. For $\{\mathbf{y}_t(\cdot)\}$, denote its autocovariance matrix functions by $\boldsymbol{\Sigma}_y^{(h)}(u,v) = \text{Cov}\{\mathbf{y}_t(u), \mathbf{y}_{t+h}(v)\}$ for $h \in \mathbb{Z}$ and $(u,v) \in \mathcal{U}^2$ and its spectral density matrix function at frequency $\theta \in [-\pi, \pi]$ by $\mathbf{f}_{y,\theta}(u,v) = (2\pi)^{-1} \sum_{h\in\mathbb{Z}} \boldsymbol{\Sigma}_y^{(h)}(u,v) \exp(-ih\theta)$. The functional stability measure of $\{\mathbf{y}_t(\cdot)\}$ is defined as

$$\mathcal{M}_y = 2\pi \cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\phi}\in\mathbb{H}_0^p} \frac{\langle \boldsymbol{\phi}, \mathbf{f}_{y,\theta}(\boldsymbol{\phi})\rangle}{\langle \boldsymbol{\phi}, \boldsymbol{\Sigma}_y(\boldsymbol{\phi})\rangle},$$

where $\boldsymbol{\Sigma}_y(\boldsymbol{\phi})(\cdot) = \int_{\mathcal{U}} \boldsymbol{\Sigma}_y(\cdot, v)\boldsymbol{\phi}(v)\mathrm{d}v$ and $\mathbb{H}_0^p = \{\boldsymbol{\phi} \in \mathbb{H}^p : \langle \boldsymbol{\phi}, \boldsymbol{\Sigma}_y(\boldsymbol{\phi})\rangle \in (0,\infty)\}$. When $\mathbf{y}_1(\cdot), \ldots, \mathbf{y}_n(\cdot)$ are independent, $\mathcal{M}_y = 1$. See also Guo and Qiao (2023) for examples satisfying $\mathcal{M}_y < \infty$, such as functional moving average model and functional linear process. Similarly, we can define $\mathcal{M}_\varepsilon$ of $\{\boldsymbol{\varepsilon}_t(\cdot)\}$. To derive relevant exponential-type tails used in convergence analysis, we assume the sub-Gaussianities for functional (or scalar) factors and idiosyncratic components. We relegate the definitions of sub-Gaussian (functional) process and multivariate (functional) linear process to Section E.3 of the Supplementary Material.

Assumption 4. *For model* (1), *(i)* $\{\mathbf{f}_t(\cdot)\}_{t\in[n]}$ *and* $\{\boldsymbol{\varepsilon}_t(\cdot)\}_{t\in[n]}$ *follow sub-Gaussian functional linear processes; (ii)* $\mathcal{M}_\varepsilon < \infty$ *and* $\mathcal{M}_\varepsilon^2 \log p = o(n)$.

Assumption 4'. *For model* (2), *(i)* $\{\boldsymbol{\gamma}_t\}_{t\in[n]}$ *follows sub-Gaussian linear process and* $\{\boldsymbol{\varepsilon}_t(\cdot)\}_{t\in[n]}$ *follows sub-Gaussian functional linear process; (ii)*$\mathcal{M}_\varepsilon < \infty$ *and* $\mathcal{M}_\varepsilon^2 \log p = o(n)$.

Assumption 5. *There exists some constant* $\tau > 0$ *such that* $\min_{i,j\in[p]} \|\text{Var}(\varepsilon_{ti}\varepsilon_{tj})\|_{\mathcal{S}} \geqslant \tau$.

Assumption 6. *The pair* (n,p) *satisfies* $\mathcal{M}_\varepsilon^2 \log p = o(n/\log n)$ *and* $n = o(p^2)$.

Assumption 5 is required when implementing AFT, however, it is weaker than the similar assumption $\inf_{(u,v)\in\mathcal{U}^2} \min_{i,j\in[p]} \text{Var}[\varepsilon_{ti}(u)\varepsilon_{tj}(v)] \geqslant \tau$ imposed in Fang et al. (2023). Assumption 6 allows the high-dimensional case, where p grows exponentially as n increases.

3.2 Convergence of estimated loadings and factors

While the main focus of this paper is to estimate $\boldsymbol{\Sigma}_y$, the estimation of factors and loadings remains a crucial aspect, encompassed by DIGIT and FPOET estimators, as well as in many other applications. We first present various convergence rates of estimated factors and loading matrix when implementing DIGIT. For the sake of simplicity, we denote

$$\varpi_{n,p} = \mathcal{M}_\varepsilon\sqrt{\log p/n} + 1/\sqrt{p}.$$

Theorem 1. *Suppose that Assumptions 1–4 hold. Then there exists an orthogonal matrix* $\mathbf{U} \in \mathbb{R}^{r\times r}$ *such that (i)* $\|\widehat{\mathbf{B}} - \mathbf{B}\mathbf{U}^{\mathrm{T}}\|_{\max} = O_p\left(\varpi_{n,p}\right)$*; (ii)* $n^{-1}\sum_{t=1}^{n}\|\widehat{\mathbf{f}}_t - \mathbf{U}\mathbf{f}_t\|^2 = O_p(\mathcal{M}_\varepsilon^2/n + 1/p)$*; (iii)* $\max_{t\in[n]}\|\widehat{\mathbf{f}}_t - \mathbf{U}\mathbf{f}\| = O_p\big(\mathcal{M}_\varepsilon\sqrt{\log n/n} + \sqrt{n^{1/2}/p}\big)$.

The orthogonal matrix $\mathbf{U}$ above is needed to ensure that $\mathbf{b}_j^{\mathrm{T}}\widehat{\mathbf{b}}_j \geqslant 0$ for each $j \in [r]$. Provided that $\widehat{\mathbf{B}}\mathbf{U}\mathbf{U}^{\mathrm{T}}\widehat{\mathbf{f}}_t = \widehat{\mathbf{B}}\widehat{\mathbf{f}}_t$, the estimation of the common components and $\boldsymbol{\Sigma}_y$ remain unaffected by the choice of $\mathbf{U}$. By Theorem 1, we can derive the following corollary, which provides the uniform convergence rate of the estimated common component. Let $\breve{\mathbf{b}}_i$ and $\breve{\mathbf{b}}_i$ denote the i-th rows of $\mathbf{B}$ and $\widehat{\mathbf{B}}$, respectively.

Corollary 1. *Under the assumptions of Theorem 1, we have* $\max_{i\in[p],t\in[n]}\|\breve{\mathbf{b}}_i^{\mathrm{T}}\widehat{\mathbf{f}}_t - \breve{\mathbf{b}}_i^{\mathrm{T}}\mathbf{f}_t\| = O_p(\varrho)$*, where* $\varrho = \mathcal{M}_\varepsilon\sqrt{\log n\log p/n} + \sqrt{n^{1/2}/p}$.

In the context of FPOET estimation of factors and loadings, we require an additional asymptotically orthogonal matrix $\mathbf{H}$ such that $\widehat{\boldsymbol{\gamma}}_t$ is a valid estimator of $\mathbf{H}\boldsymbol{\gamma}_t$. Differing from DIGIT, we follow Bai (2003) to construct $\mathbf{H}$ in a deterministic form. Let $\mathbf{V} \in \mathbb{R}^{r\times r}$ denote the diagonal matrix of the first r largest eigenvalues of $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}$ in a decreasing order. Define $\mathbf{H} = n^{-1}\mathbf{V}^{-1}\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\boldsymbol{\Gamma}\int\mathbf{Q}(u)^{\mathrm{T}}\mathbf{Q}(u)\mathrm{d}u$. By Lemma B35 of the Supplementary Material, $\mathbf{H}$ is asymptotically orthogonal such that $\mathbf{I}_r = \mathbf{H}^{\mathrm{T}}\mathbf{H} + o_p(1) = \mathbf{H}\mathbf{H}^{\mathrm{T}} + o_p(1)$.

Theorem 1'. *Suppose that Assumptions 1'–4' hold. (i)* $n^{-1}\sum_{t=1}^{n}\|\widehat{\boldsymbol{\gamma}}_t - \mathbf{H}\boldsymbol{\gamma}_t\|^2 = O_p\big(\mathcal{M}_\varepsilon^2/n + 1/p\big)$*; (ii)* $\max_{t\in[n]}\|\widehat{\boldsymbol{\gamma}}_t - \mathbf{H}\boldsymbol{\gamma}_t\| = O_p\big(\mathcal{M}_\varepsilon/\sqrt{n} + \sqrt{n^{1/2}/p}\big)$*; (iii)* $\max_{i\in[p]}\|\widehat{\mathbf{q}}_i - \mathbf{H}\mathbf{q}_i\| = O_p(\varpi_{n,p})$.

Corollary 1'. *Under the assumptions of Theorem 1', we have* $\max_{i\in[p],t\in[n]} \|\widehat{\mathbf{q}}_i^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_t - \mathbf{q}_i^{\mathrm{T}}\boldsymbol{\gamma}_t\| = O_p(\varrho)$*, where* ϱ *is specified in Corollary 1.*

The convergence rates presented in Theorem 1 and Corollary 1 for model (1) are, respectively, consistent to those established in Bai (2003) and Fan et al. (2013) when $\mathcal{M}_\varepsilon = O(1)$. Additionally, the rates in Theorem 1' and Corollary 1' for model (2) align with those in Theorem 1 and Corollary 1 . These uniform convergence rates are essential not only for estimating the FFMs but also for many subsequent high-dimensional learning tasks.

Theorem 2. *Under the assumptions of Theorems 1 and 1', we have (i)* $\mathbb{P}(\hat{r}^{\mathcal{D}} = r) \to 1$*, and (ii)* $\mathbb{P}(\hat{r}^{\mathcal{F}} = r) \to 1$ *as* $n, p \to \infty$*, where* $\hat{r}^{\mathcal{D}}$ *and* $\hat{r}^{\mathcal{F}}$ *are defined in (18) and (19), respectively.*

Remark 5. *With the aid of Theorem 2, our estimators explored in Sections 3.2 and 3.3 are asymptotically adaptive to* r*. To see this, consider, e.g., model (2), and let* $\widehat{\boldsymbol{\gamma}}_{t,\hat{r}}$ *and* $\widehat{\mathbf{q}}_{i,\hat{r}}(\cdot)$ *be constructed using* $\hat{r}^{\mathcal{F}}$ *estimated scalar factors and functional loadings. Then, for any constant* $\tilde{c} > 0$, $\mathbb{P}\big(\varrho^{-1}\max_{i\in[p],t\in[n]} \|\widehat{\mathbf{q}}_{i,\hat{r}}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_{t,\hat{r}} - \mathbf{q}_i^{\mathrm{T}}\boldsymbol{\gamma}_t\| > \tilde{c}\big) \leqslant \mathbb{P}\big(\varrho^{-1}\max_{i\in[p],t\in[n]} \|\widehat{\mathbf{q}}_i^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_t - \mathbf{q}_i^{\mathrm{T}}\boldsymbol{\gamma}_t\| > \tilde{c}|\hat{r}^{\mathcal{F}} = r\big) + \mathbb{P}(\hat{r}^{\mathcal{F}} \neq r)$*, which, combined with Corollary 1', implies that* $\max_{i\in[p],t\in[n]} \|\widehat{\mathbf{q}}_{i,\hat{r}}^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_{t,\hat{r}} - \mathbf{q}_i^{\mathrm{T}}\boldsymbol{\gamma}_t\| = O_p(\varrho)$*. Similar arguments can be applied to other estimated quantities in Sections 3.2 and 3.3. Therefore, we assume that* r *is known in our asymptotic results.*

3.3 Convergence of estimated covariance matrix functions

Estimating the idiosyncratic covariance matrix function $\boldsymbol{\Sigma}_\varepsilon$ is important in factor modeling and subsequent learning tasks. With the help of functional sparsity as specified in (7), we can obtain consistent estimators of $\boldsymbol{\Sigma}_\varepsilon$ under functional matrix ℓ_1 norm $\|\cdot\|_{\mathcal{S},1}$ in the high-dimensional scenario. The following rates of convergence based on estimated idiosyncratic components are consistent with the rate based on direct observations of independent functional data (Fang et al., 2023) when $\mathcal{M}_\varepsilon = O(1)$ and $p \log p \gtrsim n$.

Theorem 3. *Suppose that Assumptions 1–6 hold. Then, for a sufficiently large constant* $\dot{C}$ *in (10),* $\|\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}} - \boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},1} = O_p(\varpi_{n,p}^{1-q} s_p)$.

Theorem 3'. *Suppose that Assumptions 1'–4', 5, 6 hold. Then, for a sufficiently large constant* $\dot{C}$ *in* (10), $\|\widehat{\mathbf{R}}^{\mathcal{A}} - \mathbf{\Sigma}_{\varepsilon}\|_{\mathcal{S},1} = O_p(\varpi_{n,p}^{1-q} s_p)$.

When assessing the convergence criteria for our DIGIT and FPOET estimators, it is crucial to note that functional matrix norms such as $\|\cdot\|_{\mathcal{S},1}$ and $\|\cdot\|_{\mathcal{L}}$ are not suitable choices. This is because $\widehat{\mathbf{\Sigma}}_y$ may not converge to $\mathbf{\Sigma}_y$ in these norms for high-dimensional functional data, unless specific structural assumptions are directly imposed on $\mathbf{\Sigma}_y$. This issue does not arise from the poor performance of estimation methods but rather from the inherent limitation of high-dimensional models. To address this, we present convergence rates in functional elementwise ℓ_∞ norm $\|\cdot\|_{\mathcal{S},\max}$.

Theorem 4. *Under the assumptions of Theorem 3, we have* $\|\widehat{\mathbf{\Sigma}}_y^{\mathcal{D}} - \mathbf{\Sigma}_y\|_{\mathcal{S},\max} = O_p(\varpi_{n,p})$.

Theorem 4'. *Under the assumptions of Theorem 3', we have* $\|\widehat{\mathbf{\Sigma}}_y^{\mathcal{F}} - \mathbf{\Sigma}_y\|_{\mathcal{S},\max} = O_p(\varpi_{n,p})$.

Remark 6. *(i) The convergence rates of DIGIT and FPOET estimators (we use* $\widehat{\mathbf{\Sigma}}_y$ *to denote both) comprise two terms. The first term* $O_p(\mathcal{M}_{\varepsilon}\sqrt{\log p/n})$ *arises from the rate of* $\widehat{\mathbf{\Sigma}}_y^{\mathcal{S}}$*, while the second term* $O_p(p^{-1/2})$ *primarily stems from the estimation of unobservable factors. When* $\mathcal{M}_{\varepsilon} = O(1)$*, our rate aligns with the result obtained in Fan et al. (2013).*
(ii) Compared to $\widehat{\mathbf{\Sigma}}_y^{\mathcal{S}}$*, we observe that using a factor-guided approach results in the same rate in* $\|\cdot\|_{\mathcal{S},\max}$ *as long as* $p\log p \gtrsim n$*. Nevertheless, our proposed estimators offer several advantages. First, under a functional weighted quadratic norm introduced in Section 4.1, which is closely related to functional risk management,* $\widehat{\mathbf{\Sigma}}_y$ *converges to* $\mathbf{\Sigma}_y$ *in the high-dimensional case (see Theorem 6), while* $\widehat{\mathbf{\Sigma}}_y^{\mathcal{S}}$ *does not achieve this convergence. Second, as evidenced by empirical results in Sections 5 and 6,* $\widehat{\mathbf{\Sigma}}_y$ *significantly outperforms* $\widehat{\mathbf{\Sigma}}_y^{\mathcal{S}}$ *in terms of various functional matrix losses.*

Finally, we explore convergence properties of the inverse covariance matrix function estimation. Denote the null space of $\mathbf{\Sigma}_y$ and its orthogonal complement by $\ker(\mathbf{\Sigma}_y) = \{\mathbf{x} \in \mathbb{H}^p : \mathbf{\Sigma}_y(\mathbf{x}) = \mathbf{0}\}$ and $\ker(\mathbf{\Sigma}_y)^{\perp} = \{\mathbf{x} \in \mathbb{H}^p : \langle \mathbf{x}, \mathbf{y}\rangle = 0, \forall \mathbf{y} \in \ker(\mathbf{\Sigma}_y)\}$, respectively.

The inverse covariance matrix function $\boldsymbol{\Sigma}_y^{-1}$ corresponds to the inverse of the restricted covariance matrix function $\boldsymbol{\Sigma}_y\big|\ker(\boldsymbol{\Sigma}_y)^{\perp}$, which restricts the domain of $\boldsymbol{\Sigma}_y$ to $\ker(\boldsymbol{\Sigma}_y)^{\perp}$. The similar definition applies to the inverses of $\boldsymbol{\Sigma}_f$ and $\boldsymbol{\Sigma}_\varepsilon$. With the DIGIT estimator $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{D}}(\cdot,\cdot) = \widehat{\mathbf{B}}\widehat{\boldsymbol{\Sigma}}_f(\cdot,\cdot)\widehat{\mathbf{B}}^{\mathrm{T}} + \widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}}(\cdot,\cdot)$, we apply Sherman–Morrison–Woodbury identity (Theorem 3.5.6 of Hsing and Eubank, 2015) to obtain its inverse $\big(\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{D}}\big)^{-1} = \big(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}}\big)^{-1} - \big(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}}\big)^{-1}\widehat{\mathbf{B}}\big\{\widehat{\boldsymbol{\Sigma}}_f^{-1} + \widehat{\mathbf{B}}^{\mathrm{T}}\big(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}}\big)^{-1}\widehat{\mathbf{B}}\big\}^{-1}\widehat{\mathbf{B}}^{\mathrm{T}}\big(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}}\big)^{-1}$. The inverse FPOET estimator can be obtained similarly. Then, within finite-dimensional Hilbert space, both the inverse DIGIT and FPOET estimators are consistent in the operator norm, as presented in the following theorems.

Theorem 5. *Suppose that the assumptions of Theorem 4 hold, $\varpi_{n,p}^{1-q}s_p = o(1)$, and both $\lambda_{\min}(\boldsymbol{\Sigma}_\varepsilon)$ and $\lambda_{\min}(\boldsymbol{\Sigma}_f)$ are bounded away from zero. Then, $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{D}}$ has a bounded inverse with probability approaching 1, and $\big\|(\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{D}})^{-1} - \boldsymbol{\Sigma}_y^{-1}\big\|_{\mathcal{L}} = O_p(\varpi_{n,p}^{1-q}s_p)$.*

Theorem 5'. *Suppose that the assumptions of Theorem 4' hold, $\varpi_{n,p}^{1-q}s_p = o(1)$, and $\lambda_{\min}(\boldsymbol{\Sigma}_\varepsilon)$ is bounded away from zero. Then, $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{F}}$ has a bounded inverse with probability approaching 1, and $\big\|(\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{F}})^{-1} - \boldsymbol{\Sigma}_y^{-1}\big\|_{\mathcal{L}} = O_p(\varpi_{n,p}^{1-q}s_p)$.*

Remark 7. *(i) The condition that $\lambda_{\min}(\boldsymbol{\Sigma}_\varepsilon)$ and $\lambda_{\min}(\boldsymbol{\Sigma}_f)$ are bounded away from zero can also imply that $\lambda_{\min}(\boldsymbol{\Sigma}_y)$ is bounded away from zero, which means that $\boldsymbol{\Sigma}_y$ has a finite number of nonzero eigenvalues, denoted as $d_n < \infty$, i.e., $\{\mathbf{y}_t(\cdot)\}_{t\in[n]}$ are finite-dimensional functional objects (Bathia et al., 2010). While the inverse of the sample covariance matrix function fails to exhibit convergence even though it operates within finite-dimensional Hilbert space, our factor-guided methods can achieve such convergence. It should be noted that d_n can be made arbitrarily large relative to n, e.g., $d_n = 2000, n = 200$. Hence, this finite-dimensional assumption does not place a practical constraint on our method. See also applications of inverse covariance matrix function estimation including functional risk management in Section 4.1 and sparse precision matrix function estimation in Section 4.2.*

(ii) Within infinite-dimensional Hilbert space, $\boldsymbol{\Sigma}_y^{-1}$ becomes an unbounded operator, which is discontinuous and cannot be estimated in a meaningful way. However, $\boldsymbol{\Sigma}_y^{-1}$ is usually asso-

ciated with another function/operator, and the composite function/operator in ker$(\boldsymbol{\Sigma}_y)^{\perp}$ *can reasonably be assumed to be bounded, such as regression function/operator and discriminant direction function in Section 4.2. Specifically, consider the spectral decomposition* (14)*, which is truncated at* $d_n < \infty$*, i.e.,* $\boldsymbol{\Sigma}_{y,d_n}(u,v) = \sum_{i=1}^{d_n} \tau_i \boldsymbol{\varphi}_i(u)\boldsymbol{\varphi}_i(v)^{\mathrm{T}}$*. Under certain smoothness conditions, such as those on coefficient functions in multivariate functional linear regression (Chiou et al., 2016), the impact of truncation errors through* $\sum_{i=d_n+1}^{\infty} \tau_i^{-1} \boldsymbol{\varphi}_i(u)\boldsymbol{\varphi}_i(v)^{\mathrm{T}}$ *on associated functions/operators is expected to diminish, ensuring the boundedness of composite functions/operators. Consequently, the primary focus shifts towards estimating the inverse of* $\boldsymbol{\Sigma}_{y,d_n}$*, and our results in Theorems 5 and 5′ become applicable.*

Upon observation, a remarkable consistency is evident between DIGIT and FPOET methods developed under different models in terms of imposed regularity assumptions and associated convergence rates, despite the substantially different proof techniques employed.

4 Applications

4.1 Functional risk management

One main task of risk management in the stock market is to estimate the portfolio variance, which can be extended to the functional setting to account for additional intraday uncertainties. Consider a portfolio consisting of p stocks, where the i-th component of $\mathbf{y}_t(\cdot)$ represents the cumulative intraday return (CIDR) trajectory (Horváth et al., 2014) for the i-th stock on the t-th trading day. Additionally, let $\mathbf{w}(u) = \{w_1(u), \ldots, w_p(u)\}^{\mathrm{T}}$ denote the allocation vector of the functional portfolio at time $u \in \mathcal{U}$. For a given $\mathbf{w}(\cdot)$, the true and perceived variances (i.e. risks) of the functional portfolio are $\langle \mathbf{w}, \boldsymbol{\Sigma}_y(\mathbf{w})\rangle$ and $\langle \mathbf{w}, \widehat{\boldsymbol{\Sigma}}_y(\mathbf{w})\rangle$, respectively. According to Proposition S.1 of the Supplementary Material, the estimation error of the functional portfolio variance is bounded by

$$\big|\langle \mathbf{w}, \widehat{\boldsymbol{\Sigma}}_y(\mathbf{w})\rangle - \langle \mathbf{w}, \boldsymbol{\Sigma}_y(\mathbf{w})\rangle\big| \leqslant \|\widehat{\boldsymbol{\Sigma}}_y - \boldsymbol{\Sigma}_y\|_{\mathcal{S},\max}\Big(\sum_{i=1}^{p} \|w_i\|\Big)^2,$$

in which Theorems 4 and 4' quantify the maximum approximation error $\|\widehat{\mathbf{\Sigma}}_y - \mathbf{\Sigma}_y\|_{\mathcal{S},\max}$.

In addition to the absolute error between perceived and true risks, we are also interested in quantifying the relative error. To this end, we introduce the functional version of weighted quadratic norm (Fan et al., 2008), defined as $\|\mathbf{K}\|_{\mathcal{S},\Sigma_y} = p^{-1/2}\|\mathbf{\Sigma}_y^{-1/2}\mathbf{K}\mathbf{\Sigma}_y^{-1/2}\|_{\mathcal{S},\mathrm{F}}$, where $\mathbf{K} \in \mathbb{H}^p \otimes \mathbb{H}^p$ and the normalization factor $p^{-1/2}$ serves the role of $\|\mathbf{\Sigma}_y\|_{\mathcal{S},\Sigma_y} = 1$. To ensure the validity of this functional norm, we assume that $\mathbf{\Sigma}_y$ has a bounded inverse, which does not place a constraint in practice (see Remark 7(i)). With such functional norm, the relative error can be measured by

$$p^{-1/2}\|\mathbf{\Sigma}_y^{-1/2}\widehat{\mathbf{\Sigma}}_y\mathbf{\Sigma}_y^{-1/2} - \tilde{\mathbf{I}}_p\|_{\mathcal{S},\mathrm{F}} = \|\widehat{\mathbf{\Sigma}}_y - \mathbf{\Sigma}_y\|_{\mathcal{S},\Sigma_y}, \tag{21}$$

where $\tilde{\mathbf{I}}_p$ denotes the identity operator. Provided that $\|\widehat{\mathbf{\Sigma}}_y^{\mathcal{S}} - \mathbf{\Sigma}_y\|_{\mathcal{S},\Sigma_y} = O_p(\mathcal{M}_\varepsilon\sqrt{p/n})$, the sample covariance estimator fails to converge in $\|\cdot\|_{\mathcal{S},\Sigma_y}$ under the high-dimensional setting with $p > n$. On the contrary, the following theorem reveals that our DIGIT estimator $\widehat{\mathbf{\Sigma}}_y^{\mathcal{D}}$ converges to $\mathbf{\Sigma}_y$ as long as $\mathcal{M}_\varepsilon^4 p = o(n^2)$ and $\varpi_{n,p}^{1-q}s_p = o(1)$. The same result can also be extended to the FPOET estimator.

Theorem 6. *Under the assumptions of Theorem 5, we have* $\|\widehat{\mathbf{\Sigma}}_y^{\mathcal{D}} - \mathbf{\Sigma}_y\|_{\mathcal{S},\Sigma_y} = O_p\big(\mathcal{M}_\varepsilon^2 p^{1/2}n^{-1} + \varpi_{n,p}^{1-q}s_p\big)$.

By Proposition S.2 of the Supplementary Material, the relative error is bounded by

$$\big|\langle \mathbf{w}, \widehat{\mathbf{\Sigma}}_y(\mathbf{w})\rangle / \langle \mathbf{w}, \mathbf{\Sigma}_y(\mathbf{w})\rangle - 1\big| \leqslant \|\mathbf{\Sigma}_y^{-1/2}\widehat{\mathbf{\Sigma}}_y\mathbf{\Sigma}_y^{-1/2} - \tilde{\mathbf{I}}_p\|_{\mathcal{L}},$$

which, in conjunction with Theorem 6 and (21), controls the maximum relative error.

4.2 Estimation of precision matrix, regression, and discriminant direction functions

The second application considers estimating functional graphical models (Qiao et al., 2019), which aim to identify the conditional dependence structure among components in

$\mathbf{y}_t(\cdot)$. For Gaussian data, this task is equivalent to estimating the sparse inverse covariance (i.e., precision) matrix function, which is bounded for finite-dimensional functional objects. Our inverse DIGIT or FPOET estimators combined with functional thresholding can thus be utilized.

The third application explores multivariate functional linear regression (Chiou et al., 2016), which involves a scalar response z_t or a functional response

$$z_t(v) = \langle \mathbf{y}_t, \boldsymbol{\beta}(\cdot, v) \rangle + e_t(v), \quad v \in \mathcal{V},$$

where $\boldsymbol{\beta}(\cdot,\cdot) = \{\beta_1(\cdot,\cdot), \ldots, \beta_p(\cdot,\cdot)\}^{\mathrm{T}}$ is operator-valued coefficient vector to be estimated. We can impose certain smoothness condition such that $\boldsymbol{\beta}(u,v) = \sum_{i=1}^{\infty} \tilde{\tau}_i \boldsymbol{\varphi}_i(u)\boldsymbol{\varphi}_i(v)^{\mathrm{T}}$ is sufficiently smooth relative to $\boldsymbol{\Sigma}_y(u,v) = \sum_{i=1}^{\infty} \tau_i \boldsymbol{\varphi}_i(u)\boldsymbol{\varphi}_i(v)^{\mathrm{T}}$, ensuring the boundedness of the regression operator $\boldsymbol{\beta}(u,v) = \int_{\mathcal{U}} \boldsymbol{\Sigma}_y^{-1}(u,u')\mathrm{Cov}\{\mathbf{y}_t(u'), z_t(v)\}\mathrm{d}u'$. Replacing relevant terms by their (truncated) sample versions, we obtain $\widehat{\boldsymbol{\beta}}(u,v) = n^{-1}\sum_{t=1}^{n}\int_{\mathcal{U}} \widehat{\boldsymbol{\Sigma}}_{y,d_n}^{-1}(u,u')\mathbf{y}_t(u')z_t(v)\mathrm{d}u'$. This application highlights the need for estimators $\widehat{\boldsymbol{\Sigma}}_{y,d_n}^{-1}$, as studied in Theorems 5 and 5'.

The fourth application delves into linear discriminant analysis for classifying multivariate functional data (Xue et al., 2023) with class labels $w_t = \{1, 2\}$. Specifically, we assume that $\mathbf{y}_t(\cdot)|w_t = 1$ and $\mathbf{y}_t(\cdot)|w_t = 2$ follow multivariate Gaussian distributions with mean functions $\boldsymbol{\mu}_1(\cdot)$ and $\boldsymbol{\mu}_2(\cdot)$, respectively, while sharing a common covariance matrix function $\boldsymbol{\Sigma}_y$. Our goal is to determine the linear classifier by estimating the discriminant direction function $\int_{\mathcal{U}} \boldsymbol{\Sigma}_y^{-1}(u,v)\{\boldsymbol{\mu}_1(v) - \boldsymbol{\mu}_2(v)\}\mathrm{d}v$, which takes the same form as the regression function $\boldsymbol{\beta}(u) = \int_{\mathcal{U}} \boldsymbol{\Sigma}_y^{-1}(u,v)\mathrm{Cov}\{\mathbf{y}_t(v), z_t\}\mathrm{d}v$ encountered in the third application with a scalar response z_t. By similar arguments as above, both applications call for the use of estimators $\widehat{\boldsymbol{\Sigma}}_{y,d_n}^{-1}$.

4.3 Estimation of correlation matrix function

The fifth application involves estimating the correlation matrix function and its inverse, which are essential in various graphical models for truly infinite-dimensional objects, see, e.g., Solea and Li (2022) and Zapata et al. (2022). Our proposed covariance estimators can be employed to estimate the corresponding correlation matrix function and its inverse.

Specifically, let $\mathbf{D}_y(\cdot,\cdot) = \text{diag}\{\Sigma_{y,11}(\cdot,\cdot),\ldots,\Sigma_{y,pp}(\cdot,\cdot)\}$ be the $p \times p$ diagonal matrix function. According to Baker (1973), there exists a correlation matrix function $\mathbf{C}_y$ with $\|\mathbf{C}_y\|_{\mathcal{L}} \leqslant 1$ such that $\boldsymbol{\Sigma}_y = \mathbf{D}_y^{1/2}\mathbf{C}_y\mathbf{D}_y^{1/2}$. Under certain compactness and smoothness assumptions, $\mathbf{C}_y$ has a bounded inverse, denoted by $\boldsymbol{\Theta}_y$, and its functional sparsity pattern corresponds to the network (i.e., conditional dependence) structure among p components in $\mathbf{y}_t(\cdot)$; see Solea and Li (2022). In general, the estimator $\widehat{\mathbf{D}}_y = \text{diag}(\widehat{\Sigma}_{y,11},\ldots,\widehat{\Sigma}_{y,pp})$ is non-invertible, so we can adopt the Tikhonov regularization to estimate $\mathbf{C}_y$ by $\widehat{\mathbf{C}}_y^{(\kappa)} = (\widehat{\mathbf{D}}_y + \kappa\mathbf{I}_p)^{-1/2}\widehat{\boldsymbol{\Sigma}}_y(\widehat{\mathbf{D}}_y + \kappa\mathbf{I}_p)^{-1/2}$ for some regularization parameter $\kappa > 0$. The estimator of $\boldsymbol{\Theta}_y$ is then given by $\widehat{\boldsymbol{\Theta}}_y^{(\kappa)} = \widehat{\mathbf{D}}_y^{1/2}(\widehat{\boldsymbol{\Sigma}}_y + \kappa\mathbf{I}_p)^{-1}\widehat{\mathbf{D}}_y^{1/2}$. Consequently, we can plug into the DIGIT or the FPOET estimator for estimating $\mathbf{C}_y$ and its inverse $\boldsymbol{\Theta}_y$.

5 Simulations

For the first data-generating process (denoted as DGP1), we generate observed data from model (1), where the entries of $\mathbf{B} \in \mathbb{R}^{p\times r}$ are sampled independently from Uniform$[-0.75, 0.75]$, satisfying Assumption 3(i). To mimic the infinite-dimensionality of functional data, each functional factor is generated by $f_{tj}(\cdot) = \sum_{i=1}^{50}\xi_{tji}\phi_i(\cdot)$ for $j \in [r]$ over $\mathcal{U} = [0,1]$, where $\{\phi_i(\cdot)\}_{i=1}^{50}$ is a 50-dimensional Fourier basis and basis coefficients $\boldsymbol{\xi}_{ti} = (\xi_{t1i},\ldots,\xi_{tri})^{\mathrm{T}}$ are generated from a vector autoregressive model, $\boldsymbol{\xi}_{ti} = \mathbf{A}\boldsymbol{\xi}_{t-1,i} + \mathbf{u}_{ti}$ with $\mathbf{A} = \{A_{jk} = 0.4^{|j-k|+1}\}_{r\times r}$, and the innovations $\{\mathbf{u}_{ti}\}_{t\in[n]}$ being sampled independently from $\mathcal{N}(\mathbf{0}_r, i^{-2}\mathbf{I}_r)$. For the second data-generating process (denoted as DGP2), we generate observed data from model (2), where r-vector of scalar factors $\boldsymbol{\gamma}_t$ is generated from a vector autoregressive model, $\boldsymbol{\gamma}_t = \mathbf{A}\boldsymbol{\gamma}_{t-1} + \mathbf{u}_t$ with $\{\mathbf{u}_t\}_{t\in[n]}$ being sampled independently from $\mathcal{N}(\mathbf{0}_r, \mathbf{I}_r)$. The functional loading matrix $\mathbf{Q}(\cdot) = \{Q_{jk}(\cdot)\}_{p\times r}$ is generated by $Q_{jk}(\cdot) = \sum_{i=1}^{50} i^{-1}q_{ijk}\phi_i(\cdot)$, where each q_{ijk} is sampled independently from the $\mathcal{N}(0, 0.3^2)$, satisfying Assumption 3'(i).

The idiosyncratic components are generated by $\boldsymbol{\varepsilon}_t(\cdot) = \sum_{l=1}^{25} 2^{-l/2}\boldsymbol{\psi}_{tl}\phi_l(\cdot)$, where each

$\boldsymbol{\psi}_{tl}$ is independently sampled from $\mathcal{N}(\mathbf{0}_p, \mathbf{C}_\zeta)$ with $\mathbf{C}_\zeta = \mathbf{D}\mathbf{C}_0\mathbf{D}$. Here, we set $\mathbf{D} = \text{diag}(D_1, \ldots, D_p)$, where each D_i is generated from Gamma$(3, 1)$. The generation of $\mathbf{C}_0$ involves the following three steps: (i) we set the diagonal entries of $\breve{\mathbf{C}}$ to 1, and generate the off-diagonal and symmetrical entries from Uniform$[0, 0.5]$; (ii) we employ hard thresholding (Cai and Liu, 2011) on $\breve{\mathbf{C}}$ to obtain a sparse matrix $\breve{\mathbf{C}}^{\mathcal{T}}$, where the threshold level is found as the smallest value such that $\max_{i\in[p]} \sum_{j=1}^{p} I(\breve{C}_{ij}^{\mathcal{T}} \neq 0) \leqslant p^{1-\alpha}$ for $\alpha \in [0, 1]$; (iii) we set $\mathbf{C}_0 = \breve{\mathbf{C}}^{\mathcal{T}} + \tilde{\delta}\mathbf{I}_p$ where $\tilde{\delta} = \max\{-\lambda_{\min}(\breve{\mathbf{C}}), 0\} + 0.01$ to guarantee the positive-definiteness of $\mathbf{C}_0$. The parameter α controls the sparsity level with larger values yielding sparser structures in $\mathbf{C}_0$ as well as functional sparser patterns in $\boldsymbol{\Sigma}_\varepsilon(\cdot, \cdot)$. This is implied from Proposition S.3(iii) of the Supplementary Material, whose parts (i) and (ii) respectively specify the true covariance matrix functions of $\mathbf{y}_t(\cdot)$ for DGP1 and DGP2.

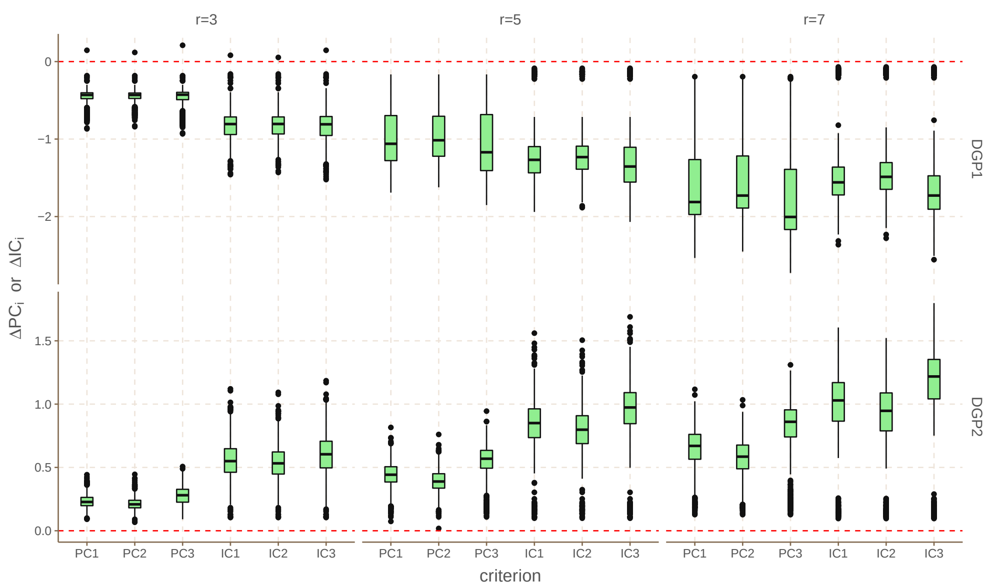

Figure 1: The boxplots of ΔPC_i and ΔIC_i ($i \in [3]$) for DGP1 and DGP2 with $p = 100, n = 100, \alpha = 0.5$, and $r = 3, 5, 7$ over 1000 simulation runs.

We firstly assess the finite-sample performance of the proposed information criteria in Section 2.4 under different combinations of p, n and α for DGP1 and DGP2. The results demonstrate that we can achieve almost 100% model selection accuracy in most cases. For instance, Figure 1 presents boxplots of ΔPC_i and ΔIC_i ($i = 1, 2, 3$) for two DGPs under the setting $p = 100, n = 100, \alpha = 0.5$, and $r = 3, 5, 7$. See also similar results for $p = 200, n = 50, \alpha = 0.5$ in Figure S.1 of the Supplementary Material. We observe that, for DGP1 (or DGP2), nearly all values of ΔPC_i and ΔIC_i are less than (or greater than) zero, indicating that the observed data are more likely to be generated by the correct model (1) (or model (2)). Furthermore, different penalty functions $g(n, p)$ have similar impacts on the information criteria when p and n are relatively large.

Table 1: The average relative frequency estimates for $\mathbb{P}(\hat{r} = r)$ over 1000 simulation runs.

			$r=3$		$r=5$		$r=7$	
α	p	n	$\mathbb{P}(\hat{r}^{\mathcal{D}} = r)$	$\mathbb{P}(\hat{r}^{\mathcal{F}} = r)$	$\mathbb{P}(\hat{r}^{\mathcal{D}} = r)$	$\mathbb{P}(\hat{r}^{\mathcal{F}} = r)$	$\mathbb{P}(\hat{r}^{\mathcal{D}} = r)$	$\mathbb{P}(\hat{r}^{\mathcal{F}} = r)$
0.25	100	100	0.854	0.828	0.762	0.715	0.618	0.597
		200	0.862	0.853	0.806	0.803	0.733	0.733
	200	100	0.922	0.868	0.832	0.792	0.739	0.667
		200	0.924	0.905	0.896	0.853	0.816	0.746
0.50	100	100	0.958	0.973	0.931	0.932	0.896	0.890
		200	0.960	0.974	0.952	0.950	0.936	0.943
	200	100	0.991	0.987	0.977	0.972	0.956	0.957
		200	0.991	0.993	0.984	0.985	0.979	0.971
0.75	100	100	0.990	0.998	0.986	0.991	0.979	0.976
		200	0.996	0.994	0.986	0.992	0.984	0.994
	200	100	0.997	1.000	0.998	1.000	0.995	0.999
		200	0.999	1.000	1.000	1.000	0.997	1.000

Once the more appropriate FFM is selected based on observed data, our next step adopts the ratio-based estimator (18) (or (19)) to determine the number of functional (or scalar) factors. The performance of proposed estimators is then examined in terms of their abilities to correctly identify the number of factors. Table 1 reports average relative frequencies $\hat{r} = r$ under different combinations of $r = 3, 5, 7$, $n = 100, 200$, $p = 100, 200$ and $\alpha = 0.25, 0.5, 0.75$ for both DGPs. Several conclusions can be drawn. First, for fixed p and n, larger values of α lead to improved accuracies in identifying r as the strength of factors (i.e. signal-to-noise ratio, see Proposition S.3(iii) of the Supplementary Material) increases. Second, we observe the phenomenon of "blessing of dimensionality" in the sense that the estimation improves as p increases, which is due to the increased information from added components on the factors.

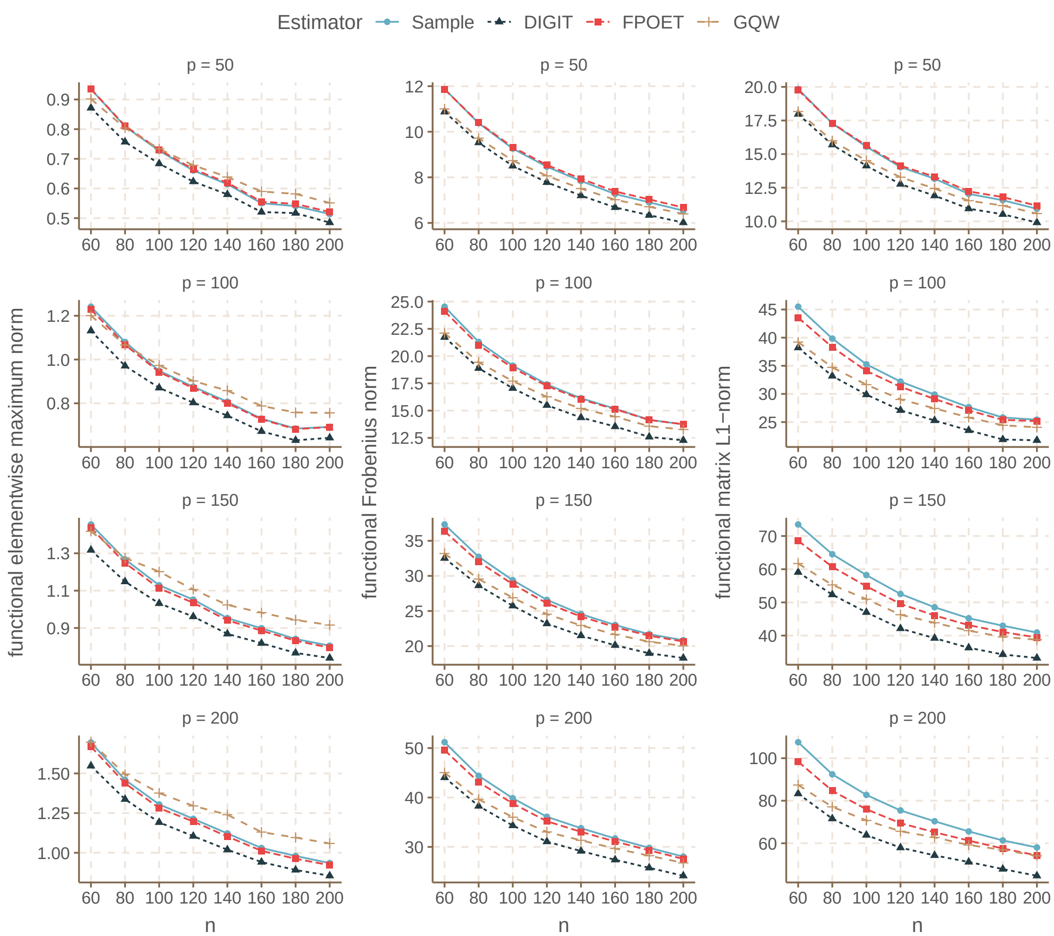

Figure 2: The average losses of $\widehat{\mathbf{\Sigma}}_y$ in functional elementwise ℓ_∞ norm (left column), Frobenius norm (middle column) and matrix ℓ_1 norm (right column) for DGP1 over 1000 simulation runs.

We next compare our proposed AFT estimator in (10) with two related methods for estimating the idiosyncratic covariance $\mathbf{\Sigma}_\varepsilon$, where the details can be found in Section F of the Supplementary Material. Following Fan et al. (2013), the threshold level for AFT is selected as $\lambda = \dot{C}(\sqrt{\log p/n} + 1/\sqrt{p})$ with $\dot{C} = 0.5$. We also implemented the cross-validation method to choose $\dot{C}$. However, such method incurred heavy computational costs and only gave a very slight improvement. We finally compare our DIGIT and FPOET estimators with two competing methods for estimating the covariance $\mathbf{\Sigma}_y$. The first competitor is the sample covariance estimator $\widehat{\mathbf{\Sigma}}_y^{\mathcal{S}}$. For comparison, we also implement the method of Guo, Qiao and Wang (2022) in conjunction with our AFT (denoted as GQW). This combined method firstly employs autocovariance-based eigenanalysis to estimate $\mathbf{B}$ and then follows the similar procedure as DIGIT to estimate $\mathbf{f}_t(\cdot)$ and $\mathbf{\Sigma}_\varepsilon$. Although DIGIT and GQW estimators (or FPOET estimator) are specifically developed to fit model (1) (or model (2)), we also use them (or it) for estimating $\mathbf{\Sigma}_y$ under DGP2 (or DGP1) to evaluate the robustness of each proposal under model misspecification. For both DGPs, we set $\alpha = 0.5$ and generate $n = 60, 80, \dots, 200$ observations of $p = 50, 100, 150, 200$ functional variables. We adopt the eigenvalue-ratio-based method to determine r. Figures 2 and 3 display the numerical summaries of losses measured by functional versions of elementwise ℓ_∞ norm, Frobenius norm, and matrix ℓ_1 norm for DGP1 and DGP2, respectively.

A few trends are observable. First, for DGP1 (or DGP2) in Figure 2 (or Figure 3), the DIGIT (or FPOET) estimator outperforms the three competitors under almost all functional matrix losses and settings we consider. In high-dimensional large p scenarios, the factor-guided estimators lead to more competitive performance, whereas the results of $\widehat{\mathbf{\Sigma}}_y^{\mathcal{S}}$ severely deteriorate especially in terms of functional matrix ℓ_1 loss. Second, although both DIGIT and GQW estimators are developed to estimate model (1) and the idiosyncratic components are generated from a white noise process, our proposed DIGIT estimator is prominently superior to the GQW estimator for DGP1 under all scenarios, as seen in Figure 2. This demonstrates the advantage of covariance-based DIGIT over autocovariance-based GQW when the factors are pervasive (i.e. strong), however, DIGIT may not perform well in the presence of weak factors. Third, the FPOET estimator exhibits enhanced robustness compared to DIGIT and GQW estimators in the case of model misspecification. In particular, for DGP2, DIGIT and GQW show substantial decline in performance measured by functional Frobenius and matrix ℓ_1 losses, while, for DGP1, FPOET still achieves reasonably good performance. This suggests a preference for FPOET over DIGIT when the model form cannot be determined

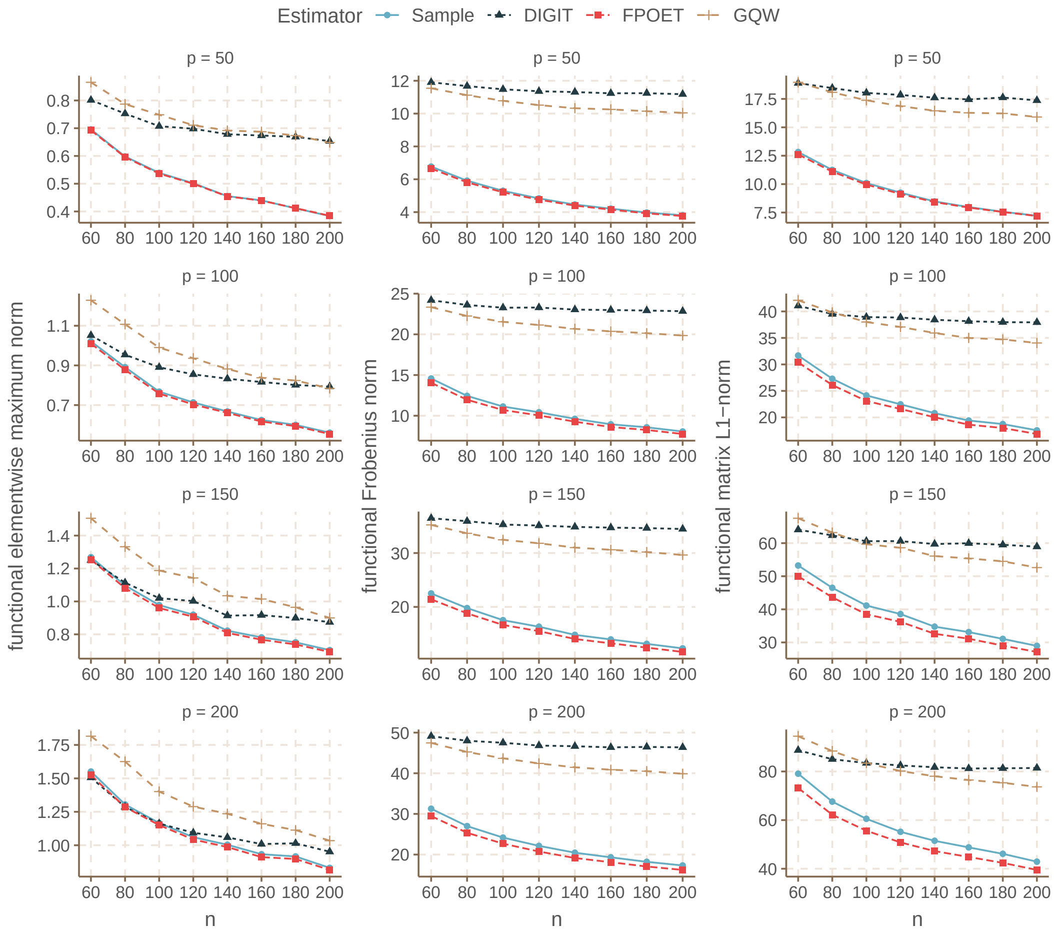

Figure 3: The average losses of $\widehat{\boldsymbol{\Sigma}}_y$ in functional elementwise ℓ_∞ norm (left column), Frobenius norm (middle column) and matrix ℓ_1 norm (right column) for DGP2 over 1000 simulation runs.

confidently (i.e. information criteria between two FFMs are relatively close).

6 Real data analysis

6.1 Age-specific mortality data

Our first dataset, available at https://www.mortality.org/, contains age- and gender-specific mortality rates for $p = 32$ countries from 1960 to 2013 ($n = 54$). Following Tang et al. (2022) which also analyzed such dataset, we apply a log transformation to mortality rates and let $y_{ti}(u_k)$ ($t \in [n], i \in [p], k \in [101]$) be the log mortality rate of people aged $u_k = k-1 \in \mathcal{U} =$

$[0, 100]$ living in the i-th country during the year $1959+t$. The observed curves are smoothed based on a 10-dimensional B-spline basis. Figure S.9 of the Supplementary Material displays rainbow plots (Hyndman and Shang, 2010) of the smoothed log-mortality rates for females in six randomly selected countries, which use different colors to represent the curves from earlier years (in red) to more recent years (in purple). We observe a similar pattern for the USA, the U.K., and Austria, with their curves being more dispersed, indicating a uniform decline in mortality over time. However, this pattern differs significantly from those for Russia, Ukraine, and Belarus, where the decreasing effect disappears, and the curves are more concentrated. It is also worth mentioning that the U.K. and Austria are far from the USA, but Austria is closer to Russia, Ukraine, and Belarus. This phenomenon inspires us to employ multivariate functional time series methods, such as two FFMs, instead of spatial-temporal models that typically rely on geographical distance as the similarity measure.

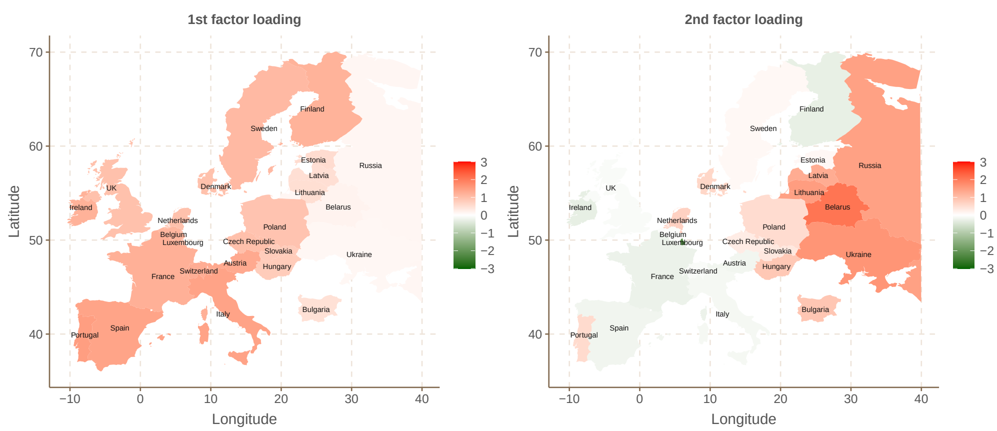

Figure 4: Spatial heatmaps of factor loadings of some European countries for females.

For model selection, we calculate the information criteria with $\mathrm{PC}_1^{\mathcal{D}} = 0.168 < \mathrm{PC}_1^{\mathcal{F}} = 0.177$, and $\mathrm{IC}_1^{\mathcal{D}} = -3.806 < \mathrm{IC}_1^{\mathcal{F}} = -3.448$. Therefore, we choose FFM (1) with age-specific factors for estimation. The leading two eigenvalues of $\widehat{\mathbf{\Omega}}$ in (8) are much larger than the rest with cumulative percentage exceeding 90%, so we choose $\hat{r}^{\mathcal{D}} = 2$ for illustration. Figures 4 and S.10 of the Supplementary Material present spatial heatmaps of factor loading of some European countries for females and males, respectively. It is apparent that the first factor mainly influences the regions of Western and Northern Europe, such as Italy, the U.K., Spain, and Sweden, while the former Soviet Union member countries such as Russia, Ukraine, Belarus, and Lithuania are heavily loaded regions on the second factor. Additionally, countries like Poland and Hungary that have experienced ideological shifts have non-negligible loadings on both factors. For countries far from Europe, such as the USA, Australia, and Japan, the first factor also serves as the main driving force.

Figures 5 and S.11 of the Supplementary Material provide the rainbow plots of the estimated age-specific factors for females and males, respectively. We observe that, for the first factor of female mortality rates, the curves of more recent years mostly lie below the curves of earlier years. This suggests a consistent improvement in mortality rates across all ages over the years. However, for the second factor, the curves of more recent years are located below the curves of earlier years when $u \leqslant 30$, and above them when $u > 30$, implying a downward trend under age 30 and an upward trend over age 30. Similar conclusions can be drawn for male mortality rates. By applying our factor-guided approach for multivariate functional time series, we achieve clustering results that are comparable to those obtained by Tang et al. (2022).

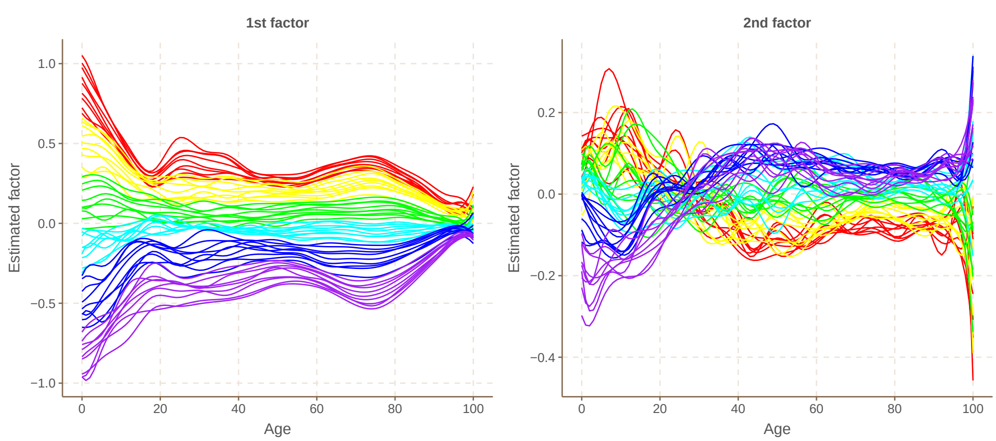

Figure 5: The estimated age-specific factors for females.

6.2 Cumulative intraday return data

Our second dataset, collected from https://wrds-www.wharton.upenn.edu/, consists of high-frequency observations of prices for a collection of S&P100 stocks from 251 trading days in the year 2017. We removed 2 stocks with missing data so $p = 98$ in our analysis. We obtain five-minute resolution prices by using the last transaction price in each five-minute interval after removing the outliers, and hence convert the trading period (9:30–16:00) to minutes $[0, 78]$. We construct CIDR (Horváth et al., 2014) trajectories, in percentage, by $y_{ti}(u_k) = 100[\log\{P_{ti}(u_k)\} - \log\{P_{ti}(u_1)\}]$, where $P_{ti}(u_k)$ ($t \in [n], i \in [p], k \in [78]$) denotes the price of the i-th stock at the k-th five-minute interval after the opening time on the t-th trading day. We obtain smoothed CIDR curves by expanding the data using a 10-dimensional B-spline basis. The CIDR curves, which always start from zero, not only have nearly the same shape as the original price curves but also enhance the plausibility of the stationarity assumption. We performed functional KPSS test (Horváth et al., 2014) for each stock, and found no overwhelming evidence (under 1% significance level) against the stationarity.

For model selection, the information criteria $\text{PC}_1^{\mathcal{D}} = 0.567 > \text{PC}_1^{\mathcal{F}} = 0.558$, and $\text{IC}_1^{\mathcal{D}} = -0.619 > \text{IC}_1^{\mathcal{F}} = -0.640$. These values suggest that FFM (2) is slightly more preferable and imply that the latent factors may not exhibit any intraday varying patterns. We consider the problem of functional risk management as discussed in Section 4.1. Our task is to obtain the optimal functional portfolio allocation $\widehat{\mathbf{w}}(\cdot)$ by minimizing the perceived risk of the functional portfolio, specifically,

$$\widehat{\mathbf{w}} = \arg\min_{\mathbf{w} \in \mathbb{H}^p} \left\langle \mathbf{w}, \widehat{\boldsymbol{\Sigma}}_y(\mathbf{w}) \right\rangle \text{ subject to } \mathbf{w}(u)^{\mathrm{T}} \mathbf{1}_p = 1 \text{ for any } u \in \mathcal{U},$$

where $\mathbf{1}_p = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^p$. Following the derivations in Section E.4 of the Supplementary Material, we obtain the solution:

$$\widehat{\mathbf{w}}(u) = \int\int \widehat{\boldsymbol{\Sigma}}_y^{-1}(u, v) \text{diag}\{H^{-1}(v, z), \cdots, H^{-1}(v, z)\} \mathbf{1}_p \text{d}v \text{d}z$$

with $H(\cdot, \cdot) = \mathbf{1}_p^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_y^{-1}(\cdot, \cdot) \mathbf{1}_p$, which allows us to obtain the actual risk. In practical implementation, we treat components of $\mathbf{y}_t(\cdot)$ as finite-dimensional functional objects and hence can obtain bounded inverse $\widehat{\boldsymbol{\Sigma}}_y^{-1}$ (and H^{-1}) using the leading eigenpairs of $\widehat{\boldsymbol{\Sigma}}_y$ (and H) such that the corresponding cumulative percentage of selected eigenvalues exceeds 95%.

Table 2: Comparisons of the risks of functional portfolios obtained by using DIGIT, FPOET, GQW, and sample estimators.

Estimator	DIGIT			FPOET			GQW			Sample
Month	$\hat{r}=1$	$\hat{r}=2$	$\hat{r}=3$	$\hat{r}=1$	$\hat{r}=2$	$\hat{r}=3$	$\hat{r}=1$	$\hat{r}=2$	$\hat{r}=3$	
July	0.052	0.060	0.058	0.057	0.060	0.057	0.083	0.062	0.052	0.099
August	0.044	0.045	0.048	0.045	0.044	0.049	0.050	0.089	0.085	0.092
September	0.092	0.051	0.065	0.093	0.053	0.058	0.108	0.056	0.060	0.097
October	0.077	0.045	0.042	0.079	0.044	0.041	0.082	0.067	0.051	0.086
November	0.078	0.060	0.043	0.079	0.063	0.045	0.063	0.073	0.076	0.090
December	0.075	0.075	0.043	0.077	0.079	0.042	0.083	0.079	0.095	0.091
Average	0.070	0.056	0.050	0.072	0.057	0.049	0.078	0.071	0.070	0.093

Following the procedure in Fan et al. (2013), on the 1st trading day of each month from July to December, we estimate $\widehat{\boldsymbol{\Sigma}}_y$ using DIGIT, FPOET, GQW and sample estimators based on the historical data comprising CIDR curves of 98 stocks for the preceding 6 months ($n = 126$). We then determine the corresponding optimal portfolio allocation $\widehat{\mathbf{w}}(u_k)$ for $k \in [78]$. At the end of the month after 21 trading days, we compare actual risks calculated by $78^{-2}\sum_{k,k'\in[78]} \widehat{\mathbf{w}}(u_k)^{\mathrm{T}}\{21^{-1}\sum_{t=1}^{21} \mathbf{y}_t(u_k)\mathbf{y}_t(v_{k'})^{\mathrm{T}}\}\widehat{\mathbf{w}}(v_{k'})$. Following Fan et al. (2013) and Wang et al. (2021), we try $\hat{r} = 1, 2$ and 3 to check the effect of r in out-of-sample performance. The numerical results are summarized in Table 2. We observe that the minimum risk functional portfolio created by DIGIT, FPOET, and GQW result in averaged risks over six months as 0.05, 0.049, and 0.07, respectively, while the sample covariance estimator gives 0.093. The risk has been significantly reduced by around 46% using our factor-guided approach.

References

Bai, J. (2003). Inferential theory for factor models of large dimensions, *Econometrica* **71**: 135–171.

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

Baker, C. R. (1973). Joint measures and cross-covariance operators, *Transactions of the American Mathematical Society* **186**: 273–289.

Bathia, N., Yao, Q. and Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series, *The Annals of Statistics* **38**: 3352–3386.

Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation, *Journal of the American Statistical Association* **106**: 672–684.

Carmeli, C., De Vito, E. and Toigo, A. (2006). Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem, *Analysis and Applications* **4**: 377–408.

Chiou, J.-M., Yang, Y.-F. and Chen, Y.-T. (2016). Multivariate functional linear regression and prediction, *Journal of Multivariate Analysis* **146**: 301–312.

Fan, J., Fan, Y. and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model, *Journal of Econometrics* **147**: 186–197.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association* **96**: 1348–1360.

Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements, *Journal of the Royal Statistical Society: Series B* **75**: 603–680.

Fan, J., Liu, H. and Wang, W. (2018). Large covariance estimation through elliptical factor models, *The Annals of Statistics* **46**: 1383.

Fang, Q., Guo, S. and Qiao, X. (2023). Adaptive functional thresholding for sparse covariance function estimation in high dimensions, *Journal of the American Statistical Association, in press* .

Guo, S. and Qiao, X. (2023). On consistency and sparsity for high-dimensional functional time series with application to autoregressions, *Bernoulli* **29**: 451–472.

Guo, S., Qiao, X. and Wang, Q. (2022). Factor modelling for high-dimensional functional time series, *arXiv:2112.13651v2* .

Hallin, M., Nisol, G. and Tavakoli, S. (2023). Factor models for high-dimensional functional time series I: Representation results, *Journal of Time Series Analysis, in press* .

Happ, C. and Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains, *Journal of the American Statistical Association* **113**: 649–659.

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series, *Journal of Econometrics* **179**: 66–82.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, John Wiley & Sons, Chichester.

Hyndman, R. J. and Shang, H. L. (2010). Rainbow plots, bagplots, and boxplots for functional data, *Journal of Computational and Graphical Statistics* **19**: 29–45.

Lam, C. and Yao, Q. (2012). Factor modelling for high-dimensional time series: Inference for the number of factors, *The Annals of Statistics* **40**: 694–726.

Qiao, X., Guo, S. and James, G. (2019). Functional graphical models, *Journal of the American Statistical Association* **114**: 211–222.

Solea, E. and Li, B. (2022). Copula gaussian graphical models for functional data, *Journal of the American Statistical Association* **117**: 781–793.

Tang, C., Shang, H. L. and Yang, Y. (2022). Clustering and forecasting multiple functional time series, *The Annals of Applied Statistics* **16**: 2523–2553.

Tavakoli, S., Nisol, G. and Hallin, M. (2023). Factor models for high-dimensional functional time series II: Estimation and forecasting, *Journal of Time Series Analysis, in press* .

Wang, H., Peng, B., Li, D. and Leng, C. (2021). Nonparametric estimation of large covariance matrices with conditional sparsity, *Journal of Econometrics* **223**: 53–72.

Wang, W. and Fan, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance, *The Annals of Statistics* **45**: 1342.

Xue, K., Yang, J. and Yao, F. (2023). Optimal linear discriminant analysis for high-dimensional functional data, *Journal of the American Statistical Association,in press* .

Zapata, J., Oh, S. Y. and Petersen, A. (2022). Partial separability and functional graphical models for multivariate Gaussian processes, *Biometrika* **109**: 665–681.

Zou, H. (2006). The adaptive lasso and its oracle properties, *Journal of the American Statistical Association* **101**: 1418–1429.

Supplementary Material to "Factor-guided estimation of large covariance matrix function with conditional functional sparsity"

Dong Li, Xinghao Qiao and Zihan Wang

This supplementary material contains technical proofs in Sections A–D, further derivations in Section E, additional simulation results in Section F and additional real data results in Section G. Throughout, we denote the multiplications of matrix kernel functions as follows. For $\mathbf{K}, \mathbf{G} \in \mathbb{H}^p \otimes \mathbb{H}^p$, we write $\mathbf{M} = \mathbf{K}\mathbf{G} \in \mathbb{H}^p \otimes \mathbb{H}^p$, where

$$\mathbf{M}(u, v) = \int_{\mathcal{U}} \mathbf{K}(u, w)\mathbf{G}(w, v)\mathrm{d}w. \tag{S.1}$$

A Proofs of theoretical results in Section 2

A.1 Technical lemmas and their proofs

We first introduce useful theorems to prove Proposition 1. In the following two lemmas, $\{\lambda_j\}_{j\in[p]}$ are the eigenvalues of $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ in a descending order and $\{\boldsymbol{\xi}_j\}_{j\in[p]}$ are the corresponding eigenvectors. Similarly, $\{\widetilde{\lambda}_j\}_{j\in[p]}$ and $\{\widetilde{\boldsymbol{\xi}}_j\}_{j\in[p]}$ are the corresponding eigenvalues and eigenvectors of $\widetilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{p\times p}$, respectively.

Lemma A1. *(Weyl's theorem; Weyl (1912)).* $|\widetilde{\lambda}_j - \lambda_j| \leqslant \|\widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|$ *for* $j \in [p]$.

Lemma A2. *(A useful variant of* $\sin(\theta)$ *theorem; Davis and Kahan (1970); Yu et al. (2015)). If* $\widetilde{\boldsymbol{\xi}}_j^{\mathrm{T}}\boldsymbol{\xi}_j \geqslant 0$ *for* $j \in [p]$*, then*

$$\|\widetilde{\boldsymbol{\xi}}_j - \boldsymbol{\xi}_j\| \leqslant \frac{\|\widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|/\sqrt{2}}{\min\left(|\widetilde{\lambda}_{j-1} - \lambda_j|, |\lambda_j - \widetilde{\lambda}_{j+1}|\right)}.$$

The functional version of Weyl's theorem has been studied in Lemmas 4.2 and 4.3 of Bosq (2000). Let $\{\tau_i\}_{i=1}^{\infty}$ be the eigenvalues of the kernel function $\boldsymbol{\Sigma}(\cdot,\cdot)$ in a descending order and $\{\boldsymbol{\varphi}_i(\cdot)\}_{i=1}^{\infty}$ are the corresponding eigenfunctions. Similary, $\{\widetilde{\tau}_i\}_{i=1}^{\infty}$ and $\{\widetilde{\boldsymbol{\varphi}}_i(\cdot)\}_{i=1}^{\infty}$ are the corresponding eigenvalues and eigenfunctions of $\widetilde{\boldsymbol{\Sigma}}(\cdot,\cdot)$, respectively.

Lemma A3. *(Lemma 4.2 in Bosq (2000)).* $|\widetilde{\tau}_i - \tau_i| \leqslant \|\widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathcal{L}}$ *for all* i.

Lemma A4. *(Lemma 4.3 in Bosq (2000)). If* $\langle \widetilde{\boldsymbol{\varphi}}_i, \boldsymbol{\varphi}_i \rangle \geqslant 0$, *then*

$$\|\widetilde{\boldsymbol{\varphi}}_i - \boldsymbol{\varphi}_i\| \leqslant \frac{2\sqrt{2}\|\widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathcal{L}}}{\min\big(|\widetilde{\tau}_{i-1} - \tau_i|, |\tau_i - \widetilde{\tau}_{i+1}|\big)}.$$

The following lemmas introduce some functional norm inequalities, which are useful in subsequent proofs.

Lemma A5. *Suppose that* $\mathbf{K} \in \mathbb{H}^p \otimes \mathbb{H}^p$ *is a Mercer's kernel with the spectral decomposition* $\mathbf{K}(u,v) = \sum_{i=1}^{\infty} \lambda_i \boldsymbol{\phi}_i(u)\boldsymbol{\phi}_i(v)^{\mathrm{T}}$, *where* $\{\lambda_i\}_{i=1}^{\infty}$ *are the eigenvalues of* $\mathbf{K}$ *in a descending order and* $\{\boldsymbol{\phi}_i(\cdot)\}$ *are the corresponding eigenfunctions. Then, we have*

(i) $\operatorname{tr}\left\{\int\int \mathbf{K}(u,v)\mathbf{K}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\} = \operatorname{tr}\{\int \mathbf{K}\mathbf{K}^{\mathrm{T}}(u,u)\mathrm{d}u\} = \|\mathbf{K}\|_{\mathcal{S},\mathrm{F}}^2 = \sum_{i=1}^{\infty}\lambda_i^2$;

(ii) $\|\int\int \mathbf{K}(u,v)\mathbf{K}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\| = \|\int \mathbf{K}\mathbf{K}^{\mathrm{T}}(u,u)\mathrm{d}u\| = \|\mathbf{K}\|_{\mathcal{L}}^2 = \lambda_1^2$.

Proof. (i) Note that $\int\int \mathbf{K}(u,v)\mathbf{K}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v = \int \sum_{i=1}^{\infty} \lambda_i^2 \boldsymbol{\phi}_i(u)\boldsymbol{\phi}_i(u)^{\mathrm{T}}\mathrm{d}u$, and thus

$$\operatorname{tr}\left\{\int\int \mathbf{K}(u,v)\mathbf{K}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\} = \operatorname{tr}\left\{\sum_{i=1}^{\infty}\lambda_i^2 \int \boldsymbol{\phi}_i(u)^{\mathrm{T}}\boldsymbol{\phi}_i(u)\mathrm{d}u\right\} = \sum_{i=1}^{\infty}\lambda_i^2.$$

The equality $\operatorname{tr}\left\{\int\int \mathbf{K}(u,v)\mathbf{K}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\} = \|\mathbf{K}\|_{\mathcal{S},\mathrm{F}}^2$ can be verified by simple calculation. The first equality can be obtained by $\mathbf{K}(u,v)^{\mathrm{T}} = \mathbf{K}^{\mathrm{T}}(v,u)$ and the multiplication of kernel functions defined in (S.1).

(ii) Similarly,

$$\begin{aligned}\left\|\int\int \mathbf{K}(u,v)\mathbf{K}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\| =& \left\|\int \sum_{i=1}^{\infty}\lambda_i^2\boldsymbol{\phi}_i(u)\boldsymbol{\phi}_i(u)^{\mathrm{T}}\mathrm{d}u\right\| = \lambda_{\max}\left\{\int \sum_{i=1}^{\infty}\lambda_i^2\boldsymbol{\phi}_i(u)\boldsymbol{\phi}_i(u)^{\mathrm{T}}\mathrm{d}u\right\}\\ =&\lambda_{\max}\left\{\boldsymbol{\Lambda}^2\int \boldsymbol{\Phi}(u)\boldsymbol{\Phi}(u)^{\mathrm{T}}\mathrm{d}u\right\} = \lambda_{\max}\left\{\boldsymbol{\Lambda}^2\int \boldsymbol{\Phi}(u)^{\mathrm{T}}\boldsymbol{\Phi}(u)\mathrm{d}u\right\}\\ =&\lambda_{\max}(\boldsymbol{\Lambda}^2) = \lambda_1^2 = \|\mathbf{K}\|_{\mathcal{L}}^2,\end{aligned}$$

where $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots)$, $\boldsymbol{\Phi}(\cdot) = \{\boldsymbol{\phi}_1(\cdot), \boldsymbol{\phi}_2(\cdot), \ldots\}$, and the fact that the $\infty \times \infty$ matrix $\int \boldsymbol{\Phi}(u)^{\mathrm{T}}\boldsymbol{\Phi}(u)\mathrm{d}u$ shares the same nonzero eigenvalues with the $p \times p$ matrix $\int \boldsymbol{\Phi}(u)\boldsymbol{\Phi}(u)^{\mathrm{T}}\mathrm{d}u$, which can be obtained following the proof of Proposition 2 in Bathia et al. (2010). □

Lemma A6. *Suppose that* $\mathbf{K}, \mathbf{G} \in \mathbb{H}^p \otimes \mathbb{H}^p$ *are Mercer's kernels, then we have*

(i) $\big\|\int\int \mathbf{K}(u,v)\mathbf{G}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\big\|_1 \leqslant \|\mathbf{K}\|_{\mathcal{S},1}\|\mathbf{G}\|_{\mathcal{S},\infty}$*;*

(ii) $\big\|\int\int \mathbf{K}(u,v)\mathbf{G}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\big\|_\infty \leqslant \|\mathbf{K}\|_{\mathcal{S},\infty}\|\mathbf{G}\|_{\mathcal{S},1}$*;*

(iii) $\big\|\int\int \mathbf{K}(u,v)\mathbf{G}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\big\| \leqslant \left(\|\mathbf{K}\|_{\mathcal{S},\infty}\|\mathbf{K}\|_{\mathcal{S},1}\right)^{1/2}\left(\|\mathbf{G}\|_{\mathcal{S},\infty}\|\mathbf{G}\|_{\mathcal{S},1}\right)^{1/2}$*;*

(iv) $\big\|\int\int \mathbf{K}(u,v)\mathbf{G}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\big\| \leqslant \big\{\big\|\int\int \mathbf{K}(u,v)\mathbf{K}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\big\|\big\}^{1/2}\big\{\big\|\int\int \mathbf{G}(u,v)\mathbf{G}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\big\|\big\}^{1/2}$.

Proof. (i) Note that

$$
\begin{aligned}
&\Big\|\int\int \mathbf{K}(u,v)\mathbf{G}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\Big\|_1 \\
=&\max_{j\in[p]}\sum_{i=1}^p\left|\int\int\sum_{k=1}^p K_{ik}(u,v)G_{jk}(u,v)\mathrm{d}u\mathrm{d}v\right| \\
\leqslant&\max_{j\in[p]}\sum_{i=1}^p\sum_{k=1}^p\|K_{ik}\|_{\mathcal{S}}\|G_{jk}\|_{\mathcal{S}} \qquad\qquad (S.2)\\
\leqslant&\Big(\max_{k\in[p]}\sum_{i=1}^p\|K_{ik}\|_{\mathcal{S}}\Big)\Big(\max_{j\in[p]}\sum_{k=1}^p\|G_{jk}\|_{\mathcal{S}}\Big) \\
=&\|\mathbf{K}\|_{\mathcal{S},1}\|\mathbf{G}\|_{\mathcal{S},\infty}.
\end{aligned}
$$

(ii) By similar arguments, we obtain that

$$
\Big\|\int\int \mathbf{K}(u,v)\mathbf{G}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\Big\|_\infty \leqslant \|\mathbf{K}\|_{\mathcal{S},\infty}\|\mathbf{G}\|_{\mathcal{S},1}. \qquad (S.3)
$$

(iii) The inequality follows immediately from (S.2), (S.3), the matrix norm inequality $\|\mathbf{A}\|^2 \leqslant \|\mathbf{A}\|_\infty\|\mathbf{A}\|_1$ for any $p\times p$ matrix $\mathbf{A}$ and the choice of $\mathbf{A} = \int\int \mathbf{K}(u,v)\mathbf{G}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v$.

(iv) An application of Hölder's inequality yields the result. □

Lemma A7. *Suppose that* $\mathbf{\Sigma} = \{\Sigma_{ij}(\cdot,\cdot)\}_{p\times p}$ *with* $\Sigma_{ij} \in \mathbb{S}$ *and* $\widetilde{\mathbf{\Sigma}} \in \mathbb{H}^p \otimes \mathbb{H}^p$ *are Mercer's kernels. Then we have (i)* $\|\mathbf{\Sigma}\widetilde{\mathbf{\Sigma}}\|_{\mathcal{L}} \leqslant \|\mathbf{\Sigma}\|_{\mathcal{L}}\cdot\|\widetilde{\mathbf{\Sigma}}\|_{\mathcal{L}}$*, (ii)* $\|\mathbf{\Sigma}\|_{\mathcal{L}} \leqslant \|\mathbf{\Sigma}\|_{\mathcal{S},\mathrm{F}}$*, and (iii)* $\|\mathbf{\Sigma}\|_{\mathcal{L}} \leqslant \|\mathbf{\Sigma}\|_{\mathcal{S},1}^{1/2}\|\mathbf{\Sigma}\|_{\mathcal{S},\infty}^{1/2}$*. Furthermore, if* $\|\Sigma_{ij}\|_{\mathcal{S}} = \|\Sigma_{ji}\|_{\mathcal{S}}$ *for all* $i,j\in[p]$*, then* $\|\mathbf{\Sigma}\|_{\mathcal{L}} \leqslant \|\mathbf{\Sigma}\|_{\mathcal{S},1}$*.*

Proof. (i) By Lemma B14 or Theorem 4.2.5 in Hsing and Eubank (2015), we have

$$\|\boldsymbol{\Sigma}\widetilde{\boldsymbol{\Sigma}}\|_{\mathcal{L}} = \max_{\mathbf{x}\in\mathbb{H}^p} \frac{\big\langle \mathbf{x}, \boldsymbol{\Sigma}\widetilde{\boldsymbol{\Sigma}}(\mathbf{x})\big\rangle}{\|\mathbf{x}\|^2} = \max_{\mathbf{x}\in\mathbb{H}^p} \frac{\Big\langle \widetilde{\boldsymbol{\Sigma}}^{1/2}(\mathbf{x}), \boldsymbol{\Sigma}\big\{\widetilde{\boldsymbol{\Sigma}}^{1/2}(\mathbf{x})\big\}\Big\rangle}{\Big\|\widetilde{\boldsymbol{\Sigma}}^{1/2}(\mathbf{x})\Big\|^2} \cdot \frac{\big\langle \mathbf{x}, \widetilde{\boldsymbol{\Sigma}}(\mathbf{x})\big\rangle}{\|\mathbf{x}\|^2}$$

$$\leqslant \max_{\mathbf{y}\in\mathbb{H}^p} \frac{\big\langle \mathbf{y}, \boldsymbol{\Sigma}(\mathbf{y})\big\rangle}{\|\mathbf{y}\|^2} \max_{\mathbf{x}\in\mathbb{H}^p} \frac{\big\langle \mathbf{x}, \widetilde{\boldsymbol{\Sigma}}(\mathbf{x})\big\rangle}{\|\mathbf{x}\|^2} = \|\boldsymbol{\Sigma}\|_{\mathcal{L}} \cdot \|\widetilde{\boldsymbol{\Sigma}}\|_{\mathcal{L}}.$$

(ii) By Lemma A5, $\|\boldsymbol{\Sigma}\|_{\mathcal{L}} = \lambda_1$ and $\|\boldsymbol{\Sigma}\|_{\mathrm{F}} = \sqrt{\sum_{i\geqslant 1}\lambda_i^2}$, where $\{\lambda_i\}_{i=1}^{\infty}$ are the eigenvalues of $\boldsymbol{\Sigma}$ in a descending order. Apparently, $\|\boldsymbol{\Sigma}\|_{\mathcal{L}} \leqslant \|\boldsymbol{\Sigma}\|_{\mathcal{S},\mathrm{F}}$ holds.

(iii) By Lemmas A5(ii) and A6(iii), $\|\boldsymbol{\Sigma}\|_{\mathcal{L}}^2 = \big\|\int\int \boldsymbol{\Sigma}(u,v)\boldsymbol{\Sigma}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\big\| \leqslant \|\boldsymbol{\Sigma}\|_{\mathcal{S},1}\|\boldsymbol{\Sigma}\|_{\mathcal{S},\infty}$. Furthermore, if $\|\Sigma_{ij}\|_{\mathcal{S}} = \|\Sigma_{ji}\|_{\mathcal{S}}$ for all $i,j \in [p]$, we have $\|\boldsymbol{\Sigma}\|_{\mathcal{S},1} = \|\boldsymbol{\Sigma}\|_{\mathcal{S},\infty}$, and thus $\|\boldsymbol{\Sigma}\|_{\mathcal{L}} \leqslant \|\boldsymbol{\Sigma}\|_{\mathcal{S},1}$ holds. □

Lemma A8. *Suppose that $\mathbf{K},\mathbf{G} \in \mathbb{H}^p \otimes \mathbb{H}^p$ are Mercer's kernels, then we have*

(i) $\mathrm{tr}\{\int \mathbf{K}\mathbf{G}(u,u)\mathrm{d}u\} = \mathrm{tr}\{\int \mathbf{G}\mathbf{K}(u,u)\mathrm{d}u\}$, *i.e.,* $\|\mathbf{K}\mathbf{G}\|_{\mathcal{N}} = \|\mathbf{G}\mathbf{K}\|_{\mathcal{N}}$*;*

(ii) $\mathrm{tr}\{\int \mathbf{K}\mathbf{G}(u,u)\mathrm{d}u\} \leqslant \|\mathbf{K}\|_{\mathcal{L}}\mathrm{tr}\{\int \mathbf{G}(u,u)\mathrm{d}u\}$, *i.e.,* $\|\mathbf{K}\mathbf{G}\|_{\mathcal{N}} \leqslant \|\mathbf{K}\|_{\mathcal{L}}\|\mathbf{G}\|_{\mathcal{N}}$*;*

(iii) $\|\mathbf{K}\mathbf{G}\|_{\mathcal{S},\mathrm{F}} \leqslant \|\mathbf{K}\|_{\mathcal{L}}\|\mathbf{G}\|_{\mathcal{S},\mathrm{F}}$.

Proof. (i) Note that

$$\begin{aligned}\mathrm{tr}\Big\{\int \mathbf{K}\mathbf{G}(u,u)\mathrm{d}u\Big\} &= \mathrm{tr}\Big\{\int\int \mathbf{K}(u,v)\mathbf{G}(v,u)\mathrm{d}u\mathrm{d}v\Big\} = \int\int \mathrm{tr}\{\mathbf{K}(u,v)\mathbf{G}(v,u)\}\mathrm{d}u\mathrm{d}v \\ &= \mathrm{tr}\Big\{\int\int \mathbf{G}(v,u)\mathbf{K}(v,u)\mathrm{d}v\mathrm{d}u\Big\} = \mathrm{tr}\Big\{\int \mathbf{G}\mathbf{K}(v,v)\mathrm{d}v\Big\}.\end{aligned}$$

(ii) Suppose that $\mathbf{K}(u,v) = \sum_{i=1}^{\infty}\lambda_i\boldsymbol{\phi}_i(u)\boldsymbol{\phi}_i(v)^{\mathrm{T}}$ and $\mathbf{G}(u,v) = \sum_{j=1}^{\infty}\omega_j\boldsymbol{\psi}_j(u)\boldsymbol{\psi}_j(v)^{\mathrm{T}}$ where $\{\boldsymbol{\phi}_i(\cdot)\}$ and $\{\boldsymbol{\psi}_j(\cdot)\}$ are both orthonormal basis functions. Then, we have

$$\begin{aligned}\mathrm{tr}\Big\{\int \mathbf{K}\mathbf{G}(u,u)\mathrm{d}u\Big\} &= \mathrm{tr}\Big\{\int\int \mathbf{K}(u,v)\mathbf{G}(v,u)\mathrm{d}u\mathrm{d}v\Big\} \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\lambda_i\omega_j\int \boldsymbol{\phi}_i(v)^{\mathrm{T}}\boldsymbol{\psi}_j(v)\mathrm{d}v\int \boldsymbol{\psi}_j(u)^{\mathrm{T}}\boldsymbol{\phi}_i(u)\mathrm{d}u \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\lambda_i\omega_j\big|\langle\boldsymbol{\phi}_i,\boldsymbol{\psi}_j\rangle\big|^2 \leqslant \sum_{i=1}^{\infty}\lambda_i\omega_i \\ &\leqslant \Big(\max_i\lambda_i\Big)\sum_{j=1}^{\infty}\omega_j = \|\mathbf{K}\|_{\mathcal{L}}\|\mathbf{G}\|_{\mathcal{N}} = \|\mathbf{K}\|_{\mathcal{L}}\mathrm{tr}\Big\{\int \mathbf{G}(u,u)\mathrm{d}u\Big\},\end{aligned}$$

where the first inequality follows by using similar arguments to prove von Neumann's trace inequality (see Carlsson, 2021).

(iii) From Lemma A5(i) and the part (ii) above, we have

$$\begin{aligned}\|\mathbf{K}\mathbf{G}\|_{\mathcal{S},\mathrm{F}}^2 =&\mathrm{tr}\Big\{\int \mathbf{K}\mathbf{G}\mathbf{G}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}(u,u)\mathrm{d}u\Big\} = \mathrm{tr}\Big\{\int \mathbf{K}^{\mathrm{T}}\mathbf{K}\mathbf{G}\mathbf{G}^{\mathrm{T}}(u,u)\mathrm{d}u\Big\}\\ \leqslant&\|\mathbf{K}\mathbf{K}^{\mathrm{T}}\|_{\mathcal{L}}\mathrm{tr}\Big\{\int \mathbf{G}\mathbf{G}^{\mathrm{T}}(u,u)\mathrm{d}u\Big\} = \|\mathbf{K}\|_{\mathcal{L}}^2\|\mathbf{G}\|_{\mathcal{S},\mathrm{F}}^2,\end{aligned}$$

which implies the desired result. $\square$

A.2 Proof of Proposition 1

(i) Note that $\{\lambda_j\}_{j=1}^p$ are the non-vanishing eigenvalues of $\mathbf{\Omega} = \iint \mathbf{\Sigma}_y(u,v)\mathbf{\Sigma}_y(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v$, and $\{p^2\theta_j\}_{j=1}^r$ are nonzero eigenvalues of $\mathbf{\Omega}_{\mathcal{L}}$, while the other $p-r$ eigenvalues are zero. Then applying Lemma A1 yields that, for each $j \in [r]$,

$$|\lambda_j - p^2\theta_j| \leqslant \|\mathbf{\Omega} - \mathbf{\Omega}_{\mathcal{L}}\| = \|\mathbf{\Omega}_{\mathcal{R}}\|,$$

and for $r+1 \leqslant j \leqslant p$, $|\lambda_j| = |\lambda_j - 0| \leqslant \|\mathbf{\Omega}_{\mathcal{R}}\|$.

(ii) By Lemma A2, for $j \in [r]$ and $\boldsymbol{\xi}_j^{\mathrm{T}}\widetilde{\mathbf{b}}_j \geqslant 0$,

$$\|\boldsymbol{\xi}_j - \widetilde{\mathbf{b}}_j\| \leqslant \frac{\|\mathbf{\Omega}_{\mathcal{R}}\|/\sqrt{2}}{\min(|\lambda_{j-1} - p^2\theta_j|, |p^2\theta_j - \lambda_{j+1}|)}.$$

Note that there exists a generic constant $c > 0$ such that $|\lambda_{j-1} - p^2\theta_j| > p^2|\theta_{j-1} - \theta_j| - |\lambda_{j-1} - p^2\theta_{j-1}| > cp^2$ since $|\lambda_{j-1} - p^2\theta_{j-1}| \leqslant \|\mathbf{\Omega}_{\mathcal{R}}\| = o(p^2)$ from part (i). If $j < r$, a similar argument implies that $|p^2\theta_j - \lambda_{j+1}| > cp^2$. If $j = r$, $|p^2\theta_r - \lambda_{r+1}| > p^2\theta_r - |\lambda_{r+1}| > cp^2$ since $|\lambda_{r+1}| \leqslant \|\mathbf{\Omega}_{\mathcal{R}}\| = o(p^2)$ by using part (i) again. Hence, $\min(|\lambda_{j-1} - p^2\theta_j|, |p^2\theta_j - \lambda_{j+1}|) \gtrsim p^2$, and if $\boldsymbol{\xi}_j^{\mathrm{T}}\widetilde{\mathbf{b}}_j \geqslant 0$, we have

$$\|\boldsymbol{\xi}_j - \widetilde{\mathbf{b}}_j\| = O(p^{-2}\|\mathbf{\Omega}_{\mathcal{R}}\|), \text{ for } j \in [r].$$

A.3 Proof of Proposition 2

To prove Proposition 2, we first present a technical lemma with its proof.

Lemma A9. *Suppose that Assumption 1′ holds. Then, $\{p\vartheta_j\}_{j\in[r]}$ are the non-vanishing eigenvalues of $\mathbf{Q}(\cdot)\mathbf{Q}(\cdot)^{\mathrm{T}} \in \mathbb{H}^p \otimes \mathbb{H}^p$ with the corresponding eigenfunctions $\{\tilde{\mathbf{q}}_j(\cdot)\}_{j\in[r]}$.*

Proof. Let $\breve{\mathbf{q}}_j(\cdot)$ for $j \in [r]$ be the columns of $\mathbf{Q}(\cdot)$. By Assumption 1′, we know that $\langle \breve{\mathbf{q}}_i, \breve{\mathbf{q}}_j \rangle = p\vartheta_i I(i = j)$, which implies that $\tilde{\mathbf{q}}_i(\cdot) = (p\vartheta_j)^{-1/2}\breve{\mathbf{q}}_j(\cdot)$. Then, for $j \in [r]$,

$$\int \{\mathbf{Q}(u)\mathbf{Q}(v)^{\mathrm{T}}\}\tilde{\mathbf{q}}_j(v)\mathrm{d}v = \sum_{i=1}^{r} \breve{\mathbf{q}}_i(u)\langle \breve{\mathbf{q}}_i, \tilde{\mathbf{q}}_j\rangle = p\vartheta_j\tilde{\mathbf{q}}_j(u),$$

which indicates that $p\vartheta_j$ is the eigenvalue of $\mathbf{Q}(\cdot)\mathbf{Q}(\cdot)^{\mathrm{T}}$ with the corresponding eigenfunction $\tilde{\mathbf{q}}_j(\cdot)$. Since $\langle \tilde{\mathbf{q}}_i, \tilde{\mathbf{q}}_j\rangle = I(i = j)$ for $i, j \in [r]$, we can expand $\{\tilde{\mathbf{q}}_j(\cdot)\}_{j\in[r]}$ into a set of orthonormal basis functions in $\mathbb{H}^p$, denoted as $\{\tilde{\mathbf{q}}_j(\cdot)\}_{j=1}^{\infty}$. Considering that $\langle \tilde{\mathbf{q}}_j, \tilde{\mathbf{q}}_k\rangle = 0$ for any $j \leqslant r$ and $k \geqslant r + 1$, we obtain that $\int\{\mathbf{Q}(u)\mathbf{Q}(v)^{\mathrm{T}}\}\tilde{\mathbf{q}}_k(v)\mathrm{d}v = 0$ for $k \geqslant r + 1$. Thus, the rest eigenvalues of $\mathbf{Q}(\cdot)\mathbf{Q}(\cdot)^{\mathrm{T}}$ are zero. □

We are now ready to prove Proposition 2.

(i) Note that $\{\tau_i\}_{i=1}^{\infty}$ are the eigenvalues of $\boldsymbol{\Sigma}_y(\cdot, \cdot)$, and $\{p\vartheta_j\}_{j=1}^{r}$ are the r non-vanishing eigenvalues of $\mathbf{Q}\mathbf{Q}^{\mathrm{T}}(\cdot, \cdot)$ by Lemma A9. Applying Lemma A3, we have, for $j \in [r]$,

$$|\tau_j - p\vartheta_j| \leqslant \|\boldsymbol{\Sigma}_y - \mathbf{Q}\mathbf{Q}^{\mathrm{T}}\|_{\mathcal{L}} = \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}},$$

and, for $j > r + 1$, $|\tau_i| = |\tau_i - 0| \leqslant \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}}$.

(ii) By Lemma A7(iii), we have $\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}} \leqslant \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},1}^{1/2}\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\infty}^{1/2} = O(s_p) = o(p)$, which yields that $\tau_j \asymp p\vartheta_j \asymp p$ for $j \in [r]$. Under Assumption 1′, ϑ_j are distinguishable and bounded away from both zero and infinity, then $\min(|p\vartheta_{j-1} - \tau_j|, |\tau_j - p\vartheta_{j+1}|) \asymp p$ for $j \in [r]$. It follows from Lemma A4 that $\|\boldsymbol{\varphi}_j - \tilde{\mathbf{q}}_j\| = O(p^{-1}\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}})$ for $j \in [r]$.

A.4 Proof of Proposition 3

The sample covariance matrix of estimated idiosyncratic components by using the constrained least squares follows that

$$\tilde{\boldsymbol{\Sigma}}_\varepsilon(u, v) = \frac{1}{n}\{\mathbf{Y}(u) - \hat{\mathbf{Q}}(u)\hat{\boldsymbol{\Gamma}}^{\mathrm{T}}\}\{\mathbf{Y}(v)^{\mathrm{T}} - \hat{\boldsymbol{\Gamma}}\hat{\mathbf{Q}}(v)^{\mathrm{T}}\} = \frac{1}{n}\mathbf{Y}(u)\mathbf{Y}(v)^{\mathrm{T}} - \hat{\mathbf{Q}}(u)\hat{\mathbf{Q}}(v)^{\mathrm{T}},$$

where we use the normalization condition $n^{-1}\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\widehat{\boldsymbol{\Gamma}} = \mathbf{I}_r$ and $\widehat{\mathbf{Q}}(\cdot) = n^{-1}\mathbf{Y}(\cdot)\widehat{\boldsymbol{\Gamma}}$. If we can show that $\widehat{\mathbf{Q}}(u)\widehat{\mathbf{Q}}(v)^{\mathrm{T}} = \sum_{j=1}^{r}\hat{\tau}_j\widehat{\boldsymbol{\varphi}}_j(u)\widehat{\boldsymbol{\varphi}}_j(v)^{\mathrm{T}}$, then by the spectral decompositions of the sample covariance estimator

$$\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v) = \frac{1}{n}\mathbf{Y}(u)\mathbf{Y}(v)^{\mathrm{T}} = \sum_{j=1}^{r}\hat{\tau}_j\widehat{\boldsymbol{\varphi}}_j(u)\widehat{\boldsymbol{\varphi}}_j(v)^{\mathrm{T}} + \widehat{\mathbf{R}}(u,v) = \widehat{\mathbf{Q}}(u)\widehat{\mathbf{Q}}(v)^{\mathrm{T}} + \widetilde{\boldsymbol{\Sigma}}_{\varepsilon}(u,v),$$

we have $\widehat{\mathbf{R}}(\cdot,\cdot) = \widetilde{\boldsymbol{\Sigma}}_{\varepsilon}(\cdot,\cdot)$. Thus, by applying the adaptive functional thresholding with the same regularization parameters to the same residual covariance matrix functions, we have $\widehat{\mathbf{R}}^{\mathcal{A}}(\cdot,\cdot) = \widetilde{\boldsymbol{\Sigma}}_{\varepsilon}^{\mathcal{A}}(\cdot,\cdot)$, and then $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{F}}(\cdot,\cdot) = \widehat{\boldsymbol{\Sigma}}_y^{\mathcal{L}}(\cdot,\cdot)$, which gives the desired result.

We next show that $\widehat{\mathbf{Q}}(u)\widehat{\mathbf{Q}}(v)^{\mathrm{T}} = \sum_{j=1}^{r}\hat{\tau}_j\widehat{\boldsymbol{\varphi}}_j(u)\widehat{\boldsymbol{\varphi}}_j(v)^{\mathrm{T}}$ holds. To do this, we impose another identifiability condition that can serve as an alternative (see also Remark 1) to Assumption 1'.

Assumption S.1. *$p^{-1}\int\mathbf{Q}(u)^{\mathrm{T}}\mathbf{Q}(u)\mathrm{d}u = \mathbf{I}_r$ and $\boldsymbol{\Sigma}_{\gamma}$ is diagonal with distinct diagonal elements being bounded away from both 0 and ∞ as $p \to \infty$.*

Note that Assumptions 1' and S.1 can be converted to each other by orthogonal transformation. Thus, for the minimization problem (16), we can use the following two equivalent normalization constraints:

$$\begin{aligned}&(i)\ n^{-1}\sum_{t=1}^{n}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}} = \mathbf{I}_r,\ \text{and}\ p^{-1}\int\mathbf{Q}(u)^{\mathrm{T}}\mathbf{Q}(u)\mathrm{d}u \text{ is diagonal},\\ &(ii)\ n^{-1}\sum_{t=1}^{n}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}} \text{ is diagonal},\ \text{and}\ p^{-1}\int\mathbf{Q}(u)^{\mathrm{T}}\mathbf{Q}(u)\mathrm{d}u = \mathbf{I}_r.\end{aligned} \tag{S.4}$$

Note that (S.4)(i) is used in Section 2.2 to obtain FPOET estimator. Following the similar procedure, we obtain that $\check{\mathbf{Q}}(\cdot) = \sqrt{p}(\widehat{\boldsymbol{\varphi}}_1(\cdot),\ldots,\widehat{\boldsymbol{\varphi}}_r(\cdot))$ and $\check{\boldsymbol{\Gamma}} = p^{-1}\int\mathbf{Y}(u)^{\mathrm{T}}\check{\mathbf{Q}}(u)\mathrm{d}u$ is the solution to (16) under (S.4)(ii). One can show that the two solutions under normalization constraints (S.4)(i) and (ii) are equivalent and can be converted to each other through an orthogonal matrix, i.e., there exists an $r \times r$ orthogonal matrix $\mathbf{H}$ such that $\check{\mathbf{Q}}(\cdot) = \widehat{\mathbf{Q}}(\cdot)\mathbf{H}$ and $\check{\boldsymbol{\Gamma}} = \widehat{\boldsymbol{\Gamma}}\mathbf{H}$. Notice that $\check{\mathbf{Q}}(\cdot) = \sqrt{p}\{\widehat{\boldsymbol{\varphi}}_1(\cdot),\ldots,\widehat{\boldsymbol{\varphi}}_r(\cdot)\}$ and $\check{\boldsymbol{\Gamma}} = p^{-1}\int\mathbf{Y}(u)^{\mathrm{T}}\check{\mathbf{Q}}(u)\mathrm{d}u$, then we have

$$
\begin{aligned}
n^{-1}\check{\boldsymbol{\Gamma}}^{\mathrm{T}}\check{\boldsymbol{\Gamma}} =& p^{-2}n^{-1}\int \check{\mathbf{Q}}(u)^{\mathrm{T}}\mathbf{Y}(u)\mathrm{d}u\int \mathbf{Y}(v)^{\mathrm{T}}\check{\mathbf{Q}}(v)\mathrm{d}v\\
=& p^{-2}\int\int \check{\mathbf{Q}}(u)^{\mathrm{T}}\{n^{-1}\mathbf{Y}(u)\mathbf{Y}(v)^{\mathrm{T}}\}\check{\mathbf{Q}}(v)\mathrm{d}u\mathrm{d}v\\
=& p^{-1}\int \{\widehat{\boldsymbol{\varphi}}_1(u)^{\mathrm{T}},\ldots,\widehat{\boldsymbol{\varphi}}_r(u)^{\mathrm{T}}\}^{\mathrm{T}}\Big[\int \widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)\{\widehat{\boldsymbol{\varphi}}_1(v),\ldots,\widehat{\boldsymbol{\varphi}}_r(v)\}\mathrm{d}v\Big]\mathrm{d}u\\
=& p^{-1}\int \{\widehat{\boldsymbol{\varphi}}_1(u)^{\mathrm{T}},\ldots,\widehat{\boldsymbol{\varphi}}_r(u)^{\mathrm{T}}\}^{\mathrm{T}}\{\hat{\tau}_1\widehat{\boldsymbol{\varphi}}_1(u),\ldots,\hat{\tau}_r\widehat{\boldsymbol{\varphi}}_r(u)\}\mathrm{d}u\\
=& p^{-1}\mathrm{diag}(\hat{\tau}_1,\ldots,\hat{\tau}_r).
\end{aligned}
$$

Since $\check{\mathbf{Q}}(\cdot)\check{\boldsymbol{\Gamma}}^{\mathrm{T}} = \widehat{\mathbf{Q}}(\cdot)\mathbf{H}\mathbf{H}^{\mathrm{T}}\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}} = \widehat{\mathbf{Q}}(\cdot)\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}$, it follows that

$$
\widehat{\mathbf{Q}}(u)\widehat{\mathbf{Q}}(v)^{\mathrm{T}} = n^{-1}\widehat{\mathbf{Q}}(u)\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\widehat{\boldsymbol{\Gamma}}\widehat{\mathbf{Q}}(v)^{\mathrm{T}} = n^{-1}\check{\mathbf{Q}}(u)\check{\boldsymbol{\Gamma}}^{\mathrm{T}}\check{\boldsymbol{\Gamma}}\check{\mathbf{Q}}(v)^{\mathrm{T}} = \sum_{j=1}^{r}\hat{\tau}_j\widehat{\boldsymbol{\varphi}}_j(u)\widehat{\boldsymbol{\varphi}}_j(u)^{\mathrm{T}}.
$$

B Proofs of theoretical results in Section 3

B.1 Technical lemmas and their proofs

Lemma B10. *For $\mathbf{A}\in\mathbb{R}^{p\times q}$ and $\mathbf{K}=\{K_{ij}(\cdot,\cdot)\}_{q\times q}\in\mathbb{H}^q\otimes\mathbb{H}^q$, we have*

(i) $\|\mathbf{A}\mathbf{K}\|_{\mathcal{S},\max}\leqslant\|\mathbf{A}\|_\infty\|\mathbf{K}\|_{\mathcal{S},\max}$, *and* $\|\mathbf{K}\mathbf{A}^{\mathrm{T}}\|_{\mathcal{S},\max}\leqslant\|\mathbf{K}\|_{\mathcal{S},\max}\|\mathbf{A}^{\mathrm{T}}\|_1=\|\mathbf{A}\|_\infty\|\mathbf{K}\|_{\mathcal{S},\max}$*;*

(ii) $\|\mathbf{A}\mathbf{K}\|_{\mathcal{S},\mathrm{F}}\leqslant\|\mathbf{A}\|_{\mathrm{F}}\|\mathbf{K}\|_{\mathcal{S},\mathrm{F}}$, *and* $\|\mathbf{K}\mathbf{A}^{\mathrm{T}}\|_{\mathcal{S},\mathrm{F}}\leqslant\|\mathbf{K}\|_{\mathcal{S},\mathrm{F}}\|\mathbf{A}^{\mathrm{T}}\|_{\mathrm{F}}=\|\mathbf{A}\|_{\mathrm{F}}\|\mathbf{K}\|_{\mathcal{S},\mathrm{F}}$*;*

(iii) $\|\mathbf{A}\mathbf{K}\|_{\mathcal{S},\infty}\leqslant\|\mathbf{A}\|_\infty\|\mathbf{K}\|_{\mathcal{S},\infty}$, *and* $\|\mathbf{K}\mathbf{A}^{\mathrm{T}}\|_{\mathcal{S},\infty}\leqslant\|\mathbf{K}\|_{\mathcal{S},\infty}\|\mathbf{A}^{\mathrm{T}}\|_\infty=\|\mathbf{A}\|_1\|\mathbf{K}\|_{\mathcal{S},\infty}$*;*

(iv) $\|\mathbf{A}\mathbf{K}\|_{\mathcal{S},1}\leqslant\|\mathbf{A}\|_1\|\mathbf{K}\|_{\mathcal{S},1}$, *and* $\|\mathbf{K}\mathbf{A}^{\mathrm{T}}\|_{\mathcal{S},1}\leqslant\|\mathbf{K}\|_{\mathcal{S},1}\|\mathbf{A}^{\mathrm{T}}\|_1=\|\mathbf{A}\|_\infty\|\mathbf{K}\|_{\mathcal{S},1}$.

Proof. (i) It follows that

$$
\begin{aligned}
\|\mathbf{A}\mathbf{K}\|_{\mathcal{S},\max} = \max_{i\in[p],j\in[q]}\Big\|\sum_{k=1}^{q}A_{ik}K_{kj}\Big\|_{\mathcal{S}} \leqslant& \max_{i\in[p],j\in[q]}\sum_{k=1}^{q}|A_{ik}|\|K_{kj}\|_{\mathcal{S}}\\
\leqslant&\Big(\max_{i\in[p]}\sum_{k=1}^{q}|A_{ik}|\Big)\cdot\|\mathbf{K}\|_{\mathcal{S},\max}=\|\mathbf{A}\|_\infty\|\mathbf{K}\|_{\mathcal{S},\max}.
\end{aligned}
$$

Further,

$$
\begin{aligned}
\|\mathbf{K}\mathbf{A}^{\mathrm{T}}\|_{\mathcal{S},\max} =& \|(\mathbf{K}\mathbf{A}^{\mathrm{T}})^{\mathrm{T}}\|_{\mathcal{S},\max} = \|\mathbf{A}\mathbf{K}^{\mathrm{T}}\|_{\mathcal{S},\max}\leqslant\|\mathbf{A}\|_\infty\|\mathbf{K}^{\mathrm{T}}\|_{\mathcal{S},\max}\\
=&\|\mathbf{A}\|_\infty\|\mathbf{K}\|_{\mathcal{S},\max}=\|\mathbf{A}^{\mathrm{T}}\|_1\|\mathbf{K}\|_{\mathcal{S},\max}.
\end{aligned}
$$

(ii) It follows that

$$
\begin{aligned}
\|\mathbf{AK}\|_{\mathcal{S},\mathrm{F}}^2 =& \sum_{i=1}^p \sum_{j=1}^q \Big\| \sum_{k=1}^q A_{ik} K_{kj} \Big\|_{\mathcal{S}}^2 \leqslant \sum_{i=1}^p \sum_{j=1}^q \Big(\sum_{k=1}^q A_{ik}^2 \sum_{k=1}^q \|K_{kj}\|_{\mathcal{S}}^2 \Big) \\
=& \sum_{i=1}^p \sum_{j=1}^q \Big(\sum_{k=1}^q \sum_{l=1}^q A_{ik}^2 \|K_{lj}\|_{\mathcal{S}}^2 \Big) = \left(\sum_{i=1}^p \sum_{k=1}^q A_{ik}^2 \right) \left(\sum_{l=1}^q \sum_{j=1}^q \|K_{lj}\|_{\mathcal{S}}^2 \right) \\
=& \|\mathbf{A}\|_{\mathrm{F}} \|\mathbf{K}\|_{\mathcal{S},\mathrm{F}},
\end{aligned}
$$

where the inequality follows from the Cauchy–Schwartz inequality. Furthermore,

$$
\|\mathbf{K}\mathbf{A}^{\mathrm{T}}\|_{\mathcal{S},\mathrm{F}} = \|(\mathbf{K}\mathbf{A}^{\mathrm{T}})^{\mathrm{T}}\|_{\mathcal{S},\mathrm{F}} = \|\mathbf{A}\mathbf{K}^{\mathrm{T}}\|_{\mathcal{S},\mathrm{F}} \leqslant \|\mathbf{A}\|_{\mathrm{F}} \|\mathbf{K}\|_{\mathcal{S},\mathrm{F}} = \|\mathbf{K}\|_{\mathcal{S},\mathrm{F}} \|\mathbf{A}^{\mathrm{T}}\|_{\mathrm{F}}.
$$

(iii) and (iv) It follows that

$$
\begin{aligned}
\|\mathbf{AK}\|_{\mathcal{S},\infty} =& \max_{i \in [p]} \sum_{k=1}^q \sum_{j=1}^q \|A_{ik} K_{kj}\|_{\mathcal{S}} = \max_{i \in [p]} \sum_{k=1}^q \sum_{j=1}^q |A_{ik}| \|K_{kj}\|_{\mathcal{S}} \\
\leqslant& \max_{i \in [p]} \sum_{k=1}^q \sum_{j=1}^q |A_{ik}| \max_{k' \in [q]} \|K_{k'j}\|_{\mathcal{S}} \\
=& \Big(\max_{i \in [p]} \sum_{k=1}^q |A_{ik}| \Big) \Big(\max_{k' \in [q]} \sum_{j=1}^q \|K_{k'j}\|_{\mathcal{S}} \Big) = \|\mathbf{A}\|_{\infty} \|\mathbf{K}\|_{\mathcal{S},\infty}.
\end{aligned}
$$

Furthermore,

$$
\|\mathbf{K}\mathbf{A}^{\mathrm{T}}\|_{\mathcal{S},1} = \|\mathbf{A}\mathbf{K}^{\mathrm{T}}\|_{\mathcal{S},\infty} \leqslant \|\mathbf{A}\|_{\infty} \|\mathbf{K}^{\mathrm{T}}\|_{\mathcal{S},\infty} = \|\mathbf{A}\|_{\infty} \|\mathbf{K}\|_{\mathcal{S},1}.
$$

The other two arguments can be proved similarly. $\square$

Lemma B11. *For* $\mathbf{f}, \mathbf{g} \in \mathbb{H}^r$, *and* $\mathbf{A} \in \mathbb{R}^{p \times r}$, *we have*

(i) $\|\mathbf{A}\mathbf{f}\| \leqslant \|\mathbf{A}\| \cdot \|\mathbf{f}\|$;

(ii) $\|K\|_{\mathcal{S}} \leqslant \|\mathbf{f}\| \cdot \|\mathbf{g}\|$ *where* $K(\cdot, \cdot) \in \mathbb{S}$ *is defined as* $K(u, v) = \mathbf{f}(u)^{\mathrm{T}} \mathbf{g}(v)$.

Proof. (i) By the definition, it follows that

$$
\|\mathbf{A}\mathbf{f}\| = \left\{ \int \mathbf{f}_t(u)^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \mathbf{A} \mathbf{f}_t(u) \mathrm{d}u \right\}^{1/2} \leqslant \left\{ \int \lambda_{\max}(\mathbf{A}^{\mathrm{T}} \mathbf{A}) \mathbf{f}_t(u)^{\mathrm{T}} \mathbf{f}_t(u) \mathrm{d}u \right\}^{1/2} = \|\mathbf{A}\| \cdot \|\mathbf{f}\|.
$$

(ii) By the Cauchy–Schwartz inequality,

$$
\begin{aligned}
\|K\|_{\mathcal{S}} =&\Big[\int\int\{\mathbf{f}(u)^{\mathrm{T}}\mathbf{g}(v)\}^2\mathrm{d}u\mathrm{d}v\Big]^{1/2} = \Big[\int\int\Big\{\sum_{j=1}^{r} f_j(u)g_j(v)\Big\}^2\mathrm{d}u\mathrm{d}v\Big]^{1/2}\\
\leqslant&\Big\{\int\int\sum_{j=1}^{r} f_j(u)^2\sum_{j=1}^{r} g_j(v)^2\mathrm{d}u\mathrm{d}v\Big\}^{1/2} = \Big\{\int\mathbf{f}(u)^{\mathrm{T}}\mathbf{f}(u)\mathrm{d}u\int\mathbf{g}(v)^{\mathrm{T}}\mathbf{g}(v)\mathrm{d}v\Big\}^{1/2}\\
=&\|\mathbf{f}\|\cdot\|\mathbf{g}\|.
\end{aligned}
$$

□

Lemma B12. *Under Assumptions 3(iv) and 4, we have that,*

(i) for any $i,j\in[r]$, $\|n^{-1}\sum_{t=1}^{n} f_{ti}f_{tj}-\Sigma_{f,ij}\|_{\mathcal{S}} = O_p(1/\sqrt{n})$, *and* $\|n^{-1}\sum_{t=1}^{n}\mathbf{f}_t\mathbf{f}_t^{\mathrm{T}}-\mathbf{\Sigma}_f\|_{\mathcal{S},\max} = O_p(1/\sqrt{n})$*;*

(ii) for any $i,j\in[p]$, $\|n^{-1}\sum_{t=1}^{n}\varepsilon_{ti}\varepsilon_{tj}-\Sigma_{\varepsilon,ij}\|_{\mathcal{S}} = O_p(\mathcal{M}_\varepsilon/\sqrt{n})$, *and* $\|n^{-1}\sum_{t=1}^{n}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\mathrm{T}}-\mathbf{\Sigma}_\varepsilon\|_{\mathcal{S},\max} = O_p(\mathcal{M}_\varepsilon\sqrt{\log p/n})$*;*

(iii) for any $i,j\in[p]$, $\|n^{-1}\sum_{t=1}^{n} y_{ti}y_{tj}-\Sigma_{y,ij}\|_{\mathcal{S}} = O_p(\mathcal{M}_\varepsilon/\sqrt{n})$, *and* $\|n^{-1}\sum_{t=1}^{n}\mathbf{y}_t\mathbf{y}_t^{\mathrm{T}}-\mathbf{\Sigma}_y\|_{\mathcal{S},\max} = O_p(\mathcal{M}_\varepsilon\sqrt{\log p/n})$.

Proof. For parts (i) and (ii), see Theorem 2, equations (12) and (14) in Guo and Qiao (2023) for the corresponding proofs.

(iii) The autocovariance matrix functions of $\{\mathbf{y}_t(\cdot)\}_{t\in\mathbb{Z}}$ at lag h satisfy $\mathbf{\Sigma}_y^{(h)}(\cdot,\cdot) = \mathbf{B}\mathbf{\Sigma}_f^{(h)}(\cdot,\cdot)\mathbf{B}^{\mathrm{T}}+\mathbf{\Sigma}_\varepsilon^{(h)}(\cdot,\cdot)$, and the corresponding spectral density matrix function at frequency $\theta\in[-\pi,\pi]$ is given by

$$
\begin{aligned}
\mathbf{f}_{y,\theta} =&\frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\mathbf{\Sigma}_y^{(h)}\exp(-ih\theta) = \frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\mathbf{B}\mathbf{\Sigma}_f^{(h)}\mathbf{B}^{\mathrm{T}}\exp(-ih\theta) + \frac{1}{2\pi}\sum_{h\in\mathbb{Z}}\mathbf{\Sigma}_\varepsilon^{(h)}\exp(-ih\theta)\\
=&\mathbf{B}\mathbf{f}_{f,\theta}\mathbf{B}^{\mathrm{T}}+\mathbf{f}_{\varepsilon,\theta}.
\end{aligned}
$$

By definition, the functional stability measure of $\{\mathbf{y}_t(\cdot)\}_{t\in\mathbb{Z}}$ is

$$
\begin{aligned}
\mathcal{M}_y =& 2\pi \cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\phi}\in\mathbb{H}_0^p} \frac{\langle \boldsymbol{\phi}, \mathbf{f}_{y,\theta}(\boldsymbol{\phi})\rangle}{\langle \boldsymbol{\phi}, \boldsymbol{\Sigma}_y(\boldsymbol{\phi})\rangle} \\
=& 2\pi \cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\phi}\in\mathbb{H}_0^p} \frac{\int\int \boldsymbol{\phi}(u)^{\mathrm{T}}\mathbf{f}_{y,\theta}(u,v)\boldsymbol{\phi}(v)\mathrm{d}u\mathrm{d}v}{\int\int \boldsymbol{\phi}(u)^{\mathrm{T}}\boldsymbol{\Sigma}_y(u,v)\boldsymbol{\phi}(v)\mathrm{d}u\mathrm{d}v} \\
\leqslant& 2\pi \cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\phi}\in\mathbb{H}_0^p} \frac{\int\int \boldsymbol{\phi}(u)^{\mathrm{T}}\mathbf{B}\mathbf{f}_{f,\theta}(u,v)\mathbf{B}^{\mathrm{T}}\boldsymbol{\phi}(v)\mathrm{d}u\mathrm{d}v}{\int\int \boldsymbol{\phi}(u)^{\mathrm{T}}\mathbf{B}\boldsymbol{\Sigma}_f(u,v)\mathbf{B}^{\mathrm{T}}\boldsymbol{\phi}(v)\mathrm{d}u\mathrm{d}v} \\
&+ 2\pi \cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\phi}\in\mathbb{H}_0^p} \frac{\int\int \boldsymbol{\phi}(u)^{\mathrm{T}}\mathbf{f}_{\varepsilon,\theta}(u,v)\boldsymbol{\phi}(v)\mathrm{d}u\mathrm{d}v}{\int\int \boldsymbol{\phi}(u)^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon(u,v)\boldsymbol{\phi}(v)\mathrm{d}u\mathrm{d}v} \\
\leqslant& 2\pi \cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\xi}\in\mathbb{H}_0^r} \frac{\langle \boldsymbol{\xi}, \mathbf{f}_{f,\theta}(\boldsymbol{\xi})\rangle}{\langle \boldsymbol{\xi}, \boldsymbol{\Sigma}_f(\boldsymbol{\xi})\rangle} + 2\pi \cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\phi}\in\mathbb{H}_0^p} \frac{\langle \boldsymbol{\phi}, \mathbf{f}_{\varepsilon,\theta}(\boldsymbol{\phi})\rangle}{\langle \boldsymbol{\phi}, \boldsymbol{\Sigma}_\varepsilon(\boldsymbol{\phi})\rangle} \\
=& \mathcal{M}_f + \mathcal{M}_\varepsilon \asymp \mathcal{M}_\varepsilon.
\end{aligned}
$$

The other conditions imposed by Guo and Qiao (2023) for $\{\mathbf{y}_t(\cdot)\}_{t\in\mathbb{Z}}$ can be easily verified. Then the desired results can be obtained immediately by combining the above results. □

We next introduce a lemma to give the perturbation rate in elementwise ℓ_∞ norm of the eigenvectors if a matrix is perturbed. Suppose that $\mathbf{A} \in \mathbb{R}^{p\times p}$ is a symmetric matrix. Let the perturbed matrix be $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}$, where $\mathbf{E} \in \mathbb{R}^{p\times p}$ is a symmetric perturbation matrix. Suppose the spectral decomposition of $\mathbf{A}$ is given by $\mathbf{A} = \sum_{i=1}^r \lambda_i \mathbf{v}_i\mathbf{v}_i^{\mathrm{T}} + \sum_{i>r} \lambda_i \mathbf{v}_i\mathbf{v}_i^{\mathrm{T}}$, where $|\lambda_1| > |\lambda_2| > \cdots > |\lambda_p|$. Clearly, $\mathbf{A}_r = \sum_{i=1}^r \lambda_i \mathbf{v}_i\mathbf{v}_i^{\mathrm{T}}$ is the best rank-r approximation of $\mathbf{A}$. Analogously, the spectral decomposition of $\widetilde{\mathbf{A}} = \sum_{i=1}^r \tilde{\lambda}_i \tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i^{\mathrm{T}} + \sum_{i>r} \tilde{\lambda}_i \tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i^{\mathrm{T}}$. Write $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_r) \in \mathbb{R}^{p\times r}$ and $\widetilde{\mathbf{V}} = (\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_r) \in \mathbb{R}^{p\times r}$.

Lemma B13. *Suppose ι satisfies $\iota > \|\mathbf{E}\|$ and for any $i \in [r]$, the interval $(\lambda_i - \iota, \lambda_i + \iota)$ does not contain any eigenvalues of $\mathbf{A}$ other than λ_i. Then, there exists an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{r\times r}$ such that*

$$
\|\widetilde{\mathbf{V}}\mathbf{U} - \mathbf{V}\|_{\max} = O\left(\frac{r^{5/2}\mu^2\|\mathbf{E}\|_\infty}{(|\lambda_r| - \|\mathbf{A} - \mathbf{A}_r\|_\infty)\sqrt{p}}\right),
$$

where $\mu = \mu(\mathbf{V})$ is the coherence of $\mathbf{V}$ defined as $\mu(\mathbf{V}) = (p/r)\max_i \sum_{j=1}^r V_{ij}^2$.

Proof. The proof can be found in Fan, Wang and Zhong (2018) and thus is omitted here. □

Lemma B14. *(Theorem 4.2.5 in Hsing and Eubank (2015)). If* $\mathbf{K}(\cdot,\cdot)$ *is a compact and nonnegative definite kernel matrix function with associated eigenvalue/eigenfunction pairs* $\{(\lambda_j, \mathbf{e}_j(\cdot)\}_{j=1}^{\infty}$*, then*

$$\lambda_k = \max_{\mathbf{e}\in \text{span}\{\mathbf{e}_1,\ldots,\mathbf{e}_{k-1}\}^{\perp}} \frac{\langle \mathbf{e}, \mathbf{K}(\mathbf{e})\rangle}{\|\mathbf{e}\|^2},$$

where $\mathbf{K}(\mathbf{e})(\cdot) = \int \mathbf{K}(\cdot, u)\mathbf{e}(u)\mathrm{d}u$.

Lemma B15. *Suppose that* $\mathbf{K}, \mathbf{G} \in \mathbb{H}^p \otimes \mathbb{H}^p$ *are Mercer's kernels, and* $\lambda_{\min}(\mathbf{G}) > c_n$ *for a sequence* $c_n > 0$*. If* $\|\mathbf{K} - \mathbf{G}\|_{\mathcal{L}} = o_p(c_n)$*, then* $\lambda_{\min}(\mathbf{K}) > c_n/2$*, and*

$$\|\mathbf{K}^{-1} - \mathbf{G}^{-1}\|_{\mathcal{L}} = O_p(c_n^{-2})\|\mathbf{K} - \mathbf{G}\|_{\mathcal{L}}.$$

Proof. Note that $\langle \mathbf{x}, \mathbf{K}(\mathbf{x})\rangle = \int\int \mathbf{x}(u)^{\mathrm{T}}\mathbf{K}(u,v)\mathbf{x}(v)\mathrm{d}u\mathrm{d}v$ $\mathbf{x} \in \mathbb{H}^p$. Then for any $\mathbf{x} \in \mathbb{H}^p$ such that $\|\mathbf{x}\| = 1$, we have

$$\left|\int\int \mathbf{x}(u)^{\mathrm{T}}[\mathbf{K}(u,v) - \mathbf{G}(u,v)]\mathbf{x}(v)\mathrm{d}u\mathrm{d}v\right| \leqslant \|\mathbf{x}\|^2 \cdot \|\mathbf{K} - \mathbf{G}\|_{\mathcal{L}} = o_p(c_n)$$

Thus, for n large enough, $\langle \mathbf{x}, \mathbf{K}(\mathbf{x})\rangle \geqslant \langle \mathbf{x}, \mathbf{G}(\mathbf{x})\rangle - c_n/2 > c_n/2$, since for any $\|\mathbf{x}\| = 1$, $\langle \mathbf{x}, \mathbf{G}(\mathbf{x})\rangle \geqslant \lambda_{\min}(\mathbf{G}) > c_n$ using Lemma B14. Hence, for all $k \geqslant 1$, the k-th eigenvalue of $\mathbf{K}$ is larger than $c_n/2$, which implies that $\lambda_{min}(\mathbf{K}) > c_n/2$. In addition,

$$\begin{aligned}\|\mathbf{K}^{-1} - \mathbf{G}^{-1}\|_{\mathcal{L}} =& \|\mathbf{K}^{-1}(\mathbf{G} - \mathbf{K})\mathbf{G}^{-1}\|_{\mathcal{L}} \leqslant \|\mathbf{K}^{-1}\|_{\mathcal{L}}\|\mathbf{K} - \mathbf{G}\|_{\mathcal{L}}\|\mathbf{G}^{-1}\|_{\mathcal{L}}\\ =& O_p(c_n^{-2})\|\mathbf{K} - \mathbf{G}\|_{\mathcal{L}},\end{aligned}$$

where the last line comes from $\|\mathbf{K}^{-1}\|_{\mathcal{L}} = \lambda_{\max}(\mathbf{K}^{-1}) = [\lambda_{\min}(\mathbf{K})]^{-1} = O_p(c_n^{-1})$. □

Lemma B16. *Under Assumptions 3′(iv) and 4′, we have that,*

(i) for any $i, j \in [r]$, $|n^{-1}\sum_{t=1}^n \gamma_{ti}\gamma_{tj} - \Sigma_{\gamma,ij}| = O_p(1/\sqrt{n})$*, and* $\|n^{-1}\sum_{t=1}^n \boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}} - \boldsymbol{\Sigma}_{\gamma}\|_{\max} = O_p(1/\sqrt{n})$*;*

(ii) for any $i, j \in [p]$, $\|n^{-1}\sum_{t=1}^n \varepsilon_{ti}\varepsilon_{tj} - \Sigma_{\varepsilon,ij}\|_{\mathcal{S}} = O_p(\mathcal{M}_{\varepsilon}/\sqrt{n})$*, and* $\|n^{-1}\sum_{t=1}^n \boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\mathrm{T}} - \boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{S},\max} = O_p(\mathcal{M}_{\varepsilon}\sqrt{\log p/n})$*;*

(iii) for any $i \in [p], j \in [r]$, $\|n^{-1}\sum_{t=1}^n \varepsilon_{ti}\gamma_{tj}\| = O_p(\mathcal{M}_{\varepsilon}/\sqrt{n})$*, and* $\max_{i\in[p],j\in[r]} \|n^{-1}\sum_{t=1}^n \varepsilon_{ti}\gamma_{tj}\| = O_p(\mathcal{M}_{\varepsilon}\sqrt{\log p/n})$.

Proof. For parts (i) and (ii), see Theorem 2, equations (12) and (14) in Guo and Qiao (2023) for the corresponding proofs.

(iii) See Remark 3 and equation (2.16) in Fang et al. (2022) for a proof. □

Lemma B17. *Under Assumptions 4 and 4′, we have (i)* $\max_{t\in[n]}\|\mathbf{f}_t\| = O_p(\sqrt{\log n})$*; (ii)* $\max_{t\in[n]}\|\boldsymbol{\gamma}_t\| = O_p(\sqrt{\log n})$.

Proof. Notice that $\mathbf{f}_t(\cdot)$ in model (1) and $\boldsymbol{\gamma}_t$ in model (2) follow the sub-Gaussian functional linear process and sub-Gaussian linear process (see Section E.3), respectively, with $\mathbb{E}\|\mathbf{f}_t\| = O(1)$ and $\mathbb{E}\|\boldsymbol{\gamma}_t\| = O(1)$. Applying Bonferroni's method yields that for $j \in [r]$ and any given $\eta > 0$,

$$P\left(\max_{t\in[n]}(\|f_{tj}\|^2 - \mathbb{E}\|f_{tj}\|^2) \geqslant \eta\right) \leqslant n\max_{t\in[n]} P\left(\|f_{tj}\|^2 - \mathbb{E}\|f_{tj}\|^2 \geqslant \eta\right)$$
$$\leqslant 2n\exp\{-c\min(\eta^2,\eta)\},$$

where $c > 0$ is some constant and the second inequality follows from Lemma 5 in Fang et al. (2022). Letting $\eta = \log n$ gives that $\max_{t\in[n]}\|\mathbf{f}_t\|^2 = O_p(\log n) + O_p(\mathbb{E}\|\mathbf{f}_t\|^2)$, which implies that $\max_{t\in[n]}\|\mathbf{f}_t\| = O_p(\sqrt{\log n})$. The second argument can be proved similarly. □

B.2 Proof of Theorem 1

The proof of part (i) of Theorem 1 mainly relies on Lemma B13. To prove Theorem 1, we first present some technical lemmas with their proofs.

Lemma B18. *Suppose that Assumption 1 holds. Then there exist some constants* $C_{\max}, C_\infty > 0$ *such that (i)* $\|\boldsymbol{\Sigma}_f\|_{\mathcal{S},\max} \leqslant C_{\max}$*, (ii)* $\max(\|\boldsymbol{\Sigma}_f\|_{\mathcal{S},\infty}, \|\boldsymbol{\Sigma}_f\|_{\mathcal{S},1}, \|\boldsymbol{\Sigma}_f\|_{\mathcal{S},\mathrm{F}}) \leqslant C_\infty$.

Proof. (i) In Assumption 1, we assume that

$$\int\int \boldsymbol{\Sigma}_f(u,v)\boldsymbol{\Sigma}_f(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v = \mathrm{diag}(\theta_1,\ldots,\theta_r),$$

i.e.,

$$\int\int\sum_{j=1}^r \Sigma_{f,ij}(u,v)^2\mathrm{d}u\mathrm{d}v = \theta_i,\ \text{for } i\in[r].$$

Then we have

$$
\begin{aligned}
\|\boldsymbol{\Sigma}_f\|_{\mathcal{S},\max}^2 = \max_{i\in[r],j\in[r]} \|\Sigma_{f,ij}\|_{\mathcal{S}}^2 &= \max_{i\in[r],j\in[r]} \int\int \Sigma_{f,ij}(u,v)^2 \mathrm{d}u\mathrm{d}v \\
&\leqslant \max_{i\in[r]} \int\int \sum_{j=1}^r \Sigma_{f,ij}(u,v)^2 \mathrm{d}u\mathrm{d}v = \max_{i\in[r]} \theta_i = \theta_1,
\end{aligned}
$$

which implies that $\|\boldsymbol{\Sigma}_f\|_{\mathcal{S},\max} \leqslant \theta_1^{1/2} \equiv C_{\max}$.

(ii) Note that $\boldsymbol{\Sigma}_f(u,v) \in \mathbb{R}^{r\times r}$, we have $\max(\|\boldsymbol{\Sigma}_f\|_{\mathcal{S},\infty}, \|\boldsymbol{\Sigma}_f\|_{\mathcal{S},1}, \|\boldsymbol{\Sigma}_f\|_{\mathcal{S},\mathrm{F}}) \leqslant r\|\boldsymbol{\Sigma}_f\|_{\mathcal{S},\max} \leqslant r\theta_1^{1/2} \equiv C_\infty$. □

Lemma B19. *Suppose that Assumptions 1–3 hold. Then we have (i) $\|\boldsymbol{\Sigma}_y\|_{\mathcal{S},\max} \lesssim 1$, (ii) $\|\boldsymbol{\Sigma}_y\|_{\mathcal{S},\infty} \lesssim p$, and (iii) $\|\boldsymbol{\Sigma}_y\|_{\mathcal{S},1} \lesssim p$.*

Proof. (i) By Lemma B10(i) and the fact $\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\max} \leqslant \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}}$, we have

$$
\begin{aligned}
\|\boldsymbol{\Sigma}_y\|_{\mathcal{S},\max} =& \|\mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}^{\mathrm{T}} + \boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\max} \leqslant \|\mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}^{\mathrm{T}}\|_{\mathcal{S},\max} + \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\max} \\
\leqslant& \|\mathbf{B}\|_\infty \|\boldsymbol{\Sigma}_f\|_{\mathcal{S},\max} \|\mathbf{B}^{\mathrm{T}}\|_1 + \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\max} \lesssim r^2 C^2 C_{\max} + O(1) \asymp 1.
\end{aligned}
$$

(ii) By Lemma B10(iii), we have

$$
\begin{aligned}
\|\boldsymbol{\Sigma}_y\|_{\mathcal{S},\infty} =& \|\mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}^{\mathrm{T}} + \boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\infty} \leqslant \|\mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}^{\mathrm{T}}\|_{\mathcal{S},\infty} + \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\infty} \\
\leqslant& \|\mathbf{B}\|_\infty \|\mathbf{B}^{\mathrm{T}}\|_\infty \|\boldsymbol{\Sigma}_f\|_{\mathcal{S},\infty} + \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\infty} \leqslant rpC^2 C_\infty + s_p \lesssim p.
\end{aligned}
$$

Part (iii) can be proved similarly. □

Lemma B20. *Supposing that Assumptions 1–3 hold, we have $\|\boldsymbol{\Omega}_{\mathcal{L}}\| \asymp p^2$ and $\|\boldsymbol{\Omega}_{\mathcal{R}}\| \lesssim p$.*

Proof. For the first part of the lemma, notice that $\|\boldsymbol{\Omega}_{\mathcal{L}}\| \leqslant \|\boldsymbol{\Omega}_{\mathcal{L}}\|_{\mathrm{F}} \leqslant \sqrt{r}\|\boldsymbol{\Omega}_{\mathcal{L}}\|$ where r is the rank of $\boldsymbol{\Omega}_{\mathcal{L}}$, so $\|\boldsymbol{\Omega}_{\mathcal{L}}\| \asymp \|\boldsymbol{\Omega}_{\mathcal{L}}\|_{\mathrm{F}}$, and

$$
\begin{aligned}
\|\boldsymbol{\Omega}_{\mathcal{L}}\|_{\mathrm{F}}^2 =& \left\| p\mathbf{B} \int\int \boldsymbol{\Sigma}_f(u,v)\boldsymbol{\Sigma}_f(u,v)^{\mathrm{T}} \mathrm{d}u\mathrm{d}v \mathbf{B}^{\mathrm{T}} \right\|_{\mathrm{F}}^2 \\
=& p^4 \mathrm{tr}\left(\mathrm{diag}\{\theta_1,\ldots,\theta_r\}\mathrm{diag}\{\theta_1,\ldots,\theta_r\}^{\mathrm{T}}\right) = p^4 \sum_{j=1}^r \theta_i^2 \asymp p^4,
\end{aligned}
$$

where the second equality follows from Assumption 1 that $\int\int \boldsymbol{\Sigma}_f(u,v)\boldsymbol{\Sigma}_f(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v = \mathrm{diag}\{\theta_1,\ldots,\theta_r\}$ and $\mathbf{B}^{\mathrm{T}}\mathbf{B} = p\mathbf{I}_r$. Thus we have $\|\boldsymbol{\Omega}_{\mathcal{L}}\| \asymp p^2$. For the second part, we have

$$
\begin{aligned}
\|\boldsymbol{\Omega}_{\mathcal{R}}\| \leqslant & \left\|\int\int \boldsymbol{\Sigma}_{\varepsilon}(u,v)\boldsymbol{\Sigma}_{\varepsilon}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\| \\
& + \left\|\int\int \mathbf{B}\boldsymbol{\Sigma}_f(u,v)\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\| + \left\|\int\int \boldsymbol{\Sigma}_{\varepsilon}(u,v)\mathbf{B}\boldsymbol{\Sigma}_f(u,v)^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\| \\
\leqslant & \|\boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{L}}^2 + 2\|\mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}^{\mathrm{T}}\|_{\mathcal{L}}\|\boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{L}} \\
= & O(p),
\end{aligned}
$$

where the second inequality follows from Lemmas A5(ii) and A6(iv), and the last line follows from Lemmas A7(i)(ii) and B18(ii). □

Lemma B21. *Under the assumptions of Theorem 1, we have* $\|\boldsymbol{\Omega}_{\mathcal{R}}\|_{\infty} \lesssim ps_p = o(p^2)$.

Proof. Notice that

$$
\begin{aligned}
\|\boldsymbol{\Omega}_{\mathcal{R}}\|_{\infty} \leqslant & \left\|\int\int \boldsymbol{\Sigma}_{\varepsilon}(u,v)\boldsymbol{\Sigma}_{\varepsilon}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\|_{\infty} \\
& + \left\|\int\int \mathbf{B}\boldsymbol{\Sigma}_f(u,v)\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\|_{\infty} + \left\|\int\int \boldsymbol{\Sigma}_{\varepsilon}(u,v)\mathbf{B}\boldsymbol{\Sigma}_f(u,v)^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\|_{\infty} \\
\leqslant & \|\boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{S},\infty}\|\boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{S},1} + 2\|\mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}^{\mathrm{T}}\|_{\mathcal{S},\infty}\|\boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{S},1} \\
\leqslant & \|\boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{S},\infty}\|\boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{S},1} + 2\|\boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{S},1}\|\mathbf{B}\|_{\infty}\|\mathbf{B}^{\mathrm{T}}\|_{\infty}\|\boldsymbol{\Sigma}_f\|_{\mathcal{S},\infty} \\
\leqslant & s_p^2 + 2rC^2C_{\infty}s_p p \lesssim s_p^2 + ps_p \lesssim ps_p = o(p^2),
\end{aligned}
$$

where the second inequality follows from Lemma A6(ii), the third inequality follows from Lemma B10(iii), and the fourth inequality follows from Lemma B18. □

Lemma B22. *Under the assumptions of Theorem 1, we have (i)* $\|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}\| = O_p(\mathcal{M}_{\varepsilon}p^2\sqrt{1/n}) = o_p(p^2)$, *and (ii)* $\|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}\|_{\infty} = O_p(\mathcal{M}_{\varepsilon}p^2\sqrt{\log p/n}) = o_p(p^2)$.

Proof. (i) Note that

$$
\begin{aligned}
\|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}\| =&\left\|\int\int \widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v-\int\int \boldsymbol{\Sigma}_y(u,v)\boldsymbol{\Sigma}_y(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\| \\
=&\left\|\int\int \left\{\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)-\boldsymbol{\Sigma}_y(u,v)\right\}\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)^{\mathrm{T}}\right. \\
&\left.+\boldsymbol{\Sigma}_y(u,v)\left\{\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)^{\mathrm{T}}-\boldsymbol{\Sigma}_y(u,v)^{\mathrm{T}}\right\}\mathrm{d}u\mathrm{d}v\right\| \\
=&\left\|\int\int \left\{\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)-\boldsymbol{\Sigma}_y(u,v)\right\}\left\{\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)^{\mathrm{T}}-\boldsymbol{\Sigma}_y(u,v)^{\mathrm{T}}+\boldsymbol{\Sigma}_y(u,v)^{\mathrm{T}}\right\}\right. \\
&\left.+\boldsymbol{\Sigma}_y(u,v)\left\{\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)^{\mathrm{T}}-\boldsymbol{\Sigma}_y(u,v)^{\mathrm{T}}\right\}\mathrm{d}u\mathrm{d}v\right\| \\
\leqslant&\left\|\int\int \left\{\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)-\boldsymbol{\Sigma}_y(u,v)\right\}\left\{\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)-\boldsymbol{\Sigma}_y(u,v)\right\}^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\| \\
&+2\left\|\int\int \left\{\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)-\boldsymbol{\Sigma}_y(u,v)\right\}\boldsymbol{\Sigma}_y(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\| \\
\leqslant&\|\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}-\boldsymbol{\Sigma}_y\|_{\mathcal{L}}^2+2\|\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}-\boldsymbol{\Sigma}_y\|_{\mathcal{L}}\|\boldsymbol{\Sigma}_y\|_{\mathcal{L}} \\
\leqslant&\|\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}-\boldsymbol{\Sigma}_y\|_{\mathcal{S},\mathrm{F}}^2+2\|\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}-\boldsymbol{\Sigma}_y\|_{\mathcal{S},\mathrm{F}}\|\boldsymbol{\Sigma}_y\|_{\mathcal{S},\mathrm{F}} \\
=&\sum_{i=1}^p\sum_{j=1}^p\|n^{-1}\sum_{t=1}^n y_{ti}y_{tj}-\Sigma_{y,ij}\|_{\mathcal{S}}^2+2\left(\sum_{i=1}^p\sum_{j=1}^p\|n^{-1}\sum_{t=1}^n y_{ti}y_{tj}-\Sigma_{y,ij}\|_{\mathcal{S}}^2\right)^{1/2}\|\boldsymbol{\Sigma}_y\|_{\mathcal{S},\mathrm{F}} \\
=&O_p\left(\mathcal{M}_\varepsilon p^2\sqrt{1/n}\right)=o_p(p^2),
\end{aligned}
\tag{S.5}
$$

where the second inequality follows from Lemmas A5(ii) and A6(iv), the third inequality follows from Lemma A7(ii), and the last line follows from Lemmas B12(iii), the fact that $\|\mathbf{K}\|_{\mathcal{S},\mathrm{F}}\leqslant p\|\mathbf{K}\|_{\mathcal{S},\max}$ and Assumption 4(ii).

(ii) The argument can be proved in matrix ℓ_∞ norm following the similar procedure. Specifically,

$$
\begin{aligned}
\|\widehat{\boldsymbol{\Omega}}-\boldsymbol{\Omega}\|_\infty \leqslant&\left\|\int\int \left\{\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)-\boldsymbol{\Sigma}_y(u,v)\right\}\left\{\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)-\boldsymbol{\Sigma}_y(u,v)\right\}^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\|_\infty \\
&+2\left\|\int\int \left\{\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)-\boldsymbol{\Sigma}_y(u,v)\right\}\boldsymbol{\Sigma}_y(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v\right\|_\infty \\
\leqslant&\|\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}-\boldsymbol{\Sigma}_y\|_{\mathcal{S},\infty}\|\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}-\boldsymbol{\Sigma}_y\|_{\mathcal{S},1}+2\|\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}-\boldsymbol{\Sigma}_y\|_{\mathcal{S},\infty}\|\boldsymbol{\Sigma}_y\|_{\mathcal{S},1} \\
\asymp&p^2\|\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}-\boldsymbol{\Sigma}_y\|_{\mathcal{S},\max}=O_p\Big(\mathcal{M}_\varepsilon p^2\sqrt{\frac{\log p}{n}}\Big)=o_p(p^2),
\end{aligned}
$$

where the first inequality can be obtained in a way similar to (S.5), the second inequality follows from Lemma A6(ii), and the last line follows from Lemma B12(iii) and Assumption 4(ii). $\square$

Lemma B23. *Let $\{\hat{\lambda}_j\}_{j=1}^p$ be the eigenvalues of $\widehat{\mathbf{\Omega}}$ in a descending order. Under the assumptions of Theorem 1, it holds that $\hat{\lambda}_r \gtrsim p^2$ with probability approaching 1. Furthermore, $\hat{\lambda}_i - \hat{\lambda}_j \gtrsim p^2$ for all $1 \leqslant i < j \leqslant r$ with probability approaching 1.*

Proof. By Proposition 1 and Lemma B20, the r-th largest eigenvalue λ_r of $\mathbf{\Omega}$ satisfies

$$\lambda_r \geqslant p^2\theta_r - |\lambda_r - p^2\theta_r| \geqslant p^2\theta_r - \|\mathbf{\Omega}_{\mathcal{R}}\| \gtrsim p^2.$$

Applying Lemma A1 yields that

$$|\hat{\lambda}_j - \lambda_j| \leqslant \|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}\|, \text{ for } j \in [p].$$

From Lemma B22(i), we have $\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}\| = o_p(p^2)$ and hence $\hat{\lambda}_r \gtrsim p^2$ with probability approaching 1. Furthermore, for all $1 \leqslant i < j \leqslant r$,

$$\hat{\lambda}_i - \hat{\lambda}_j \geqslant (\lambda_i - \lambda_j) - |\hat{\lambda}_i - \lambda_i| - |\hat{\lambda}_j - \lambda_j| = p^2(\theta_i - \theta_j) - o_p(p^2) \gtrsim p^2$$

with probability approaching 1. $\square$

Lemma B24. *Under the assumptions of Theorem 1, we have*

$$\|\widehat{\mathbf{B}} - \mathbf{B}\mathbf{U}^{\mathrm{T}}\|_{\mathrm{F}} = O_p(\mathcal{M}_{\varepsilon}\sqrt{p/n} + 1/\sqrt{p}).$$

Proof. By Proposition 1(ii) and Lemma B20, if $\boldsymbol{\xi}_j^{\mathrm{T}}\widetilde{\mathbf{b}}_j \geqslant 0$, then

$$\|\boldsymbol{\xi}_j - \widetilde{\mathbf{b}}_j\| = O_p(p^{-2}\|\mathbf{\Omega}_{\mathcal{R}}\|) = O_p(p^{-1}), \text{ for } j \in [r]. \tag{S.6}$$

Applying Lemma A2 yields that, if $\widehat{\boldsymbol{\xi}}_j^{\mathrm{T}}\boldsymbol{\xi}_j \geqslant 0$, we have

$$\|\widehat{\boldsymbol{\xi}}_j - \boldsymbol{\xi}_j\| \leqslant \frac{\|\widehat{\mathbf{\Omega}} - \mathbf{\Omega}\|/\sqrt{2}}{\min\Big(|\hat{\lambda}_{j-1} - \lambda_j|, |\lambda_j - \hat{\lambda}_{j+1}|\Big)}, \tag{S.7}$$

where $\{\hat{\lambda}_j\}_{j=1}^p$ are the eigenvalues of $\widehat{\mathbf{\Omega}}$ in a descending order, and $\{\widehat{\boldsymbol{\xi}}_j\}_{j=1}^p$ are their corresponding eigenvectors. Then, for $j \in [r]$, we have $|\hat{\lambda}_{j-1} - \lambda_j| \geqslant |\hat{\lambda}_{j-1} - \hat{\lambda}_j| - |\lambda_j - \hat{\lambda}_j|$, where the first term $|\hat{\lambda}_{j-1} - \hat{\lambda}_j| \gtrsim p^2$ with probability approaching 1 by Lemma B23, and the second term $|\lambda_j - \hat{\lambda}_j| = o_p(p^2)$ by Lemmas A1 and B22(i). Hence, $|\hat{\lambda}_{j-1} - \lambda_j| \gtrsim p^2$ with probability approaching 1 for all $j \in [r]$. We can also show the similar result for $|\lambda_j - \hat{\lambda}_{j+1}|$ if $j \in [r-1]$. If $j = r$, $|\lambda_r - \hat{\lambda}_{r+1}| > \lambda_r - \hat{\lambda}_{r+1} = p^2\theta_r - \hat{\lambda}_{r+1} \gtrsim p^2$ since $\hat{\lambda}_{r+1} = o_p(p^2)$, which can be implied by Proposition 1 and Lemma B20 that $\lambda_{r+1} = o(p^2)$, and Lemmas A1 and B22(i) that $|\hat{\lambda}_{r+1} - \lambda_{r+1}| = o_p(p^2)$. Thus,

$$\min\left(|\hat{\lambda}_{j-1} - \lambda_j|, |\lambda_j - \hat{\lambda}_{j+1}|\right) \gtrsim p^2.$$

Applying (S.7), Lemma B22(i) and the above argument, we have, if $\widehat{\boldsymbol{\xi}}_j^{\mathrm{T}} \boldsymbol{\xi}_j \geqslant 0$, then

$$\|\widehat{\boldsymbol{\xi}}_j - \boldsymbol{\xi}_j\| = O_p\Big(\mathcal{M}_\varepsilon \sqrt{1/n}\Big), \text{ for } j \in [r].$$

Combing with (S.6) we have, if $\widehat{\boldsymbol{\xi}}_j^{\mathrm{T}} \widetilde{\mathbf{b}}_j \geqslant 0$, then

$$\|\widehat{\boldsymbol{\xi}}_j - \widetilde{\mathbf{b}}_j\| = O_p\Big(\mathcal{M}_\varepsilon \sqrt{1/n} + 1/p\Big), \text{ for } j \in [r].$$

Since $\widehat{\mathbf{b}}_j = \sqrt{p}\widehat{\boldsymbol{\xi}}_j$ and $\mathbf{b}_j = \sqrt{p}\widetilde{\mathbf{b}}_j$, one can obtain that there exists an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{r\times r}$ such that

$$\|\widehat{\mathbf{B}} - \mathbf{B}\mathbf{U}^{\mathrm{T}}\|_{\mathrm{F}} = O_p\Big(\mathcal{M}_\varepsilon \sqrt{p/n} + 1/\sqrt{p}\Big),$$

where the matrix $\mathbf{U}$ is used to adjust the direction so that each $\mathbf{b}_j^{\mathrm{T}} \widehat{\mathbf{b}}_j \geqslant 0$ for $j \in [r]$. □

We are now ready to prove Theorem 1.

(i) Let $\mathbf{E} = \widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}$ be the $p \times p$ perturbation matrix. By Lemma B22, we have

$$\|\mathbf{E}\|_\infty \leqslant \|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\|_\infty = O_p\left(\mathcal{M}_\varepsilon p^2 \sqrt{\frac{\log p}{n}}\right) = o_p(p^2).$$

Corresponding to Lemma B13, here $\mathbf{A} = \boldsymbol{\Omega}$, $\widetilde{\mathbf{A}} = \widehat{\boldsymbol{\Omega}}$, and the r-th eigenvalue of $\mathbf{A}$ satisfies $\lambda_r \asymp p^2$ by Proposition 1 and Lemma B20. Then, $\|\mathbf{A} - \mathbf{A}_r\|_\infty \leqslant \|\boldsymbol{\Omega} - \boldsymbol{\Omega}_{\mathcal{L}}\|_\infty = \|\boldsymbol{\Omega}_{\mathcal{R}}\|_\infty = ps_p = o(p^2)$ from Lemma B21. Note that $\mathbf{V} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_r) \in \mathbb{R}^{p\times r}$, and denote $\boldsymbol{\xi}_j = (\xi_{1j}, \ldots, \xi_{pj})^{\mathrm{T}}$. The coherence of $\mathbf{V}$ is given by

$$\mu = \mu(\mathbf{V}) = \frac{p}{r} \max_{i\in[p]} \sum_{j=1}^r \xi_{ij}^2 \leqslant \frac{p}{r} \max_{i\in[p]} \sum_{j=1}^r \left(\widetilde{b}_{ij}^2 + \|\boldsymbol{\xi}_j - \widetilde{\mathbf{b}}_j\|^2\right) = O(1),$$

since $\max_{i\in[p]} \widetilde{b}_{ij} = \max_{i\in[p]} p^{-1/2} b_{ij} \leqslant p^{-1/2}\|\mathbf{B}\|_{\max} = O(p^{-1/2})$ and $\|\boldsymbol{\xi}_j - \widetilde{\mathbf{b}}_j\| = O(p^{-1})$ if $\boldsymbol{\xi}_j^{\mathrm{T}}\widetilde{\mathbf{b}}_j \geqslant 0$ by Proposition 1(ii) and Lemma B20. In addition, supposing that $\iota \asymp \|\mathbf{E}\| = o_p(p^2)$ but $\iota > \|\mathbf{E}\|$, we can show that for any $j \in [r]$, the the interval $(\lambda_j - \iota, \lambda_j + \iota)$ does not contain any eigenvalues of $\boldsymbol{\Omega}$ other than λ_j for efficient large p. Thus by applying Lemma B13, we have for $j \in [r]$, if $\widehat{\boldsymbol{\xi}}_j^{\mathrm{T}}\boldsymbol{\xi}_j \geqslant 0$,

$$\|\widehat{\boldsymbol{\xi}}_j - \boldsymbol{\xi}_j\|_{\max} = O\left(\frac{r^{5/2}\mu^2\|\mathbf{E}\|_{\infty}}{p^2\sqrt{p}}\right) = O_p\left(\mathcal{M}_{\varepsilon}\sqrt{\frac{\log p}{pn}}\right).$$

For $j \in [r]$, if $\boldsymbol{\xi}_j^{\mathrm{T}}\widetilde{\mathbf{b}}_j \geqslant 0$, we have $\|\boldsymbol{\xi}_j - \widetilde{\mathbf{b}}_j\|_{\max} \leqslant \|\boldsymbol{\xi}_j - \widetilde{\mathbf{b}}_j\| = O(p^{-1})$, which implies that $\|\widehat{\boldsymbol{\xi}}_j - \widetilde{\mathbf{b}}_j\|_{\max} = O_p(\mathcal{M}_{\varepsilon}\sqrt{\log p/pn} + 1/p)$. Since $\widehat{\mathbf{B}} = \sqrt{p}(\widehat{\boldsymbol{\xi}}_1, \ldots, \widehat{\boldsymbol{\xi}}_r)$ and $\mathbf{B} = \sqrt{p}(\widetilde{\mathbf{b}}_1, \ldots, \widetilde{\mathbf{b}}_r)$, one can obtain that there exists an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{r\times r}$ (=the same as that in Lemma B24) such that

$$\|\widehat{\mathbf{B}} - \mathbf{B}\mathbf{U}^{\mathrm{T}}\|_{\max} = O_p\left(\mathcal{M}_{\varepsilon}\sqrt{\frac{\log p}{n}} + \frac{1}{\sqrt{p}}\right) = o_p(1),$$

where the matrix $\mathbf{U}$ is used to adjust the direction so that each $\mathbf{b}_j^{\mathrm{T}}\widehat{\mathbf{b}}_j \geqslant 0$ for $j \in [r]$.

(ii) Note that $\widehat{\mathbf{f}}_t(\cdot) = p^{-1}\widehat{\mathbf{B}}^{\mathrm{T}}\mathbf{y}_t(\cdot) = p^{-1}\widehat{\mathbf{B}}^{\mathrm{T}}\{\mathbf{B}\mathbf{f}_t(\cdot) + \boldsymbol{\varepsilon}_t(\cdot)\}$for $t \in [n]$ and then

$$\widehat{\mathbf{f}}_t(\cdot) - \mathbf{U}\mathbf{f}_t(\cdot) = p^{-1}(\widehat{\mathbf{B}}^{\mathrm{T}}\mathbf{B} - \mathbf{U}\mathbf{B}^{\mathrm{T}}\mathbf{B})\mathbf{f}_t(\cdot) + p^{-1}\widehat{\mathbf{B}}^{\mathrm{T}}\boldsymbol{\varepsilon}_t(\cdot). \tag{S.8}$$

For the first term of (S.8), applying Lemmas B11 and B24 yields that

$$\|p^{-1}(\widehat{\mathbf{B}}^{\mathrm{T}}\mathbf{B} - \mathbf{U}\mathbf{B}^{\mathrm{T}}\mathbf{B})\mathbf{f}_t\| \leqslant p^{-1}\|\widehat{\mathbf{B}}^{\mathrm{T}} - \mathbf{U}\mathbf{B}^{\mathrm{T}}\| \cdot \|\mathbf{B}\| \cdot \|\mathbf{f}_t\| = O_p\left(\mathcal{M}_{\varepsilon}\sqrt{1/n} + 1/p\right),$$

since $\|\widehat{\mathbf{B}}^{\mathrm{T}} - \mathbf{U}\mathbf{B}^{\mathrm{T}}\| \leqslant \|\widehat{\mathbf{B}} - \mathbf{B}\mathbf{U}^{\mathrm{T}}\|_{\mathrm{F}} = O_p(\mathcal{M}_{\varepsilon}\sqrt{p/n} + 1/\sqrt{p}), \|\mathbf{B}\| = \lambda_{\max}^{1/2}(\mathbf{B}^{\mathrm{T}}\mathbf{B}) = \sqrt{p}$ and $\|\mathbf{f}_t\| = O_p(1)$. For the second term of (S.8), denote $\breve{\mathbf{b}}_i \in \mathbb{R}^r$ the i-th row of $\widehat{\mathbf{B}}$ and

$$\begin{aligned}\|p^{-1}\widehat{\mathbf{B}}^{\mathrm{T}}\boldsymbol{\varepsilon}_t\| =& \Big\|p^{-1}\sum_{i=1}^p \breve{\mathbf{b}}_i\varepsilon_{ti}\Big\| = p^{-1}\Big\{\int\sum_{i=1}^p \varepsilon_{ti}(u)^2\breve{\mathbf{b}}_i\breve{\mathbf{b}}_i^{\mathrm{T}}\mathrm{d}u\Big\}^{1/2}\\ \leqslant& r^{1/2}\|\widehat{\mathbf{B}}\|_{\max}p^{-1}\|\boldsymbol{\varepsilon}_t\| = O_p(1/\sqrt{p}),\end{aligned}$$

since $\|\widehat{\mathbf{B}}\|_{\max} \leqslant C + o_p(1)$ and $\|\boldsymbol{\varepsilon}_t\| = O_p(p^{1/2})$ by Assumption 3(ii). The result follows immediately that for $t \in [n]$,

$$\|\widehat{\mathbf{f}}_t - \mathbf{U}\mathbf{f}_t\| = O_p\left(\mathcal{M}_{\varepsilon}/\sqrt{n} + 1/\sqrt{p}\right),$$

and thus $n^{-1}\sum_{t=1}^{n}\|\widehat{\mathbf{f}}_t-\mathbf{U}\mathbf{f}_t\|^2=O_p(\mathcal{M}_\varepsilon^2/n+1/p)$.

(iii) The proof procedure is similar to part (ii). We only need to notice that $\max_{t\in[n]}\|\mathbf{f}_t\|=O_p(\sqrt{\log n})$ by Lemma B17 and $\max_{t\in[n]}\|\boldsymbol{\varepsilon}_t\|=O_p(n^{1/4}p^{1/2})$ by applying the Chebyshev's inequality and Bonferroni's method combined with Assumption 3(ii).

B.3 Proof of Corollary 1

By Theorem 1(i)(iii), Lemma B17, and Assumption 3(i), we have

$$
\begin{aligned}
\max_{i\in[p],t\in[n]}\|\breve{\mathbf{b}}_i^{\mathrm{T}}\widehat{\mathbf{f}}_t-\breve{\mathbf{b}}_i^{\mathrm{T}}\mathbf{f}_t\| \leqslant & \max_{i\in[p]}\|\breve{\mathbf{b}}_i-\mathbf{U}\breve{\mathbf{b}}_i\|\cdot\max_{t\in[n]}\|\widehat{\mathbf{f}}_t\|+\max_{i\in[p]}\|\breve{\mathbf{b}}_i\|\cdot\max_{t\in[n]}\|\mathbf{U}^{\mathrm{T}}\widehat{\mathbf{f}}_t-\mathbf{f}_t\| \\
=&O_p(\varpi_{n,p}\cdot\sqrt{\log n})+O_p(\mathcal{M}_\varepsilon\sqrt{\log n/n}+n^{1/4}/p^{1/2}) \\
=&O_p\Big\{(\log n)^{1/2}\mathcal{M}_\varepsilon\sqrt{\frac{\log p}{n}}+\frac{n^{1/4}}{\sqrt{p}}\Big\}.
\end{aligned}
$$

B.4 Proof of Theorem 3

To prove Theorem 3, we first present some technical lemmas with their proofs.

Lemma B25. *Under the assumptions of Theorem 3, it holds that*

(i) $\max_{i\in[p]}n^{-1}\sum_{t=1}^{n}\|\widehat{\varepsilon}_{ti}-\varepsilon_{ti}\|^2=O_p(\varpi_{n,p}^2)$*;*

(ii) $\max_{i,j\in[p]}\|n^{-1}\sum_{t=1}^{n}\widehat{\varepsilon}_{ti}\widehat{\varepsilon}_{tj}-n^{-1}\sum_{t=1}^{n}\varepsilon_{ti}\varepsilon_{tj}\|_{\mathcal{S}}=O_p(\varpi_{n,p})$*;*

(iii) $\|\widehat{\boldsymbol{\Sigma}}_\varepsilon-\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\max}=O_p(\varpi_{n,p})$*.*

Proof. (i) Notice that $\widehat{\varepsilon}_{ti}(\cdot)-\varepsilon_{ti}(\cdot)=\{y_{ti}(\cdot)-\breve{\mathbf{b}}_i^{\mathrm{T}}\mathbf{f}_t(\cdot)\}-\{y_{ti}(\cdot)-\breve{\mathbf{b}}_i^{\mathrm{T}}\widehat{\mathbf{f}}_t(\cdot)\}=(\breve{\mathbf{b}}_i-\mathbf{U}\breve{\mathbf{b}}_i)^{\mathrm{T}}\widehat{\mathbf{f}}_t(\cdot)-\breve{\mathbf{b}}_i^{\mathrm{T}}(\mathbf{U}^{\mathrm{T}}\widehat{\mathbf{f}}_t-\mathbf{f}_t)(\cdot)$, where $\breve{\mathbf{b}}_i$ and $\breve{\mathbf{b}}_i$ are the i-th rows of $\mathbf{B}$ and $\widehat{\mathbf{B}}$, respectively. Applying the inequality $(a+b)^2\leqslant 2a^2+2b^2$ and the Cauchy–Schwartz inequality yields that

$$
\begin{aligned}
\max_{i\in[p]}\frac{1}{n}\sum_{t=1}^{n}\|\widehat{\varepsilon}_{ti}-\varepsilon_{ti}\|^2 \leqslant & 2\max_{i\in[p]}\|\breve{\mathbf{b}}_i-\mathbf{U}\breve{\mathbf{b}}_i\|^2\frac{1}{n}\sum_{t=1}^{n}\|\widehat{\mathbf{f}}_t\|^2+2\max_{i\in[p]}\|\breve{\mathbf{b}}_i\|^2\frac{1}{n}\sum_{t=1}^{n}\|\mathbf{U}^{\mathrm{T}}\widehat{\mathbf{f}}_t-\mathbf{f}_t\|^2 \\
=&O_p(\varpi_{n,p}^2)+O_p(\mathcal{M}_\varepsilon^2/n+1/p)=O_p(\varpi_{n,p}^2).
\end{aligned}
$$

(ii) Notice that $\max_{i\in[p]}\mathbb{E}\|\varepsilon_{ti}\|^2=\max_{i\in[p]}\mathbb{E}\int\varepsilon_{ti}(u)^2\mathrm{d}u=\max_{i\in[p]}\int\Sigma_{\varepsilon,ii}(u,u)\mathrm{d}u=O(1)$ from Assumption 3(iv), thus we have $\max_{i\in[p]}n^{-1}\sum_{t=1}^{n}\|\varepsilon_{ti}\|^2=O_p(1)$. By the Cauchy–Schwartz inequality,

$$\begin{aligned}\max_{i,j\in[p]}\Big\|\frac{1}{n}\sum_{t=1}^{n}\widehat{\varepsilon}_{ti}\widehat{\varepsilon}_{tj}-\frac{1}{n}\sum_{t=1}^{n}\varepsilon_{ti}\varepsilon_{tj}\Big\|_{\mathcal{S}}=&\max_{i,j\in[p]}\Big\|\frac{1}{n}\sum_{t=1}^{n}(\widehat{\varepsilon}_{ti}-\varepsilon_{ti})\widehat{\varepsilon}_{tj}+\varepsilon_{ti}(\widehat{\varepsilon}_{tj}-\varepsilon_{tj})\Big\|_{\mathcal{S}}\\ \leqslant&\max_{i\in[p]}\frac{1}{n}\sum_{t=1}^{n}\|\widehat{\varepsilon}_{ti}-\varepsilon_{ti}\|^2\\ &+2\Big(\max_{i\in[p]}\frac{1}{n}\sum_{t=1}^{n}\|\varepsilon_{ti}\|^2\Big)^{1/2}\Big(\max_{j\in[p]}\frac{1}{n}\sum_{t=1}^{n}\|\widehat{\varepsilon}_{tj}-\varepsilon_{tj}\|^2\Big)^{1/2}\\ =&O_p(\varpi_{n,p}^2)+O_p(\varpi_{n,p})=O_p(\varpi_{n,p}).\end{aligned}$$

(iii) The result is immediately implied by part (ii) above and Lemma B12(ii). □

Lemma B26. *Under the assumptions of Theorem 3, there exist some constants* $\Theta_1,\Theta_2>0$ *such that with probability approaching 1,*

$$\Theta_1\leqslant\min_{i\in[p],j\in[p]}\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}\leqslant\max_{i\in[p],j\in[p]}\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}\leqslant\Theta_2.$$

Proof. We first prove the upper bound. By the definition of $\widehat{\Theta}_{ij}$, we have

$$\begin{aligned}\widehat{\Theta}_{ij}(u,v)=&\frac{1}{n}\sum_{t=1}^{n}\Big\{\widehat{\varepsilon}_{ti}(u)\widehat{\varepsilon}_{tj}(v)-\frac{1}{n}\sum_{s=1}^{n}\widehat{\varepsilon}_{si}(u)\widehat{\varepsilon}_{sj}(v)\Big\}^2\\ \leqslant&\frac{2}{n}\sum_{t=1}^{n}\Big\{\widehat{\varepsilon}_{ti}(u)\widehat{\varepsilon}_{tj}(v)-\Sigma_{\varepsilon,ij}(u,v)\Big\}^2+2\max_{i\in[p],j\in[p]}\Big\{\Sigma_{\varepsilon,ij}(u,v)-\frac{1}{n}\sum_{s=1}^{n}\widehat{\varepsilon}_{si}(u)\widehat{\varepsilon}_{sj}(v)\Big\}^2,\end{aligned}$$

which implies that

$$\begin{aligned}\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}^2=&\int\int\widehat{\Theta}_{ij}(u,v)\mathrm{d}u\mathrm{d}v\leqslant\frac{2}{n}\int\int\sum_{t=1}^{n}\{\widehat{\varepsilon}_{ti}(u)\widehat{\varepsilon}_{tj}(v)-\Sigma_{\varepsilon,ij}(u,v)\}^2\,\mathrm{d}u\mathrm{d}v+2\|\widehat{\boldsymbol{\Sigma}}_{\varepsilon}-\boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{S},\max}^2\\ =&\frac{2}{n}\int\int\sum_{t=1}^{n}\Big\{\widehat{\varepsilon}_{ti}(u)\widehat{\varepsilon}_{tj}(v)-\Sigma_{\varepsilon,ij}(u,v)\Big\}^2\mathrm{d}u\mathrm{d}v+o_p(1),\end{aligned}$$

where the last line follows from Lemma B25. Moreover

$$
\begin{aligned}
&\sum_{t=1}^{n}\Big\{\widehat{\varepsilon}_{ti}(u)\widehat{\varepsilon}_{tj}(v)-\Sigma_{\varepsilon,ij}(u,v)\Big\}^2\\
=&\sum_{t=1}^{n}\Big[\big\{\widehat{\varepsilon}_{ti}(u)-\varepsilon_{ti}(u)\big\}\widehat{\varepsilon}_{tj}(v)+\varepsilon_{ti}(u)\big\{\widehat{\varepsilon}_{tj}(v)-\varepsilon_{tj}(v)\big\}\\
&+\varepsilon_{ti}(u)\varepsilon_{tj}(v)-\Sigma_{\varepsilon,ij}(u,v)\Big]^2\\
\leqslant&4\sum_{t=1}^{n}\big\{\widehat{\varepsilon}_{ti}(u)-\varepsilon_{ti}(u)\big\}^2\widehat{\varepsilon}_{tj}(v)^2+4\sum_{t=1}^{n}\varepsilon_{ti}(u)^2\big\{\widehat{\varepsilon}_{tj}(v)-\varepsilon_{tj}(v)\big\}^2\\
&+2\sum_{t=1}^{n}\big\{\varepsilon_{ti}(u)\varepsilon_{tj}(v)-\Sigma_{\varepsilon,ij}(u,v)\big\}^2\\
\leqslant&4\max_{i\in[p],t\in[n]}\big\{\widehat{\varepsilon}_{ti}(u)-\varepsilon_{ti}(u)\big\}^2\max_{j\in[p]}\Big[\sum_{t=1}^{n}2\big\{\widehat{\varepsilon}_{tj}(v)-\varepsilon_{tj}(v)\big\}^2+3\varepsilon_{tj}(v)^2\Big]\\
&+2\sum_{t=1}^{n}\big\{\varepsilon_{ti}(u)\varepsilon_{tj}(v)-\Sigma_{\varepsilon,ij}(u,v)\big\}^2.
\end{aligned}
$$

Here, we bound each term above as follows: (a) by Corollary 1, we have $\max_{i\in[p],t\in[n]}\|\widehat{\varepsilon}_{ti}-\varepsilon_{ti}\|^2=O_p\big\{(\log n)^{1/2}\mathcal{M}_\varepsilon\sqrt{\log p/n}+n^{1/4}/\sqrt{p}\big\}=o_p(1)$ under Assumption 6; (b) by Lemma B25(i), $\max_{j\in[p]}n^{-1}\sum_{t=1}^{n}\|\widehat{\varepsilon}_{tj}-\varepsilon_{tj}\|^2=O_p(\varpi_{n,p}^2)=o_p(1)$; (c) by Lemma B12(ii) and Assumption 3(iv), $\max_{j\in[p]}n^{-1}\sum_{t=1}^{n}\|\varepsilon_{tj}\|^2\leqslant o_p(1)+\max_{j\in[p]}\int\Sigma_{\varepsilon,jj}(u,u)\mathrm{d}u=O_p(1)$. Combing these results yields that

$$\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}^2\leqslant\frac{2}{n}\int\int\sum_{t=1}^{n}\big\{\varepsilon_{ti}(u)\varepsilon_{tj}(v)-\Sigma_{\varepsilon,ij}(u,v)\big\}^2\mathrm{d}u\mathrm{d}v+o_p(1).$$

Similar arguments as those in the proof of Lemma Cai and Liu (2011) results in

$$\max_{i\in[p],j\in[p]}\Big\|\frac{1}{n}\sum_{t=1}^{n}(\varepsilon_{ti}\varepsilon_{tj}-\Sigma_{\varepsilon,ij})^2-\mathrm{Var}(\varepsilon_{ti}\varepsilon_{tj})\Big\|_{\mathcal{S}}=o_p(1).$$

Combining with Assumption 5 implies that $\max_{i\in[p],j\in[p]}\|n^{-1}\sum_{t=1}^{n}(\varepsilon_{ti}\varepsilon_{tj}-\Sigma_{\varepsilon,ij})^2\|_{\mathcal{S}}$ is bounded away from both zero and infinity with probability approaching 1. Therefore, $\max_{i,j\in[p]}\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}$ is bounded away from infinity with probability approaching 1.

We next prove the lower bound. Notice that

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\Big\{\varepsilon_{ti}(u)\varepsilon_{tj}(v)-\Sigma_{\varepsilon,ij}(u,v)\Big\}^2 \leqslant & 4\sum_{t=1}^{n}\Big\{\varepsilon_{ti}(u)\varepsilon_{tj}(v)-\widehat{\varepsilon}_{ti}(u)\widehat{\varepsilon}_{tj}(v)\Big\}^2\\ &+4\sum_{t=1}^{n}\Big\{\widehat{\varepsilon}_{ti}(u)\widehat{\varepsilon}_{tj}(v)-\frac{1}{n}\sum_{s=1}^{n}\widehat{\varepsilon}_{si}(u)\widehat{\varepsilon}_{sj}(v)\Big\}^2\\ &+2\sum_{t=1}^{n}\Big\{\frac{1}{n}\sum_{s=1}^{n}\widehat{\varepsilon}_{si}(u)\widehat{\varepsilon}_{sj}(v)-\Sigma_{\varepsilon,ij}(u,v)\Big\}^2,\end{aligned}$$

which implies that

$$\begin{aligned}\frac{1}{n}\int\int\sum_{t=1}^{n}\Big\{\varepsilon_{ti}(u)\varepsilon_{tj}(v)-\Sigma_{\varepsilon,ij}(u,v)\Big\}^2\mathrm{d}u\mathrm{d}v \leqslant & \frac{4}{n}\int\int\sum_{t=1}^{n}[\varepsilon_{ti}(u)\varepsilon_{tj}(v)-\widehat{\varepsilon}_{ti}(u)\widehat{\varepsilon}_{tj}(v)]^2\mathrm{d}u\mathrm{d}v\\ &+4\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}^2+o_p(1),\end{aligned}$$

where the LHS is bounded away from both zero and infinity uniformly in i,j. Then,

$$\begin{aligned}&\sum_{t=1}^{n}\Big\{\varepsilon_{ti}(u)\varepsilon_{tj}(v)-\widehat{\varepsilon}_{ti}(u)\widehat{\varepsilon}_{tj}(v)\Big\}^2\\ \leqslant & 2\sum_{t=1}^{n}\varepsilon_{ti}(u)^2\big\{\varepsilon_{tj}(v)-\widehat{\varepsilon}_{tj}(v)\big\}^2+2\sum_{t=1}^{n}\widehat{\varepsilon}_{jt}(v)^2\big\{\varepsilon_{ti}(v)-\widehat{\varepsilon}_{ti}(u)\big\}^2\\ \leqslant & 4\max_{i\in[p],t\in[n]}\big\{\widehat{\varepsilon}_{ti}(u)-\varepsilon_{ti}(u)\big\}^2\max_{j\in[p]}\sum_{t=1}^{n}\Big[\big\{[\widehat{\varepsilon}_{jt}(v)-\varepsilon_{tj}(v)\big\}^2+\varepsilon_{jt}(v)^2\Big].\end{aligned}$$

As demonstrated in the proof of the upper bound above, we have

$$\frac{1}{n}\int\int\sum_{t=1}^{n}\big\{\varepsilon_{ti}(u)\varepsilon_{tj}(v)-\widehat{\varepsilon}_{ti}(u)\widehat{\varepsilon}_{tj}(v)\big\}^2\mathrm{d}u\mathrm{d}v=o_p(1).$$

Hence, $\min_{i\in[p],j\in[p]}\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}$ is bounded away from zero with probability approaching 1. □

We are now ready to prove Theorem 3. By Lemmas B25(iii) and B26, we have $\|\widehat{\mathbf{\Sigma}}_{\varepsilon}-\mathbf{\Sigma}_{\varepsilon}\|_{\mathcal{S},\max}=O_p(\varpi_{n,p})$ and $\max_{ij\in[p]}\|\Theta_{ij}\|_{\mathcal{S}}=O_p(1)$. Consequently, for any $\epsilon>0$, there exist some positive constants N,Θ_1 and Θ_2 such that each of events

$$\Upsilon_1=\left\{\max_{i\in[p],j\in[p]}\Big\|\widehat{\Sigma}_{\varepsilon,ij}-\Sigma_{\varepsilon,ij}\Big\|_{\mathcal{S}}<N\varpi_{n,p}\right\},\ \Upsilon_2=\Big\{\Theta_1\leqslant\big\|\widehat{\Theta}_{ij}^{1/2}\big\|_{\mathcal{S}}\leqslant\Theta_2,\ \text{all } i,j\in[p]\Big\}$$

hold with probability at least $1-\epsilon$. The thresholding in (10) is equivalent to $\widehat{\Sigma}_{\varepsilon,ij}^{\mathcal{A}}=s_{ij}\big(\widehat{\Sigma}_{\varepsilon,ij}\big)$, where $s_{ij}(\cdot)\equiv s_{\lambda_{ij}}(\cdot)$ with $\lambda_{ij}=\dot{C}\omega_{n,p}\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}$ and $\omega_{n,p}=\sqrt{\log p/n}+1/\sqrt{p}$ which is smaller than $\varpi_{n,p}$. For $\dot{C} > 2N\Theta_1^{-1}(\varpi_{n,p}/\omega_{n,p})$, under the event $\Upsilon_1 \cap \Upsilon_2$, we obtain that

$$
\begin{aligned}
\|\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}} - \boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},1} =& \max_{i\in[p]} \sum_{j=1}^p \|\widehat{\Sigma}_{\varepsilon,ij}^{\mathcal{A}} - \Sigma_{\varepsilon,ij}\|_{\mathcal{S}} = \max_{i\in[p]} \sum_{j=1}^p \|s_{ij}(\widehat{\Sigma}_{\varepsilon,ij}) - \Sigma_{\varepsilon,ij}\|_{\mathcal{S}} \\
\leqslant& \max_{i\in[p]} \sum_{j=1}^p \|s_{ij}(\widehat{\Sigma}_{\varepsilon,ij}) - \widehat{\Sigma}_{\varepsilon,ij}\|_{\mathcal{S}} I\big(\|\widehat{\Sigma}_{\varepsilon,ij}\|_{\mathcal{S}} > \dot{C}\omega_{n,p}\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}\big) \\
&+ \max_{i\in[p]} \sum_{j=1}^p \|\widehat{\Sigma}_{\varepsilon,ij} - \Sigma_{\varepsilon,ij}\|_{\mathcal{S}} I\big(\|\widehat{\Sigma}_{\varepsilon,ij}\|_{\mathcal{S}} > \dot{C}\omega_{n,p}\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}\big) \\
&+ \max_{i\in[p]} \sum_{j=1}^p \|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}} I\big(\|\widehat{\Sigma}_{\varepsilon,ij}\|_{\mathcal{S}} \leqslant \dot{C}\omega_{n,p}\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}\big) \\
\leqslant& \max_{i\in[p]} \sum_{j=1}^p \lambda_{ij} I\big(\|\widehat{\Sigma}_{\varepsilon,ij}\|_{\mathcal{S}} > \dot{C}\omega_{n,p}\Theta_1\big) + \max_{i\in[p]} \sum_{j=1}^p N\varpi_{n,p} I\big(\|\widehat{\Sigma}_{\varepsilon,ij}\|_{\mathcal{S}} > \dot{C}\omega_{n,p}\Theta_1\big) \\
&+ \max_{i\in[p]} \sum_{j=1}^p \|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}} I\big(\|\widehat{\Sigma}_{\varepsilon,ij}\|_{\mathcal{S}} \leqslant \dot{C}\omega_{n,p}\Theta_2\big) \\
\leqslant& (\dot{C}\Theta_2 + N)\varpi_{n,p} \max_{i\in[p]} \sum_{j=1}^p I\big(\|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}} > N\varpi_{n,p}\big) \\
&+ \max_{i\in[p]} \sum_{j=1}^p \|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}} I\big(\|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}} \leqslant (\dot{C}\Theta_2 + N)\varpi_{n,p}\big) \\
\leqslant& (\dot{C}\Theta_2 + N)\varpi_{n,p} \max_{i\in[p]} \sum_{j=1}^p \frac{\|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}}^q}{N^q\varpi_{n,p}^q} I\big(\|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}} > N\varpi_{n,p}\big) \\
&+ \max_{i\in[p]} \sum_{j=1}^p \|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}} \frac{(\dot{C}\Theta_2 + N)^{1-q}\varpi_{n,p}^{1-q}}{\|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}}^{1-q}} I\big(\|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}} \leqslant (\dot{C}\Theta_2 + N)\varpi_{n,p}\big) \\
\leqslant& (\dot{C}\Theta_2 + N)\big\{N^{-q} + (\dot{C}\Theta_2 + N)^{-q}\big\}\varpi_{n,p}^{1-q} \max_{i\in[p]} \sum_{j=1}^p \|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}}^q \\
\asymp& \varpi_{n,p}^{1-q} s_p,
\end{aligned}
$$

where the third inequality follows from $\dot{C}\Theta_1\omega_{n,p} > 2N\varpi_{n,p}$, and the last line follows from the fact that $s_p = \max_{i\in[p]} \sum_{j=1}^p \|\sigma_i\|_{\mathcal{N}}^{(1-q)/2}\|\sigma_j\|_{\mathcal{N}}^{(1-q)/2}\|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}}^q \asymp \max_{i\in[p]} \sum_{j=1}^p \|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}}^q$ since $\max_{i\in[p]} \|\sigma_i\|_{\mathcal{N}} = \max_{i\in[p]} \int \Sigma_\varepsilon(u,u)\mathrm{d}u = O(1)$ by Assumption 3(iv). Therefore, with probability at least $1 - 2\epsilon$, $\|\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}} - \boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},1} \lesssim \varpi_{n,p}^{1-q} s_p$. Considering that $\epsilon > 0$ can be arbitrarily small, we have the desired result

$$
\|\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}} - \boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}} \leqslant \|\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}} - \boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},1} = O_p(\varpi_{n,p}^{1-q} s_p).
$$

B.5 Proof of Theorem 4

To prove Theorem 4, we first present a technical lemma with its proof.

Lemma B27. *Suppose that the assumptions of Theorem 4 hold. For the sample covariance of* $\widehat{\mathbf{f}}_t$*, i.e.,* $\widehat{\mathbf{\Sigma}}_f(u,v) = n^{-1}\sum_{t=1}^n \widehat{\mathbf{f}}_t(u)\widehat{\mathbf{f}}_t(v)^{\mathrm{T}}$*, we have*

$$\|\widehat{\mathbf{\Sigma}}_f - \mathbf{U}\mathbf{\Sigma}_f\mathbf{U}^{\mathrm{T}}\|_{\mathcal{S},\max} = O_p(\mathcal{M}_\varepsilon/\sqrt{n} + 1/\sqrt{p}).$$

Proof. Consider $\widehat{\mathbf{f}}_t(u)\widehat{\mathbf{f}}_t(v)^{\mathrm{T}} - \mathbf{U}\mathbf{f}_t(u)\mathbf{f}_t(v)^{\mathrm{T}}\mathbf{U}^{\mathrm{T}} = \big\{\widehat{\mathbf{f}}_t(u) - \mathbf{U}\mathbf{f}_t(u)\big\}\widehat{\mathbf{f}}_t(v)^{\mathrm{T}} + \mathbf{U}\mathbf{f}_t(u)\big\{\widehat{\mathbf{f}}_t(v) - \mathbf{U}\mathbf{f}_t(v)]\big\}^{\mathrm{T}}$. Then

$$\begin{aligned}
\Big\|\frac{1}{n}\sum_{t=1}^n(\widehat{\mathbf{f}}_t\widehat{\mathbf{f}}_t^{\mathrm{T}} - \mathbf{U}\mathbf{f}_t\mathbf{f}_t^{\mathrm{T}}\mathbf{U}^{\mathrm{T}})\Big\|_{\mathcal{S},\max} \leqslant& \Big\|\frac{1}{n}\sum_{t=1}^n(\widehat{\mathbf{f}}_t - \mathbf{U}\mathbf{f}_t)\widehat{\mathbf{f}}_t^{\mathrm{T}}\Big\|_{\mathcal{S},\max} + \Big\|\frac{1}{n}\sum_{t=1}^n\mathbf{U}\mathbf{f}_t(\widehat{\mathbf{f}}_t - \mathbf{U}\mathbf{f}_t)^{\mathrm{T}}\Big\|_{\mathcal{S},\max}\\
\leqslant&\Big(\frac{1}{n}\sum_{t=1}^n\|\widehat{\mathbf{f}}_t - \mathbf{U}\mathbf{f}_t\|^2\Big)^{1/2}\left(\frac{1}{n}\sum_{t=1}^n\|\widehat{\mathbf{f}}_t\|^2\right)^{1/2}\\
&+\Big(\frac{1}{n}\sum_{t=1}^n\|\widehat{\mathbf{f}}_t - \mathbf{U}\mathbf{f}_t\|^2\Big)^{1/2}\left(\frac{1}{n}\sum_{t=1}^n\|\mathbf{U}\mathbf{f}_t\|^2\right)^{1/2}\\
=&O_p(\varpi_{n,p}),
\end{aligned}$$

where the second inequality follows from the Cauchy–Schwartz inequality, and the last line follows from $n^{-1}\sum_{t=1}^n\|\widehat{\mathbf{f}}_t - \mathbf{U}\mathbf{f}_t\|^2 = O_p(\mathcal{M}_\varepsilon^2/n + 1/p)$ by Theorem 1(ii), and $n^{-1}\sum_{t=1}^n\|\mathbf{U}\mathbf{f}_t\|^2 = O_p(1)$ since $\|\mathbf{U}\| = 1$ and $\mathbb{E}\|\mathbf{f}_t\|^2 = O(1)$. Together with Lemma B12(i), the desired result follows immediately. $\square$

We are now ready to prove Theorem 4. Consider that

$$\begin{aligned}
&\mathbf{B}\mathbf{\Sigma}_f\mathbf{B}^{\mathrm{T}} - \widehat{\mathbf{B}}\widehat{\mathbf{\Sigma}}_f\widehat{\mathbf{B}}^{\mathrm{T}} = \mathbf{B}\mathbf{U}^{\mathrm{T}}\mathbf{U}\mathbf{\Sigma}_f\mathbf{U}^{\mathrm{T}}\mathbf{U}\mathbf{B}^{\mathrm{T}} - \widehat{\mathbf{B}}\widehat{\mathbf{\Sigma}}_f\widehat{\mathbf{B}}^{\mathrm{T}}\\
=&\mathbf{B}\mathbf{U}^{\mathrm{T}}(\mathbf{U}\mathbf{\Sigma}_f\mathbf{U}^{\mathrm{T}} - \widehat{\mathbf{\Sigma}}_f)\mathbf{U}\mathbf{B}^{\mathrm{T}} + (\mathbf{B}\mathbf{U}^{\mathrm{T}} - \widehat{\mathbf{B}})\widehat{\mathbf{\Sigma}}_f\mathbf{U}\mathbf{B}^{\mathrm{T}} + \widehat{\mathbf{B}}\widehat{\mathbf{\Sigma}}_f(\mathbf{U}\mathbf{B}^{\mathrm{T}} - \widehat{\mathbf{B}}^{\mathrm{T}}).
\end{aligned}$$

Then we have

$$
\begin{aligned}
&\|\mathbf{B}\boldsymbol{\Sigma}_f\mathbf{B}^{\mathrm{T}}-\widehat{\mathbf{B}}\widehat{\boldsymbol{\Sigma}}_f\widehat{\mathbf{B}}^{\mathrm{T}}\|_{\mathcal{S},\max}\\
\leqslant&\|\mathbf{B}\mathbf{U}^{\mathrm{T}}\|_{\infty}\|\mathbf{U}\boldsymbol{\Sigma}_f\mathbf{U}^{\mathrm{T}}-\widehat{\boldsymbol{\Sigma}}_f\|_{\mathcal{S},\max}\|\mathbf{U}\mathbf{B}^{\mathrm{T}}\|_1\\
&+2\|\mathbf{B}\mathbf{U}^{\mathrm{T}}-\widehat{\mathbf{B}}\|_{\infty}(\|\mathbf{U}\boldsymbol{\Sigma}_f\mathbf{U}^{\mathrm{T}}\|_{\mathcal{S},\max}+\|\mathbf{U}\boldsymbol{\Sigma}_f\mathbf{U}^{\mathrm{T}}-\widehat{\boldsymbol{\Sigma}}_f\|_{\mathcal{S},\max})\|\mathbf{B}\mathbf{U}^{\mathrm{T}}\|_{\infty}\\
\leqslant&r^3C^2\|\mathbf{U}\boldsymbol{\Sigma}_f\mathbf{U}^{\mathrm{T}}-\widehat{\boldsymbol{\Sigma}}_f\|_{\mathcal{S},\max}+2r^{5/2}C(rC_{\max}+\|\mathbf{U}\boldsymbol{\Sigma}_f\mathbf{U}^{\mathrm{T}}-\widehat{\boldsymbol{\Sigma}}_f\|_{\mathcal{S},\max})\|\mathbf{B}\mathbf{U}^{\mathrm{T}}-\widehat{\mathbf{B}}\|_{\max}\\
=&O_p(\mathcal{M}_{\varepsilon}/\sqrt{n}+1/\sqrt{p})+O_p(\varpi_{n,p})=O_p(\varpi_{n,p}),
\end{aligned}
\tag{S.9}
$$

where the first inequality follows from Lemma B10(i), the second inequality follows from $\|\mathbf{B}\mathbf{U}^{\mathrm{T}}\|_{\infty}\leqslant r\|\mathbf{B}\mathbf{U}^{\mathrm{T}}\|_{\max}\leqslant r\|\mathbf{B}\|_{\max}\|\mathbf{U}\|_{\infty}\leqslant r^{3/2}C$ provided that $\|\mathbf{U}\|_{\infty}\leqslant\sqrt{r}\|\mathbf{U}\|=\sqrt{r}$, $\|\mathbf{B}\mathbf{U}^{\mathrm{T}}-\widehat{\mathbf{B}}\|_{\infty}\leqslant r\|\mathbf{B}\mathbf{U}^{\mathrm{T}}-\widehat{\mathbf{B}}\|_{\max}$ and $\|\mathbf{U}\boldsymbol{\Sigma}_f\mathbf{U}^{\mathrm{T}}\|_{\mathcal{S},\max}\leqslant C_{\max}\|\mathbf{U}\|_{\infty}^2\leqslant rC_{\max}$ in Lemma B18(i), and the last line follows from Lemma B27 and $\|\mathbf{B}\mathbf{U}^{\mathrm{T}}-\widehat{\mathbf{B}}\|_{\max}=O_p(\varpi_{n,p})$ in Theorem 1(i). Then note that

$$
\begin{aligned}
\|\widehat{\boldsymbol{\Sigma}}_{\varepsilon}^{\mathcal{A}}-\boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{S},\max}\leqslant&\|\widehat{\boldsymbol{\Sigma}}_{\varepsilon}^{\mathcal{A}}-\widehat{\boldsymbol{\Sigma}}_{\varepsilon}\|_{\mathcal{S},\max}+\|\widehat{\boldsymbol{\Sigma}}_{\varepsilon}-\boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{S},\max}\\
\leqslant&\max_{i,j\in[p]}(\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}\lambda)+O_p(\varpi_{n,p})=O_p(\varpi_{n,p}),
\end{aligned}
$$

where the last line follows from Lemma B25(iii), the choice of $\lambda=\dot{C}(\sqrt{\log p/n}+\sqrt{1/p})\lesssim\varpi_{n,p}$, and the fact $\max_{i,j\in[p]}\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}=O_p(1)$ by Lemma B26. By combining (5), (12) and (S.9), we obtain the desired result.

B.6 Proof of Theorem 5

For the sake of brevity, in this section, we suppose that the orthogonal matrix $\mathbf{U}$ in Theorem 1 and Lemmas B24–B27 is an identity matrix, which means, when we perform eigen-decomposition on $\widehat{\boldsymbol{\Omega}}$, we can always select the correct direction of $\widehat{\boldsymbol{\xi}}_j$ to ensure $\widehat{\boldsymbol{\xi}}_j^{\mathrm{T}}\widetilde{\mathbf{b}}_j\geqslant 0$. The proofs of Theorems 4 and 3 verify that the choice of $\mathbf{U}$ does not affect the theoretical results. To prove Theorem 5, we first present some technical lemmas with their proofs.

Lemma B28. *Under the assumptions of Theorem 5, then, $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}^{\mathcal{A}}$ has a bounded inverse with probability approaching 1, and $\|(\widehat{\boldsymbol{\Sigma}}_{\varepsilon}^{\mathcal{A}})^{-1}-\boldsymbol{\Sigma}_{\varepsilon}^{-1}\|_{\mathcal{L}}=O_p(\varpi_{n,p}^{1-q}s_p)$.*

Proof. Provided that $\varpi_{n,p}^{1-q}s_p = o(1)$ and $\lambda_{\min}(\boldsymbol{\Sigma}_\varepsilon) > c_1$ for some constant $c_1 > 0$, we combine Lemma B15 and Theorem 3 to yield that $\lambda_{\min}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}}) > c_1/2$ with probability approaching 1, and thus $\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}}$ has a bounded inverse with probability approaching 1 together with the desired result $\|(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1} - \boldsymbol{\Sigma}_\varepsilon^{-1}\|_{\mathcal{L}} = O_p(\varpi_{n,p}^{1-q}s_p)$. □

Lemma B29. *Under the assumptions of Theorem 5,*

$$\|\widehat{\mathbf{B}}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\widehat{\mathbf{B}} - \mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}\|_{\mathcal{L}} = O_p(p\varpi_{n,p}^{1-q}s_p) = o_p(p).$$

Proof. Consider

$$\begin{aligned}\|\widehat{\mathbf{B}}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\widehat{\mathbf{B}} - \mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}\|_{\mathcal{L}} \leqslant& 2\|(\widehat{\mathbf{B}} - \mathbf{B})^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\widehat{\mathbf{B}}\|_{\mathcal{L}} + \|\mathbf{B}^{\mathrm{T}}\{(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1} - \boldsymbol{\Sigma}_\varepsilon^{-1}\}\mathbf{B}\|_{\mathcal{L}}\\ \leqslant& 2\|\widehat{\mathbf{B}} - \mathbf{B}\|\{\lambda_{\min}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})\}^{-1}\|\mathbf{B}\| + \|\mathbf{B}\|^2\|(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1} - \boldsymbol{\Sigma}_\varepsilon^{-1}\|_{\mathcal{L}}\\ =& O_p(p\varpi_{n,p}) + O_p(p\varpi_{n,p}^{1-q}s_p) = O_p(p\varpi_{n,p}^{1-q}s_p) = o_p(p),\end{aligned}$$

where the last line follows from Lemmas B24 and B28. □

Lemma B30. *Under the assumptions of Theorem 5, then, with probability approaching 1,*

(i) $\lambda_{\min}(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}) \gtrsim p$;

(ii) $\lambda_{\min}(\widehat{\boldsymbol{\Sigma}}_f^{-1} + \widehat{\mathbf{B}}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\widehat{\mathbf{B}}) \gtrsim p$.

Proof. (i) Note that

$$\lambda_{\min}(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}) \geqslant \lambda_{\min}(\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}) \geqslant \lambda_{\min}(\boldsymbol{\Sigma}_\varepsilon^{-1})\lambda_{\min}(\mathbf{B}^{\mathrm{T}}\mathbf{B}) \gtrsim p,$$

where the first inequality follows from the fact $\boldsymbol{\Sigma}_f$ is a Mercer kernel.

(ii) Since $\lambda_{\min}(\boldsymbol{\Sigma}_f) > c_2$ and $\|\widehat{\boldsymbol{\Sigma}}_f - \boldsymbol{\Sigma}_f\|_{\mathcal{L}} = O_p(\varpi_{n,p}) = o_p(1)$, by using Lemma B15, we have $\|\widehat{\boldsymbol{\Sigma}}_f^{-1} - \boldsymbol{\Sigma}_f^{-1}\|_{\mathcal{L}} = O_p(\varpi_{n,p})$. Thus, by Lemma B29,

$$\left\|\left\{\widehat{\boldsymbol{\Sigma}}_f^{-1} + \widehat{\mathbf{B}}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\widehat{\mathbf{B}}\right\} - \left\{\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}\right\}\right\|_{\mathcal{L}} = o_p(p).$$

Combing with Lemma A3, we obtain that $\lambda_{\min}(\widehat{\boldsymbol{\Sigma}}_f^{-1} + \widehat{\mathbf{B}}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\widehat{\mathbf{B}}) \gtrsim p$. □

We are now ready to prove Theorem 5. Using the functional version of Sherman–Morrison–Woodbury identity, we have $\|(\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{D}})^{-1} - \boldsymbol{\Sigma}_y^{-1}\|_{\mathcal{L}} \leqslant \sum_{k=1}^4 L_k$, where

$$
\begin{aligned}
L_1 &= \left\|(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1} - \boldsymbol{\Sigma}_\varepsilon^{-1}\right\|_{\mathcal{L}}, \\
L_2 &= \left\|\left\{(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\widehat{\mathbf{B}} - \boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}\right\}\left\{\widehat{\boldsymbol{\Sigma}}_f^{-1} + \widehat{\mathbf{B}}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\widehat{\mathbf{B}}\right\}^{-1}\widehat{\mathbf{B}}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\right\|_{\mathcal{L}}, \\
L_3 &= \left\|\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}\left\{\widehat{\boldsymbol{\Sigma}}_f^{-1} + \widehat{\mathbf{B}}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\widehat{\mathbf{B}}\right\}^{-1}\left\{\widehat{\mathbf{B}}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1} - \mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\right\}\right\|_{\mathcal{L}}, \\
L_4 &= \left\|\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}\left[\left\{\widehat{\boldsymbol{\Sigma}}_f^{-1} + \widehat{\mathbf{B}}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\widehat{\mathbf{B}}\right\}^{-1} - \left\{\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}\right\}^{-1}\right]\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\right\|_{\mathcal{L}}.
\end{aligned}
$$

Clearly, $L_1 = O_p(\varpi_{n,p}^{1-q}s_p)$ by Lemma B28. Then, note that $\|(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\widehat{\mathbf{B}} - \boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}\|_{\mathcal{L}} \leqslant \|(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1} - \boldsymbol{\Sigma}_\varepsilon^{-1}\|_{\mathcal{L}}\|\widehat{\mathbf{B}}\| + \|\boldsymbol{\Sigma}_\varepsilon^{-1}\|_{\mathcal{L}}\|\widehat{\mathbf{B}} - \mathbf{B}\| = O_p(\sqrt{p}\varpi_{n,p}^{1-q}s_p)$. From Lemma B30, we obtain that $L_2 \asymp L_3 = O_p(\varpi_{n,p}^{1-q}s_p)$. Lastly, since $\lambda_{\min}(\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}) \gtrsim p$ and $\|\{\widehat{\boldsymbol{\Sigma}}_f^{-1} + \widehat{\mathbf{B}}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\widehat{\mathbf{B}}\} - \{\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}\}\|_{\mathcal{L}} = O_p(p\varpi_{n,p}^{1-q}s_p) = o_p(p)$, we apply Lemma B15 to obtain that

$$
\left\|\{\widehat{\boldsymbol{\Sigma}}_f^{-1} + \widehat{\mathbf{B}}^{\mathrm{T}}(\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}})^{-1}\widehat{\mathbf{B}}\}^{-1} - \{\boldsymbol{\Sigma}_f^{-1} + \mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{B}\}^{-1}\right\|_{\mathcal{L}} = O_p(p^{-2})O_p(p\varpi_{n,p}^{1-q}s_p) = O_p(p^{-1}\varpi_{n,p}^{1-q}s_p),
$$

which implies that $L_4 = O_p(\varpi_{n,p}^{1-q}s_p)$. Combining the above results, $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{D}}$ has a bounded inverse with probability approaching one, and

$$
\left\|(\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{D}})^{-1} - \boldsymbol{\Sigma}_y^{-1}\right\|_{\mathcal{L}} = O_p(\varpi_{n,p}^{1-q}s_p).
$$

B.7 Proof of Theorem 1'

To prove Theorem 1', we first present some technical lemmas with their proofs.

Lemma B31. *Under Assumption 4', it holds that*

$$
\max_{t\in[n]}\sum_{t'=1}^{n}\frac{|\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle|}{p} = O(\mathcal{M}_\varepsilon), \quad \text{and} \quad \max_{t',t\in[n]}\frac{|\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle|}{p} = O(\mathcal{M}_\varepsilon).
$$

Proof. From Assumption 4', the functional stability measure of $\{\boldsymbol{\varepsilon}_t(\cdot)\}_{t\in\mathbb{Z}}$ is bounded ($\mathcal{M}_\varepsilon < \infty$), and we would like to associate it with the equation of interest in this lemma. Since

$\{\boldsymbol{\varepsilon}_t(\cdot)\}_{t\in\mathbb{Z}}$ is stationary, we have, uniformly in n,

$$\begin{aligned}
\max_{t\in[n]}\sum_{t'=1}^{n}\frac{|\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle|}{p} &\leqslant \max_{t\in[n]}\frac{1}{p}\sum_{i=1}^{p}\sum_{t'=1}^{n}|\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'i},\boldsymbol{\varepsilon}_{ti}\rangle| \leqslant \max_{t\in[n]}\max_{i\in[p]}\sum_{t'=1}^{n}|\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'i},\boldsymbol{\varepsilon}_{ti}\rangle| \\
&\leqslant \max_{i\in[p]}\sum_{t'=-\infty}^{\infty}\left|\mathbb{E}\int\varepsilon_{1i}(u)\varepsilon_{t'i}(u)\mathrm{d}u\right| \\
&\leqslant \max_{i\in[p]}\sum_{t'=-\infty}^{\infty}\left\{\mathbb{E}\int\varepsilon_{1i}(u)\varepsilon_{t'i}(u)\mathrm{d}u\cdot\int\varepsilon_{1i}(v)\varepsilon_{t'i}(v)\mathrm{d}v\right\}^{1/2} \\
&\leqslant \max_{i\in[p]}\sum_{t'=-\infty}^{\infty}\mathbb{E}\int\int\varepsilon_{1i}(u)\varepsilon_{t'i}(v)\mathrm{d}u\mathrm{d}v \\
&= \max_{i\in[p]}\sum_{h\in\mathbb{Z}}\int\int\boldsymbol{\phi}_i(u)^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon}^{(h)}(u,v)\boldsymbol{\phi}_i(v)\mathrm{d}u\mathrm{d}v \\
&=2\pi\cdot\max_{i\in[p]}\langle\boldsymbol{\phi}_i,\mathbf{f}_{\varepsilon,\theta=0}(\boldsymbol{\phi}_i)\rangle \leqslant 2\pi\omega_0^{\varepsilon}\cdot\max_{i\in[p]}\frac{\langle\boldsymbol{\phi}_i,\mathbf{f}_{\varepsilon,\theta=0}(\boldsymbol{\phi}_i)\rangle}{\langle\boldsymbol{\phi}_i,\boldsymbol{\Sigma}_{\varepsilon}(\boldsymbol{\phi}_i)\rangle} \\
&\leqslant 2\pi\omega_0^{\varepsilon}\cdot\operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\boldsymbol{\phi}\in\mathbb{H}_{0,\varepsilon}^{p}}\frac{\langle\boldsymbol{\phi},\mathbf{f}_{\varepsilon,\theta}(\boldsymbol{\phi})\rangle}{\langle\boldsymbol{\phi},\boldsymbol{\Sigma}_{\varepsilon}(\boldsymbol{\phi})\rangle} = \omega_0^{\varepsilon}\mathcal{M}_{\varepsilon} = O(\mathcal{M}_{\varepsilon}),
\end{aligned}$$

where $\boldsymbol{\phi}_i(\cdot) = (0,\ldots,1,\ldots)^{\mathrm{T}}$ with its i-th element being 1 and the rest being 0, $\mathbb{H}_{0,\varepsilon}^{p} = \{\boldsymbol{\phi}\in\mathbb{H}^p : \langle\boldsymbol{\phi},\boldsymbol{\Sigma}_{\varepsilon}(\boldsymbol{\phi})\rangle\in(0,\infty)\}$, $\mathbf{f}_{\varepsilon,\theta}$ is the spectral density matrix function of $\{\boldsymbol{\varepsilon}_t(\cdot)\}_{t\in\mathbb{Z}}$ defined in Section 3.1, and $\omega_0^{\varepsilon} = \max_{j\in[p]}\int\Sigma_{\varepsilon,jj}(u,u)\mathrm{d}u$. Furthermore, we also obtain that

$$\max_{t',t\in[n]}\frac{|\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle|}{p} \leqslant \max_{t\in[n]}\sum_{t'=1}^{n}\frac{|\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle|}{p} = O(\mathcal{M}_{\varepsilon}).$$

□

Recall the definition of the asymptotically orthogonal matrix $\mathbf{H}$ introduced in Section 3.2. Applying the equation (C.2) in Fan et al. (2013) or (A.1) in Bai (2003), we have

$$\widehat{\boldsymbol{\gamma}}_t - \mathbf{H}\boldsymbol{\gamma}_t = \left(\frac{\mathbf{V}}{p}\right)^{-1}\left\{\frac{1}{n}\sum_{t'=1}^{n}\widehat{\boldsymbol{\gamma}}_{t'}\frac{\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle}{p} + \frac{1}{n}\sum_{t'=1}^{n}\widehat{\boldsymbol{\gamma}}_{t'}\zeta_{t't} + \frac{1}{n}\sum_{t'=1}^{n}\widehat{\boldsymbol{\gamma}}_{t'}\eta_{t't} + \frac{1}{n}\widehat{\boldsymbol{\gamma}}_{t'}\xi_{t't}\right\}, \tag{S.10}$$

where

$$\begin{aligned}
\zeta_{t't} &= \frac{1}{p}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle - \frac{1}{p}\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle, \\
\eta_{t't} &= \frac{1}{p}\boldsymbol{\gamma}_{t'}^{\mathrm{T}}\sum_{i=1}^{p}\int\mathbf{q}_i(u)\varepsilon_{ti}(u)\mathrm{d}u, \\
\xi_{t't} &= \frac{1}{p}\boldsymbol{\gamma}_{t}^{\mathrm{T}}\sum_{i=1}^{p}\int\mathbf{q}_i(u)\varepsilon_{t'i}(u)\mathrm{d}u.
\end{aligned}$$

Lemma B32. *Under the assumptions of Theorem 1', it holds that*

(i) $\max_{t\in[n]} \|(np)^{-1}\sum_{t'=1}^{n} \widehat{\boldsymbol{\gamma}}_{t'}\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_{t}\rangle\| = O_p(\mathcal{M}_\varepsilon/\sqrt{n})$*;*

(ii) $\max_{t\in[n]} \|n^{-1}\sum_{t'=1}^{n} \widehat{\boldsymbol{\gamma}}_{t'}\zeta_{t't}\| = O_p(\sqrt{n^{1/2}/p})$*;*

(iii) $\max_{t\in[n]} \|n^{-1}\sum_{t'=1}^{n} \widehat{\boldsymbol{\gamma}}_{t'}\eta_{t't}\| = O_p(\sqrt{n^{1/2}/p})$*;*

(iv) $\max_{t\in[n]} \|n^{-1}\sum_{t'=1}^{n} \widehat{\boldsymbol{\gamma}}_{t'}\xi_{t't}\| = O_p(\sqrt{n^{1/2}/p})$*.*

Proof. (i) By the Cauchy–Schwartz inequality and the fact that $n^{-1}\sum_{t=1}^{n}\|\widehat{\boldsymbol{\gamma}}_t\|^2 = O_p(1)$,

$$\begin{aligned}
\max_{t\in[n]}\left\|\frac{1}{np}\sum_{t'=1}^{n}\widehat{\boldsymbol{\gamma}}_{t'}\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_{t}\rangle\right\| \leqslant& \max_{t\in[n]}\left[\frac{1}{n}\sum_{t'=1}^{n}\|\widehat{\boldsymbol{\gamma}}_{t'}\|^2\frac{1}{n}\sum_{t'=1}^{n}\left\{\frac{\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_{t}\rangle}{p}\right\}^2\right]^{1/2}\\
\leqslant& O_p(1)\max_{t\in[n]}\left[\frac{1}{n}\sum_{t'=1}^{n}\left\{\frac{\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_{t}\rangle}{p}\right\}^2\right]^{1/2}\\
\leqslant& O_p(1)\max_{t',t\in[n]}\left|\frac{\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_{t}\rangle}{p}\right|^{1/2}\max_{t\in[n]}\left\{\frac{1}{n}\sum_{t'=1}^{n}\left|\frac{\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_{t}\rangle}{p}\right|\right\}^{1/2}\\
=& O_p(\mathcal{M}_\varepsilon/\sqrt{n}),
\end{aligned}$$

where the last equality follows from Lemma B31.

(ii) By the Cauchy–Schwartz inequality and the fact that $n^{-1}\sum_{t=1}^{n}\|\widehat{\boldsymbol{\gamma}}_t\|^2 = O_p(1)$,

$$\begin{aligned}
\max_{t\in[n]}\left\|\frac{1}{n}\sum_{t'=1}^{n}\widehat{\boldsymbol{\gamma}}_{t'}\zeta_{t't}\right\| \leqslant& \max_{t\in[n]}\frac{1}{n}\Big(\sum_{t'=1}^{n}\|\widehat{\boldsymbol{\gamma}}_{t'}\|^2\sum_{t'=1}^{n}\zeta_{t't}^2\Big)^{1/2} = O_p(1)\Big(\max_{t\in[n]}\frac{1}{n}\sum_{t'=1}^{n}\zeta_{t't}^2\Big)^{1/2}\\
=& O_p(1)\left\{\max_{t\in[n]}\frac{1}{n}\sum_{t'=1}^{n}\Big(\frac{1}{p}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_{t}\rangle - \frac{1}{p}\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_{t}\rangle\Big)^2\right\}^{1/2} = O_p(n^{1/4}/\sqrt{p}),
\end{aligned}$$

where the last equality follows from Assumption 3'(ii) that $\mathbb{E}(n^{-1}\sum_{t'=1}^{n}\zeta_{t't}^2)^2 \leqslant \max_{t',t\in[n]} E\zeta_{t't}^4 = O(1/p^2)$, and then using Chebyshev's inequality and Bonferroni's method that leads to $\max_{t\in[n]} n^{-1}\sum_{t'=1}^{n}\zeta_{t't}^2 = O_p(\sqrt{n}/p)$.

(iii) By the Cauchy–Schwartz inequality and the fact that $\|n^{-1}\sum_{t'=1}^{n}\widehat{\boldsymbol{\gamma}}_{t'}\boldsymbol{\gamma}_{t'}^{\mathrm{T}}\| = O_p(1)$,

$$\max_{t\in[n]}\left\|\frac{1}{n}\sum_{t'=1}^{n}\widehat{\boldsymbol{\gamma}}_{t'}\eta_{t't}\right\| \leqslant \left\|\frac{1}{n}\sum_{t'=1}^{n}\widehat{\boldsymbol{\gamma}}_{t'}\boldsymbol{\gamma}_{t'}^{\mathrm{T}}\right\|\max_{t\in[n]}\left\|\frac{1}{p}\sum_{i=1}^{p}\int\mathbf{q}_i(u)\varepsilon_{ti}(u)\mathrm{d}u\right\| = O_p(n^{1/4}/\sqrt{p}),$$

where the last equality follows from Assumption 3'(ii) that $\mathbb{E}\|p^{-1/2}\sum_{i=1}^{p}\int\mathbf{q}_i(u)\varepsilon_{ti}(u)\mathrm{d}u\|^4 = O(1)$, and then using Chebyshev's inequality and Bonferroni's method that leads to

$\max_{t\in[n]} \|p^{-1}\sum_{i=1}^p \int \mathbf{q}_i(u)\varepsilon_{ti}(u)\mathrm{d}u\| = O_p(n^{1/4}/\sqrt{p})$.

(iv) Similar to (iii), we can show that $\|(np)^{-1}\sum_{t'=1}^n\sum_{i=1}^p \int \mathbf{q}_i(u)\varepsilon_{t'i}(u)\mathrm{d}u\widehat{\boldsymbol{\gamma}}_{t'}\| = O_p(1/\sqrt{p})$. Additionally, $\max_{t\in[n]}\|\boldsymbol{\gamma}_t\| = O_p(n^{1/4})$, implied by $\mathbb{E}\|\boldsymbol{\gamma}_t\|^4 = O(1)$ and the use of Bonferroni's method. The desired result follows immediately that

$$\max_{t\in[n]}\Big\|\frac{1}{n}\sum_{t'=1}^n \widehat{\boldsymbol{\gamma}}_{t'}\xi_{t't}\Big\| \leqslant \max_{t\in[n]}\|\boldsymbol{\gamma}_t\|\Big\|\frac{1}{np}\sum_{t'=1}^n\sum_{i=1}^p\int \mathbf{q}_i(u)\varepsilon_{t'i}(u)\mathrm{d}u\widehat{\boldsymbol{\gamma}}_{t'}\Big\| = O_p(n^{1/4}/\sqrt{p}).$$

$\square$

Lemma B33. *Denote $\widehat{\boldsymbol{\gamma}}_t = (\widehat{\gamma}_{t1},\ldots,\widehat{\gamma}_{tr})^{\mathrm{T}}$. Under the assumptions of Theorem 1′, it holds that, for $i\in[r]$,*

(i) $n^{-1}\sum_{t=1}^n[(np)^{-1}\sum_{t'=1}^n \widehat{\gamma}_{t'i}\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle]^2 = O_p(\mathcal{M}_\varepsilon^2/n)$;

(ii) $n^{-1}\sum_{t=1}^n(n^{-1}\sum_{t'=1}^n \widehat{\gamma}_{t'i}\zeta_{t't})^2 = O_p(1/p)$;

(iii) $n^{-1}\sum_{t=1}^n(n^{-1}\sum_{t'=1}^n \widehat{\gamma}_{t'i}\eta_{t't})^2 = O_p(1/p)$;

(iv) $n^{-1}\sum_{t=1}^n(n^{-1}\sum_{t'=1}^n \widehat{\gamma}_{t'i}\xi_{t't})^2 = O_p(1/p)$.

Proof. (i) By the Cauchy–Schwartz inequality and the fact that $\sum_{t'=1}^n \widehat{\gamma}_{t'i}^2 = n$,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^n\left(\frac{1}{n}\sum_{t'=1}^n\widehat{\gamma}_{t'i}\frac{\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle}{p}\right)^2 \leqslant& \frac{1}{n}\sum_{t=1}^n\frac{1}{n}\Big(\sum_{t'=1}^n\widehat{\gamma}_{t'i}^2\Big)\frac{1}{n}\sum_{t'=1}^n\left(\frac{\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle}{p}\right)^2\\ =&\frac{1}{n}\sum_{t=1}^n\frac{1}{n}\sum_{t'=1}^n\left(\frac{\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle}{p}\right)^2 \leqslant \max_{t\in[n]}\frac{1}{n}\sum_{t'=1}^n\left(\frac{\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle}{p}\right)^2\\ \leqslant& \max_{t',t\in[n]}\left|\frac{\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle}{p}\right|\max_{t\in[n]}\frac{1}{n}\sum_{t'=1}^n\left|\frac{\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle}{p}\right| = O(\mathcal{M}_\varepsilon^2/n),\end{aligned}$$

where the last equality follows from Lemma B31.

(ii) By the Cauchy–Schwartz inequality and the fact that $\sum_{t'=1}^n \widehat{\gamma}_{t'i}^2 = n$,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^n\Big(\frac{1}{n}\sum_{t'=1}^n\widehat{\gamma}_{t'i}\zeta_{t't}\Big)^2 =& \frac{1}{n^3}\sum_{t',l\in[n]}\Big\{\widehat{\gamma}_{t'i}\widehat{\gamma}_{ti}\Big(\sum_{t=1}^n\zeta_{t't}\zeta_{lt}\Big)\Big\}\\ \leqslant&\frac{1}{n^3}\Big\{\sum_{t',l\in[n]}\widehat{\gamma}_{t'i}^2\widehat{\gamma}_{ti}^2\sum_{t',l\in[n]}\Big(\sum_{t=1}^n\zeta_{t't}\zeta_{lt}\Big)^2\Big\}^{1/2}\\ \leqslant&\frac{1}{n^3}\sum_{t'=1}^n\widehat{\gamma}_{t'i}^2\Big\{\sum_{t',l\in[n]}\Big(\sum_{t=1}^n\zeta_{t't}\zeta_{lt}\Big)^2\Big\}^{1/2} = \frac{1}{n^2}\Big\{\sum_{t',l\in[n]}\Big(\sum_{t=1}^n\zeta_{t't}\zeta_{lt}\Big)^2\Big\}^{1/2}.\end{aligned}$$

Notice that $\mathbb{E}\big\{\sum_{t',l\in[n]}(\sum_{t=1}^n \zeta_{t't}\zeta_{lt})^2\big\} = n^2\mathbb{E}(\sum_{t=1}^n \zeta_{t't}\zeta_{lt})^2 \leqslant n^4 \max_{t',t}\mathbb{E}|\zeta_{t't}|^4$, and by Assumption 3'(ii) we have $\max_{t',t}\mathbb{E}|\zeta_{t't}|^4 = O(1/p^2)$, which yields the desired result by using Chebyshev's inequality.

(iii) By the Cauchy–Schwartz inequality, and the facts that $\sum_{t'=1}^n \widehat{\gamma}_{t'i}^2 = n$ and $n^{-1}\sum_{t'=1}^n \|\boldsymbol{\gamma}_{t'}\|^2 = O_p(1)$,

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^n\Big(\frac{1}{n}\sum_{t'=1}^n \widehat{\gamma}_{t'i}\eta_{t't}\Big)^2 \leqslant & \Big\|\frac{1}{n}\sum_{t'=1}^n \widehat{\gamma}_{t'i}\boldsymbol{\gamma}_{t'}^{\mathrm{T}}\Big\|^2 \frac{1}{n}\sum_{t=1}^n\Big\|\frac{1}{p}\sum_{j=1}^p\int \mathbf{q}_j(u)\varepsilon_{tj}(u)\mathrm{d}u\Big\|^2\\
\leqslant & \Big(\frac{1}{n}\sum_{t'=1}^n \widehat{\gamma}_{t'i}^2\frac{1}{n}\sum_{t'=1}^n\|\boldsymbol{\gamma}_{t'}\|^2\Big)\frac{1}{np^2}\sum_{t=1}^n\Big\|\sum_{j=1}^p\int \mathbf{q}_j(u)\varepsilon_{tj}(u)\mathrm{d}u\Big\|^2\\
= & O_p(1)\cdot\frac{1}{np^2}\sum_{t=1}^n\Big\|\sum_{j=1}^p\int \mathbf{q}_j(u)\varepsilon_{tj}(u)\mathrm{d}u\Big\|^2.
\end{aligned}$$

Notice that we have $\mathbb{E}\|\sum_{j=1}^p\int \mathbf{q}_j(u)\varepsilon_{tj}(u)\mathrm{d}u\|^2 = O(p)$ by Assumption 3'(ii), which implies that $n^{-1}\mathbb{E}[\sum_{t=1}^n\|\sum_{j=1}^p\int \mathbf{q}_j(u)\varepsilon_{tj}(u)\mathrm{d}u\|^2] = O(p)$ and yields the result.

(iv) By the Cauchy–Schwartz inequality, and the facts that $\sum_{t'=1}^n \widehat{\gamma}_{t'i}^2 = n$ and $n^{-1}\sum_{t'=1}^n \|\boldsymbol{\gamma}_{t'}\|^2 = O_p(1)$,

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^n\Big(\frac{1}{n}\sum_{t'=1}^n \widehat{\gamma}_{t'i}\xi_{t't}\Big)^2 \leqslant & \Big(\frac{1}{n}\sum_{t=1}^n\|\boldsymbol{\gamma}_t\|^2\Big)\Big\|\frac{1}{np}\sum_{t'=1}^n\widehat{\gamma}_{t'i}\sum_{j=1}^p\int \mathbf{q}_j(u)\varepsilon_{t'j}(u)\mathrm{d}u\Big\|^2\\
\leqslant & O_p(1)\cdot\Big(\frac{1}{n}\sum_{t'=1}^n\widehat{\gamma}_{t'i}^2\Big)\frac{1}{np^2}\sum_{t'=1}^n\Big\|\sum_{j=1}^p\int \mathbf{q}_j(u)\varepsilon_{t'j}(u)\mathrm{d}u\Big\|^2 = O_p(1/p),
\end{aligned}$$

where the result in the last equality has been obtained in part (iii). □

Lemma B34. *Let $\{\hat{\tau}_j\}_{j=1}^r$ be the first r largest eigenvalues of $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(\cdot,\cdot)$ in a descending order. Under the assumptions of Theorem 1', it holds that $\hat{\tau}_r \gtrsim p$ with probability approaching 1.*

Proof. By Proposition 2, we obtain that

$$\tau_r \geqslant \|p\vartheta_r\|^2 - |\tau_r - p\vartheta_r| \gtrsim p - s_p \asymp p.$$

To show $\hat{\tau}_r \gtrsim p$ with probability approaching 1, it suffices to show that $|\hat{\tau}_r - \tau_r| = o_p(p)$. By applying Lemma A3 again, we only need to show $\|\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}} - \boldsymbol{\Sigma}_y\|_{\mathcal{S},\mathrm{F}} = o_p(p)$. Note that

$$\begin{aligned}
\|\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}} - \boldsymbol{\Sigma}_y\|_{\mathcal{S},\mathrm{F}} =& \Big\|\frac{1}{n}\sum_{t=1}^n (\mathbf{Q}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)(\mathbf{Q}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)^{\mathrm{T}} - \mathbf{Q}\mathbf{Q}^{\mathrm{T}} - \boldsymbol{\Sigma}_\varepsilon\Big\|_{\mathcal{S},\mathrm{F}} \\
\leqslant & \Big\|\mathbf{Q}\big(\frac{1}{n}\sum_{t=1}^n \boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}} - \mathbf{I}_r\big)\mathbf{Q}^{\mathrm{T}}\Big\|_{\mathcal{S},\mathrm{F}} + \Big\|\frac{1}{n}\sum_{t=1}^n \boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t^{\mathrm{T}} - \boldsymbol{\Sigma}_\varepsilon\Big\|_{\mathcal{S},\mathrm{F}} \\
&+ \Big\|\mathbf{Q}\big(n^{-1}\sum_{t=1}^n \boldsymbol{\gamma}_t\boldsymbol{\varepsilon}_t^{\mathrm{T}}\big)\Big\|_{\mathcal{S},\mathrm{F}} + \Big\|\big(\frac{1}{n}\sum_{t=1}^n \boldsymbol{\varepsilon}_t\boldsymbol{\gamma}_t^{\mathrm{T}}\big)\mathbf{Q}^{\mathrm{T}}\Big\|_{\mathcal{S},\mathrm{F}} \\
\leqslant & \Big\|n^{-1}\sum_{t=1}^n \boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}} - \mathbf{I}_r\Big\|_{\mathrm{F}} \cdot \|\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\|_{\mathcal{S},\mathrm{F}} + \Big(\sum_{i=1}^p\sum_{j=1}^p \Big\|n^{-1}\sum_{t=1}^n \boldsymbol{\varepsilon}_{ti}\boldsymbol{\varepsilon}_{tj} - \Sigma_{\varepsilon,ij}\Big\|_{\mathcal{S}}^2\Big)^{1/2} \\
&+ 2\Big(\sum_{i=1}^p\sum_{j=1}^r \Big\|n^{-1}\sum_{t=1}^n \varepsilon_{ti}\gamma_{tj}\Big\|^2\Big)^{1/2} \cdot \sqrt{p}\max_{i\in[p]}\|\mathbf{q}_i\| \\
=& O_p(p/\sqrt{n}) + O_p(p\mathcal{M}_\varepsilon\sqrt{1/n}) + O_p(p\mathcal{M}_\varepsilon\sqrt{1/n}) = o_p(p),
\end{aligned}$$

where the second inequality follows from Lemma B10(ii), the fact $\|\mathbf{K}\|_{\mathcal{S},\mathrm{F}} \leqslant p\|\mathbf{K}\|_{\mathcal{S},\max}$ any $\mathbf{K}(\cdot,\cdot) \in \mathbb{H}^p \otimes \mathbb{H}^p$, and the Cauchy–Schwartz inequality. The last line of the above equation follows from Lemma B16, $\mathcal{M}_\varepsilon^2 = o(n)$, and the fact

$$\|\mathbf{Q}\mathbf{Q}^{\mathrm{T}}\|_{\mathcal{S},\mathrm{F}} = \Big[\sum_{i=1}^p\sum_{j=1}^p \int \big\{\mathbf{q}_i(u)^{\mathrm{T}}\mathbf{q}_j(v)\big\}^2 \mathrm{d}u\mathrm{d}v\Big]^{1/2} \leqslant p\max_{i\in[p]}\|\mathbf{q}_i\|^2 \asymp p.$$

Therefore, we have obtained that $\hat{\tau}_r \gtrsim p$ with probability approaching 1. $\square$

Lemma B35. *Under the assumptions of Theorem 1′, it holds that*

(i) $\|\mathbf{H}\| = O_p(1)$;

(ii) $\mathbf{H}\mathbf{H}^{\mathrm{T}} = \mathbf{I}_r + O_p(\mathcal{M}_\varepsilon/\sqrt{n} + 1/\sqrt{p})$;

(iii) $\mathbf{H}^{\mathrm{T}}\mathbf{H} = \mathbf{I}_r + O_p(\mathcal{M}_\varepsilon/\sqrt{n} + 1/\sqrt{p})$.

Proof. (i) By Lemma B34, $\|\mathbf{V}^{-1}\| = \hat{\tau}_r^{-1} = O_p(p^{-1})$. Also, $\|\widehat{\boldsymbol{\Gamma}}\| = \lambda_{\max}^{1/2}(\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\widehat{\boldsymbol{\Gamma}}) = \lambda_{\max}^{1/2}(n\mathbf{I}_r) = \sqrt{n}$ from the normalization (S.4), and $\|\boldsymbol{\Gamma}\| = \lambda_{\max}^{1/2}(\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Gamma}) = \lambda_{\max}^{1/2}(\sum_{t=1}^n \boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}}) = O_p(\sqrt{n})$ by Lemma B16(i). In addition, $\|\int \mathbf{Q}(u)^{\mathrm{T}}\mathbf{Q}(u)\mathrm{d}u\| = O(p)$. By the definition of $\mathbf{H}$, i.e., $\mathbf{H} = n^{-1}\mathbf{V}^{-1}\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\boldsymbol{\Gamma}\int \mathbf{Q}(u)^{\mathrm{T}}\mathbf{Q}(u)\mathrm{d}u$, we have $\|\mathbf{H}\| = O_p(1)$, which is also satisfied for $\|\mathbf{H}\|_{\mathrm{F}}$ since $\mathbf{H} \in \mathbb{R}^{r\times r}$.

(ii) Notice that

$$\|\mathbf{H}\mathbf{H}^{\mathrm{T}}-\mathbf{I}_r\|_{\mathrm{F}} \leqslant \Big\|\mathbf{H}\mathbf{H}^{\mathrm{T}}-\frac{1}{n}\sum_{t=1}^{n}\mathbf{H}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\Big\|_{\mathrm{F}}+\Big\|\frac{1}{n}\sum_{t=1}^{n}\mathbf{H}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}-\mathbf{I}_r\Big\|_{\mathrm{F}}. \tag{S.11}$$

In (S.11), the first term can be bound by $\|\mathbf{H}\mathbf{H}^{\mathrm{T}}-n^{-1}\sum_{t=1}^{n}\mathbf{H}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\|_{\mathrm{F}} \leqslant \|\mathbf{H}\|_{\mathrm{F}}^2\|\mathbf{I}_r - n^{-1}\sum_{t=1}^{n}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}}\|_{\mathrm{F}} = O_p(1/\sqrt{n})$ by Lemma B16(i). The second term can be bounded by

$$\begin{aligned}
\Big\|\frac{1}{n}\sum_{t=1}^{n}\mathbf{H}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}-\mathbf{I}_r\Big\|_{\mathrm{F}} =& \Big\|\frac{1}{n}\sum_{t=1}^{n}\mathbf{H}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}-\frac{1}{n}\sum_{t=1}^{n}\widehat{\boldsymbol{\gamma}}_t\widehat{\boldsymbol{\gamma}}_t^{\mathrm{T}}\Big\|_{\mathrm{F}} \\
\leqslant& \Big\|\frac{1}{n}\sum_{t=1}^{n}(\mathbf{H}\boldsymbol{\gamma}_t-\widehat{\boldsymbol{\gamma}}_t)\boldsymbol{\gamma}_t^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\Big\|_{\mathrm{F}}+\Big\|\frac{1}{n}\sum_{t=1}^{n}\widehat{\boldsymbol{\gamma}}_t(\widehat{\boldsymbol{\gamma}}_t^{\mathrm{T}}-\boldsymbol{\gamma}_t^{\mathrm{T}}\mathbf{H}^{\mathrm{T}})\Big\|_{\mathrm{F}} \\
\leqslant& \Big(\frac{1}{n}\sum_{t=1}^{n}\|\mathbf{H}\boldsymbol{\gamma}_t-\widehat{\boldsymbol{\gamma}}_t\|^2\frac{1}{n}\sum_{t=1}^{n}\|\mathbf{H}\boldsymbol{\gamma}_t\|^2\Big)^{1/2} \\
&+\Big(\frac{1}{n}\sum_{t=1}^{n}\|\mathbf{H}\boldsymbol{\gamma}_t-\widehat{\boldsymbol{\gamma}}_t\|^2\frac{1}{n}\sum_{t=1}^{n}\|\widehat{\boldsymbol{\gamma}}_t\|^2\Big)^{1/2} \\
=& O_p(\mathcal{M}_\varepsilon/\sqrt{n}+1/\sqrt{p}),
\end{aligned}$$

where the third line follows from Cauchy–Schwartz inequality, and the last line follows from Theorem 1'(i) and the fact $n^{-1}\sum_{t=1}^{n}\|\widehat{\boldsymbol{\gamma}}_t\|^2=O_p(1)$.

(iii) From part (ii), we have $\mathbf{H}\mathbf{H}^{\mathrm{T}}=\mathbf{I}_r+O_p(\mathcal{M}_\varepsilon/\sqrt{n}+1/\sqrt{p})$ and $\|\mathbf{H}\|=O_p(1)$. Therefore

$$\mathbf{H}\mathbf{H}^{\mathrm{T}}\mathbf{H}=\mathbf{H}+O_p(\mathcal{M}_\varepsilon/\sqrt{n}+1/\sqrt{p}).$$

Also, $\|\mathbf{H}^{-1}\| \leqslant \|\mathbf{H}\|+o_p(1)\|\mathbf{H}^{-1}\|$, which implies that $\|\mathbf{H}^{-1}\|=O_p(1)$. Multiplying the LHS of the above by $\mathbf{H}^{-1}$ yields that $\mathbf{H}^{\mathrm{T}}\mathbf{H}=\mathbf{I}_r+O_p(\mathcal{M}_\varepsilon/\sqrt{n}+1/\sqrt{p})$. □

We are now ready to prove Theorem 1'.

(i) By Lemma B34, the diagonal elements of $\mathbf{V}/p=\mathrm{diag}(\hat{\tau}_1/p,\ldots,\hat{\tau}_r/p)$ are bounded away from 0. By the inequality $(a+b+c+d)^2 \leqslant 4(a^2+b^2+c^2+d^2)$, equation (S.10) and Lemma B33, we have

$$\begin{aligned}
\max_{i\in[r]}\frac{1}{n}\sum_{t=1}^{n}(\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}\boldsymbol{\gamma}_t)_i^2 \lesssim& \max_{i\in[r]}\frac{1}{n}\sum_{t=1}^{n}\Big(\frac{1}{n}\sum_{t'=1}^{n}\widehat{\gamma}_{t'i}\frac{\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle}{p}\Big)^2+\max_{i\in[r]}\frac{1}{n}\sum_{t=1}^{n}\Big(\frac{1}{n}\sum_{t'=1}^{n}\widehat{\gamma}_{t'i}\zeta_{t't}\Big)^2 \\
&+\max_{i\in[r]}\frac{1}{n}\sum_{t=1}^{n}\Big(\frac{1}{n}\sum_{t'=1}^{n}\widehat{\gamma}_{t'i}\eta_{t't}\Big)^2+\max_{i\in[r]}\frac{1}{n}\sum_{t=1}^{n}\Big(\frac{1}{n}\sum_{t'=1}^{n}\widehat{\gamma}_{t'i}\xi_{t't}\Big)^2 \\
=& O_p\left(\mathcal{M}_\varepsilon^2/n+1/p\right).
\end{aligned}$$

The desired result immediately follows that

$$\frac{1}{n}\sum_{t=1}^{n}\|\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}\boldsymbol{\gamma}_t\|^2\leqslant r\max_{i\in[r]}\frac{1}{n}\sum_{t=1}^{n}(\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}\boldsymbol{\gamma}_t)_i^2=O_p\left(\mathcal{M}_{\varepsilon}^2/n+1/p\right).$$

(ii) Note $\|(\mathbf{V}/p)^{-1}\|=O(1)$. Applying the inequality $(a+b+c+d)^2\leqslant 4(a^2+b^2+c^2+d^2)$, equation (S.10) and Lemma B32, we have

$$\begin{aligned}\max_{t\in[n]}\|\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}\boldsymbol{\gamma}_t\|\lesssim&\max_{t\in[n]}\Big\|\frac{1}{np}\sum_{t'=1}^{n}\widehat{\boldsymbol{\gamma}}_{t'}\mathbb{E}\langle\boldsymbol{\varepsilon}_{t'},\boldsymbol{\varepsilon}_t\rangle\Big\|+\max_{t\in[n]}\Big\|\frac{1}{n}\sum_{t'=1}^{n}\widehat{\boldsymbol{\gamma}}_{t'}\zeta_{t't}\Big\|\\&+\max_{t\in[n]}\Big\|\frac{1}{n}\sum_{t'=1}^{n}\widehat{\boldsymbol{\gamma}}_{t'}\eta_{t't}\Big\|+\max_{t\in[n]}\Big\|\frac{1}{n}\sum_{t'=1}^{n}\widehat{\boldsymbol{\gamma}}_{t'}\xi_{t't}\Big\|\\=&O_p\Big(\mathcal{M}_{\varepsilon}/\sqrt{n}+\sqrt{n^{1/2}/p}\Big).\end{aligned}$$

(iii) Using the facts that $\widehat{\mathbf{q}}_i(\cdot)=n^{-1}\sum_{t=1}^{n}y_{ti}(\cdot)\widehat{\boldsymbol{\gamma}}_t$ and $y_{ti}(\cdot)=\mathbf{q}_i(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}_t+\varepsilon_{ti}(\cdot)$, we have, for $i\in[p]$

$$\begin{aligned}\widehat{\mathbf{q}}_i(\cdot)-\mathbf{H}\mathbf{q}_i(\cdot)=&\frac{1}{n}\sum_{t=1}^{n}y_{ti}(\cdot)\widehat{\boldsymbol{\gamma}}_t-\frac{1}{n}\sum_{t=1}^{n}\mathbf{H}\boldsymbol{\gamma}_t\big\{y_{ti}(\cdot)-\mathbf{q}_i(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}_t+\varepsilon_{ti}(\cdot)\big\}-\mathbf{H}\mathbf{q}_i(\cdot)\\=&\frac{1}{n}\sum_{t=1}^{n}\mathbf{H}\boldsymbol{\gamma}_t\varepsilon_{ti}(u)+\frac{1}{n}\sum_{t=1}^{n}y_{ti}(\cdot)(\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}\boldsymbol{\gamma}_t)+\mathbf{H}\Big(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}}-\mathbf{I}_r\Big)\mathbf{q}_i(\cdot).\end{aligned}\tag{S.12}$$

The first term in (S.12) can be bounded by

$$\max_{i\in[p]}\Big\|\frac{1}{n}\sum_{t=1}^{n}\mathbf{H}\boldsymbol{\gamma}_t\varepsilon_{ti}\Big\|\leqslant\|\mathbf{H}\|\max_{i\in[p]}\Big\{\sum_{j=1}^{r}\Big\|\frac{1}{n}\sum_{t=1}^{n}\gamma_{tj}\varepsilon_{ti}\Big\|^2\Big\}^{1/2}=O_p\left(\mathcal{M}_{\varepsilon}\sqrt{\frac{\log p}{n}}\right),$$

where the inequality follows from Lemma B11(i), and the last equality follows from Lemmas B16(iii) and B35(i). For the second term, since $\Sigma_{y,ii}=\mathbf{q}_i^{\mathrm{T}}\Sigma_{\gamma,ii}\mathbf{q}_i+\Sigma_{\varepsilon,ii}$ with $\|\mathbf{q}_i\|=O(1)$ and $\|\boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{S},\max}\leqslant\|\boldsymbol{\Sigma}_{\varepsilon}\|_{\mathcal{L}}=O(1)$, we have $\|\Sigma_{y,ii}\|_{\mathcal{S}}=\mathbb{E}\|y_{ti}\|^2=O(1)$, and thus $n^{-1}\sum_{t=1}^{n}\|y_{ti}\|^2=O_p(1)$ by Chebyshev's inequality. Using the Cauchy–Schwartz inequality in the second term of (S.12), we obtain that

$$\max_{i\in[p]}\Big\|\frac{1}{n}\sum_{t=1}^{n}y_{ti}(\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}\boldsymbol{\gamma}_t)\Big\|\leqslant\max_{i\in[p]}\Big(\frac{1}{n}\sum_{t=1}^{n}\|y_{ti}\|^2\Big)^{1/2}\Big(\frac{1}{n}\sum_{t=1}^{n}\|\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}\boldsymbol{\gamma}_t\|^2\Big)^{1/2}=O_p\left(\frac{\mathcal{M}_{\varepsilon}}{\sqrt{n}}+\frac{1}{\sqrt{p}}\right).$$

In addition, $\|\mathbf{H}\|=O_p(1)$ from Lemma B35(i), $\|n^{-1}\sum_{t=1}^{n}\boldsymbol{\gamma}_t\boldsymbol{\gamma}_t^{\mathrm{T}}-\mathbf{I}_r\|=O_p(1/\sqrt{n})$ from Lemma B16(i) and $\max_{i\in[p]}\|\mathbf{q}_i\|=O(1)$ from Assumption 3'(i) yield that the third term is of order $O_p(1/\sqrt{n})$. Combining the above results, we obtain that

$$\max_{i\in[p]}\|\widehat{\mathbf{q}}_i-\mathbf{H}\mathbf{q}_i\|=O_p\left(\mathcal{M}_{\varepsilon}\sqrt{\frac{\log p}{n}}+\frac{1}{\sqrt{p}}\right)=O_p(\varpi_{n,p}).$$

B.8 Proof of Corollary 1'

By Theorem 1'(ii)(iii), Lemmas B35(ii) and B17, we have

$$\begin{aligned}
\max_{i\in[p],t\in[n]}\|\widehat{\mathbf{q}}_i^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_t-\mathbf{q}_i^{\mathrm{T}}\boldsymbol{\gamma}_t\| \leqslant& \max_{i\in[p]}\|\widehat{\mathbf{q}}_i-\mathbf{H}\mathbf{q}_i\|\cdot\max_{t\in[n]}\|\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}\boldsymbol{\gamma}_t\|\\
&+\max_{i\in[p]}\|\mathbf{H}\mathbf{q}_i\|\cdot\max_{t\in[n]}\|\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}\boldsymbol{\gamma}_t\|\\
&+\max_{i\in[p]}\|\widehat{\mathbf{q}}_i-\mathbf{H}\mathbf{q}_i\|\cdot\max_{t\in[n]}\|\mathbf{H}\boldsymbol{\gamma}_t\|\\
&+\max_{i\in[p]}\|\mathbf{q}_i\|\cdot\max_{t\in[n]}\|\boldsymbol{\gamma}_t\|\cdot\|\mathbf{H}^{\mathrm{T}}\mathbf{H}-\mathbf{I}_r\|\\
=&O_p\Big\{\varpi_{n,p}\cdot(\mathcal{M}_\varepsilon/n^{1/2}+n^{1/4}/p^{1/2})\Big\}+O_p\big(\mathcal{M}_\varepsilon/n^{1/2}+n^{1/4}/p^{1/2}\big)\\
&+O_p\big(\varpi_{n,p}\cdot\sqrt{\log n}\big)+O_p\Big\{\sqrt{\log n}\cdot(\mathcal{M}_\varepsilon/n^{1/2}+p^{1/2})\Big\}\\
=&O_p\Big\{(\log n)^{1/2}\mathcal{M}_\varepsilon\sqrt{\frac{\log p}{n}}+\frac{n^{1/4}}{\sqrt{p}}\Big\}.
\end{aligned}$$

B.9 Proof of Theorem 3'

To prove Theorem 3', we first present a technical lemma with its proof.

Lemma B36. *Under the assumptions of Theorem 3', it holds that*

(i) $\max_{i\in[p]} n^{-1}\sum_{t=1}^n\|\widehat{\varepsilon}_{ti}-\varepsilon_{ti}\|^2=O_p(\varpi_{n,p}^2)$;

(ii) $\max_{i,j\in[p]}\|n^{-1}\sum_{t=1}^n\widehat{\varepsilon}_{ti}\widehat{\varepsilon}_{tj}-n^{-1}\sum_{t=1}^n\varepsilon_{ti}\varepsilon_{tj}\|_{\mathcal{S}}=O_p(\varpi_{n,p})$;

(iii) $\|\widetilde{\boldsymbol{\Sigma}}_\varepsilon-\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\max}=O_p(\varpi_{n,p})$.

Proof. (i) Note that $\varepsilon_{ti}(\cdot)-\widehat{\varepsilon}_{ti}(\cdot)=\big\{y_{ti}(\cdot)-\mathbf{q}_i(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}_t\big\}-\big\{y_{ti}(\cdot)-\widehat{\mathbf{q}}_i(\cdot)^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_t\big\}=\widehat{\mathbf{q}}_i(\cdot)^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_t-\mathbf{q}_i(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}_t$, which can be decomposed as $\widehat{\mathbf{q}}_i(\cdot)^{\mathrm{T}}\widehat{\boldsymbol{\gamma}}_t-\mathbf{q}_i(\cdot)^{\mathrm{T}}\boldsymbol{\gamma}_t=\big\{\widehat{\mathbf{q}}_i(\cdot)^{\mathrm{T}}-\mathbf{q}_i(\cdot)^{\mathrm{T}}\mathbf{H}\big\}\widehat{\boldsymbol{\gamma}}_t+\mathbf{q}_i(\cdot)^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}(\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}\boldsymbol{\gamma}_t)+\mathbf{q}_i(\cdot)^{\mathrm{T}}(\mathbf{H}^{\mathrm{T}}\mathbf{H}-\mathbf{I}_r)\boldsymbol{\gamma}_t$. Applying the inequality $(a+b+c)^2\leqslant 3a^2+3b^2+3c^2$ and the

Cauchy–Schwartz inequality yields that

$$
\begin{aligned}
\max_{i\in[p]}\frac{1}{n}\sum_{t=1}^{n}\|\widehat{\varepsilon}_{ti}-\varepsilon_{ti}\|^2 \leqslant & 3\max_{i\in[p]}\|\widehat{\mathbf{q}}_i-\mathbf{H}\mathbf{q}_i\|^2\frac{1}{n}\sum_{t=1}^{n}\|\widehat{\boldsymbol{\gamma}}_t\|^2 \\
& + 3\max_{i\in[p]}\|\mathbf{q}_i\|^2\|\mathbf{H}\|^2\frac{1}{n}\sum_{t=1}^{n}\|\widehat{\boldsymbol{\gamma}}_t-\mathbf{H}\boldsymbol{\gamma}_t\|^2 \\
& + 3\max_{i\in[p]}\|\mathbf{q}_i\|^2\|\mathbf{H}^{\mathrm{T}}\mathbf{H}-\mathbf{I}_r\|^2\sum_{t=1}^{n}\|\boldsymbol{\gamma}_t\|^2 \\
= & O_p(\varpi_{n,p}^2)+O_p(\mathcal{M}_\varepsilon^2/n+1/p)=O_p(\varpi_{n,p}^2).
\end{aligned}
$$

(ii) Notice that $\max_{i\in[p]}\mathbb{E}\|\varepsilon_{ti}\|^2=\max_{i\in[p]}\mathbb{E}\int\varepsilon_{ti}(u)^2\mathrm{d}u=\max_{i\in[p]}\int\Sigma_{\varepsilon,ii}(u,u)\mathrm{d}u=O(1)$ from Assumption 3'(iv), thus we have $\max_{i\in[p]}n^{-1}\sum_{t=1}^{n}\|\varepsilon_{ti}\|^2=O_p(1)$. By the Cauchy–Schwartz inequality, we have

$$
\begin{aligned}
\max_{i,j\in[p]}\Big\|\frac{1}{n}\sum_{t=1}^{n}\widehat{\varepsilon}_{ti}\widehat{\varepsilon}_{tj}-\frac{1}{n}\sum_{t=1}^{n}\varepsilon_{ti}\varepsilon_{tj}\Big\|_{\mathcal{S}} = & \max_{i,j\in[p]}\Big\|\frac{1}{n}\sum_{t=1}^{n}(\widehat{\varepsilon}_{ti}-\varepsilon_{ti})\widehat{\varepsilon}_{tj}+\varepsilon_{ti}(\widehat{\varepsilon}_{tj}-\varepsilon_{tj})\Big\|_{\mathcal{S}} \\
\leqslant & \max_{i\in[p]}\frac{1}{n}\sum_{t=1}^{n}\|\widehat{\varepsilon}_{ti}-\varepsilon_{ti}\|^2 \\
& + 2\Big(\max_{i\in[p]}\frac{1}{n}\sum_{t=1}^{n}\|\varepsilon_{ti}\|^2\Big)^{1/2}\Big(\max_{j\in[p]}\frac{1}{n}\sum_{t=1}^{n}\|\widehat{\varepsilon}_{tj}-\varepsilon_{tj}\|^2\Big)^{1/2} \\
= & O_p(\varpi_{n,p}^2)+O_p(\varpi_{n,p})=O_p(\varpi_{n,p}).
\end{aligned}
$$

(iii) By part (ii) and Lemma B16(ii), the result follows immediately. □

We are now ready to prove Theorem 3'. By Corollary 1' and Lemma B36(i), we can follow nearly the same procedure as in the proof of Theorem 3 to show the similar argument that there exist some constants $C_1, C_2>0$ such that with probability approaching 1,

$$
C_1\leqslant\min_{i\in[p],j\in[p]}\|\widetilde{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}\leqslant\max_{i\in[p],j\in[p]}\|\widetilde{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}\leqslant C_2.
$$

Together with Lemmas B36(iii), we can show that for any $\epsilon>0$, there exist some positive constant N such that each of events

$$
\widetilde{\Upsilon}_1=\Big\{\max_{i\in[p],j\in[p]}\Big\|\widetilde{\Sigma}_{\varepsilon,ij}-\Sigma_{\varepsilon,ij}\Big\|_{\mathcal{S}}<N\varpi_{n,p}\Big\},\ \widetilde{\Upsilon}_2=\Big\{C_1\leqslant\Big\|\widetilde{\Theta}_{ij}^{1/2}\Big\|_{\mathcal{S}}\leqslant C_2,\ \text{all } i,j\in[p]\Big\}
$$

hold with probability at least $1-\epsilon$. Then for $\dot{C}>2NC_1^{-1}(\varpi_{n,p}/\omega_{n,p})$ and under the event $\widetilde{\Upsilon}_1\cap\widetilde{\Upsilon}_2$, we obtain that $\|\widetilde{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}}-\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},1}\lesssim\varpi_{n,p}^{1-q}s_p$ by using the same way as the proof of

Theorem 3. By Proposition 3, we know that $\widehat{\mathbf{R}}^{\mathcal{A}} = \widetilde{\mathbf{\Sigma}}_{\varepsilon}^{\mathcal{A}}$. Therefore, with probability at least $1-2\epsilon$, $\|\widehat{\mathbf{R}}^{\mathcal{A}} - \mathbf{\Sigma}_{\varepsilon}\|_{\mathcal{S},1} \lesssim \varpi_{n,p}^{1-q} s_p$. Considering that $\epsilon > 0$ can be arbitrarily small, we have $\|\widehat{\mathbf{R}}^{\mathcal{A}} - \mathbf{\Sigma}_{\varepsilon}\|_{\mathcal{L}} \leqslant \|\widehat{\mathbf{R}}^{\mathcal{A}} - \mathbf{\Sigma}_{\varepsilon}\|_{\mathcal{S},1} = O_p(\varpi_{n,p}^{1-q} s_p)$.

B.10 Proof of Theorem 4'

Under Assumption 1', we have $\mathbf{\Sigma}_y(u,v) = \mathbf{Q}(u)\mathbf{Q}(v)^{\mathrm{T}} + \mathbf{\Sigma}_{\varepsilon}(u,v)$. By the Cauchy–Schwartz inequality and Lemma B11, we have

$$
\begin{aligned}
\|\widehat{\mathbf{Q}}\widehat{\mathbf{Q}}^{\mathrm{T}} - \mathbf{Q}\mathbf{Q}^{\mathrm{T}}\|_{\mathcal{S},\max} &= \max_{i,j\in[p]} \|\widehat{\mathbf{q}}_i^{\mathrm{T}}\widehat{\mathbf{q}}_j - \mathbf{q}_i^{\mathrm{T}}\mathbf{q}_j\|_{\mathcal{S}} \\
&\leqslant \max_{i,j\in[p]} \{\|(\widehat{\mathbf{q}}_i - \mathbf{H}\mathbf{q}_i)^{\mathrm{T}}\widehat{\mathbf{q}}_j\|_{\mathcal{S}} + \|\mathbf{q}_i^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}(\widehat{\mathbf{q}}_j - \mathbf{H}\mathbf{q}_j)\|_{\mathcal{S}} + \|\mathbf{q}_i^{\mathrm{T}}(\mathbf{H}^{\mathrm{T}}\mathbf{H} - \mathbf{I}_r)\mathbf{q}_j\|_{\mathcal{S}}\} \\
&\leqslant \max_{i\in[p]} \|\widehat{\mathbf{q}}_i - \mathbf{H}\mathbf{q}_i\|^2 + 2\|\mathbf{H}\| \max_{i\in[p]} \|\mathbf{q}_j\|\|\widehat{\mathbf{q}}_i - \mathbf{H}\mathbf{q}_i\| \\
&\quad + \|\mathbf{H}^{\mathrm{T}}\mathbf{H} - \mathbf{I}_r\| \max_{i\in[p]} \|\mathbf{q}_i\|^2 \\
&= O_p(\varpi_{n,p}^2) + O_p(\varpi_{n,p}) + O_p(\mathcal{M}_{\varepsilon}/\sqrt{n} + 1/\sqrt{p}) = O_p(\varpi_{n,p}),
\end{aligned}
$$

where the last line follows from Theorem 1'(iii) and Lemma B35. Then by Lemma B36, we have $\|\widetilde{\mathbf{\Sigma}}_{\varepsilon} - \mathbf{\Sigma}_{\varepsilon}\|_{\mathcal{S},\max} = \max_{i,j\in[p]} \|\widetilde{\Sigma}_{\varepsilon,ij} - \Sigma_{\varepsilon,ij}\|_{\mathcal{S}} = O_p(\varpi_{n,p})$, and hence

$$
\begin{aligned}
\|\widetilde{\mathbf{\Sigma}}_{\varepsilon}^{\mathcal{A}} - \mathbf{\Sigma}_{\varepsilon}\|_{\mathcal{S},\max} &\leqslant \|\widetilde{\mathbf{\Sigma}}_{\varepsilon}^{\mathcal{A}} - \widetilde{\mathbf{\Sigma}}_{\varepsilon}\|_{\mathcal{S},\max} + \|\widetilde{\mathbf{\Sigma}}_{\varepsilon} - \mathbf{\Sigma}_{\varepsilon}\|_{\mathcal{S},\max} \\
&\leqslant \max_{i,j\in[p]} (\|\widetilde{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}\lambda) + O_p(\varpi_{n,p}) = O_p(\varpi_{n,p}),
\end{aligned}
$$

where $\widetilde{\Theta}_{ij}(u,v) \equiv n^{-1}\sum_{t=1}^{n} \left\{\widehat{\varepsilon}_{ti}(u)\widehat{\varepsilon}_{tj}(v) - \widetilde{\Sigma}_{\varepsilon,ij}(u,v)\right\}^2$, the last line follows from Lemma B36(iii), the choice of $\lambda = \dot{C}(\sqrt{\log p/n} + \sqrt{1/p}) \lesssim \varpi_{n,p}$, and the fact $\max_{i,j\in[p]} \|\widetilde{\Theta}_{ij}^{1/2}\|_{\mathcal{S}} = O_p(1)$ that can be proved following a similar argument compared to the proof of Lemma B26. The desired result follows immediately.

B.11 Proof of Theorem 5'

By Proposition 3, we can rewrite the FPOET estimator as $\widehat{\mathbf{\Sigma}}_y^{\mathcal{F}}(u,v) = \widehat{\mathbf{Q}}(u)\widehat{\mathbf{Q}}(v)^{\mathrm{T}} + \widehat{\mathbf{R}}^{\mathcal{A}}(u,v), (u,v) \in \mathcal{U}^2$. By Sherman–Morrison–Woodbury identity (Theorem 4.2.5 in Hsing and Eubank (2015)) to obtain the inverse FPOET $(\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{F}})^{-1} = (\widehat{\mathbf{R}}^{\mathcal{A}})^{-1} - (\widehat{\mathbf{R}}^{\mathcal{A}})^{-1}\widehat{\mathbf{Q}}\{\mathbf{I}_r + \widehat{\mathbf{Q}}^{\mathrm{T}}(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1}\widehat{\mathbf{Q}}\}^{-1}\widehat{\mathbf{Q}}^{\mathrm{T}}(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1}$. Note that under Assumption 1', $\boldsymbol{\Sigma}_\gamma = \mathbf{I}_r$ and $\widehat{\boldsymbol{\Sigma}}_\gamma = \mathbf{I}_r$. To prove Theorem 5', we first present some technical lemmas with their proofs.

Lemma B37. *Under the assumptions of Theorem 5', then, $\widehat{\mathbf{R}}^{\mathcal{A}}$ has a bounded inverse with probability approaching 1, and $\|(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1} - \boldsymbol{\Sigma}_\varepsilon^{-1}\|_{\mathcal{L}} = O_p(\varpi_{n,p}^{1-q}s_p)$.*

Proof. Provided that $\varpi_{n,p}^{1-q}s_p = o(1)$ and $\lambda_{\min}(\boldsymbol{\Sigma}_\varepsilon) > c_1$ for some constant $c_1 > 0$, we combine Lemma B15 and Theorem 3' to yield that $\lambda_{\min}(\widehat{\mathbf{R}}^{\mathcal{A}}) > c_1/2$ with probability approaching 1, and thus $\widehat{\mathbf{R}}^{\mathcal{A}}$ has a bounded inverse with probability approaching 1 together with the desired result $\|(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1} - \boldsymbol{\Sigma}_\varepsilon^{-1}\|_{\mathcal{L}} = O_p(\varpi_{n,p}^{1-q}s_p)$. □

Lemma B38. *Under the assumptions of Theorem 5',*

$$\|\widehat{\mathbf{Q}}^{\mathrm{T}}(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1}\widehat{\mathbf{Q}} - \mathbf{H}\mathbf{Q}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{Q}\mathbf{H}^{\mathrm{T}}\|_{\mathcal{L}} = O_p(p\varpi_{n,p}^{1-q}s_p) = o_p(p).$$

Proof. In model (2), $\mathbf{Q}(\cdot)$ can be viewed as a bounded linear operator from $\mathbb{R}^r$ to $\mathbb{H}^p$, and thus we can also regard it as a kernel matrix function satisfying $\mathbf{Q}(u,v) \equiv \mathbf{Q}(u), \forall u, v \in \mathcal{U}$. From this perspective, $\|\mathbf{Q}\|_{\mathcal{S},\mathrm{F}}^2 = \sum_{i=1}^p \|\mathbf{q}_i\|^2 = \int \mathbf{Q}(u)^{\mathrm{T}}\mathbf{Q}(u)\mathrm{d}u = p(\vartheta_1 + \cdots + \vartheta_r) \asymp p$ under Assumption 1'. By Theorem 1'(iii), $\|\widehat{\mathbf{Q}} - \mathbf{Q}\mathbf{H}^{\mathrm{T}}\|_{\mathcal{S},\mathrm{F}} = \big\{\sum_{i=1}^p \|\widehat{\mathbf{q}}_i - \mathbf{H}\mathbf{q}_i\|^2\big\}^{1/2} = O_p(\sqrt{p}\varpi_{n,p})$. Hence,

$$\begin{aligned}
\|\widehat{\mathbf{Q}}^{\mathrm{T}}(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1}\widehat{\mathbf{Q}} - \mathbf{H}\mathbf{Q}^{\mathrm{T}}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{Q}\mathbf{H}^{\mathrm{T}}\|_{\mathcal{L}} \leqslant & 2\|(\widehat{\mathbf{Q}} - \mathbf{Q}\mathbf{H}^{\mathrm{T}})^{\mathrm{T}}(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1}\widehat{\mathbf{Q}}\|_{\mathcal{L}} \\
& + \|\mathbf{H}\mathbf{Q}^{\mathrm{T}}\{(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1} - \boldsymbol{\Sigma}_\varepsilon^{-1}\}\mathbf{Q}\mathbf{H}^{\mathrm{T}}\|_{\mathcal{L}} \\
\leqslant & 2\|\widehat{\mathbf{Q}} - \mathbf{Q}\mathbf{H}^{\mathrm{T}}\|_{\mathcal{S},\mathrm{F}}\{\lambda_{\min}(\widehat{\mathbf{R}}^{\mathcal{A}})\}^{-1}\|\mathbf{Q}\|_{\mathcal{S},\mathrm{F}} \\
& + \|\mathbf{Q}\|_{\mathcal{S},\mathrm{F}}^2\|\mathbf{H}\|^2\|(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1} - \boldsymbol{\Sigma}_\varepsilon^{-1}\|_{\mathcal{L}} \\
= & O_p(p\varpi_{n,p}) + O_p(p\varpi_{n,p}^{1-q}s_p) = O_p(p\varpi_{n,p}^{1-q}s_p) = o_p(p),
\end{aligned}$$

where the second inequality follows from Lemma A7(i)(ii), and the last line follows from Lemmas B35 and B37. □

Lemma B39. *Under the assumptions of Theorem 5′, then, with probability approaching 1,*

(i) $\lambda_{\min}\{\mathbf{I}_r + \mathbf{H}\mathbf{Q}^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{Q}\mathbf{H}^{\mathrm{T}}\} \gtrsim p$;

(ii) $\lambda_{\min}\{\mathbf{I}_r + \widehat{\mathbf{Q}}^{\mathrm{T}}(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1}\widehat{\mathbf{Q}}\} \gtrsim p$;

(iii) $\lambda_{\min}(\mathbf{I}_r + \mathbf{Q}^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{Q}) \gtrsim p$;

(iv) $\lambda_{\min}\{(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1} + \mathbf{Q}^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{Q}\} \gtrsim p$.

Proof. (i) By Lemma B35, with probability approaching 1, $\lambda_{\min}(\mathbf{H}\mathbf{H}^{\mathrm{T}})$ is bounded away from 0. Hence,

$$\begin{aligned}\lambda_{\min}\{\mathbf{I}_r + \mathbf{H}\mathbf{Q}^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{Q}\mathbf{H}^{\mathrm{T}}\} \geqslant & \lambda_{\min}\{\mathbf{H}\mathbf{Q}^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{Q}\mathbf{H}^{\mathrm{T}}\} \\ & \geqslant \lambda_{\min}(\boldsymbol{\Sigma}_{\varepsilon}^{-1})\lambda_{\min}(\mathbf{Q}\mathbf{Q}^{\mathrm{T}})\lambda_{\min}(\mathbf{H}^{\mathrm{T}}\mathbf{H}) \gtrsim p,\end{aligned}$$

where $\lambda_{\min}(\mathbf{Q}\mathbf{Q}^{\mathrm{T}}) = p\vartheta_r$ by Assumption 1′.

(ii) The result follows from part (i) and Lemmas B15 and B38.

(iii) The result follows from a similar argument to that for part (i).

(iv) The result follows from part (iii) and Lemmas B15 and B35. □

We are now ready to prove Theorem 5′. Using the functional version of Sherman—Morrison—Woodbury identity, we have $\|(\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{F}})^{-1} - \boldsymbol{\Sigma}_y^{-1}\|_{\mathcal{L}} \leqslant \sum_{k=1}^{6} L_k$, where

$$\begin{aligned}
L_1 &= \left\|(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1} - \boldsymbol{\Sigma}_{\varepsilon}^{-1}\right\|_{\mathcal{L}},\\
L_2 &= \left\|\{(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1} - \boldsymbol{\Sigma}_{\varepsilon}^{-1}\}\widehat{\mathbf{Q}}\{\mathbf{I}_r + \widehat{\mathbf{Q}}^{\mathrm{T}}(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1}\widehat{\mathbf{Q}}\}^{-1}\widehat{\mathbf{Q}}^{\mathrm{T}}(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1}\right\|_{\mathcal{L}},\\
L_3 &= \left\|\{(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1} - \boldsymbol{\Sigma}_{\varepsilon}^{-1}\}\widehat{\mathbf{Q}}\{\mathbf{I}_r + \widehat{\mathbf{Q}}^{\mathrm{T}}(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1}\widehat{\mathbf{Q}}\}^{-1}\widehat{\mathbf{Q}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon}^{-1}\right\|_{\mathcal{L}},\\
L_4 &= \left\|\boldsymbol{\Sigma}_{\varepsilon}^{-1}(\widehat{\mathbf{Q}} - \mathbf{Q}\mathbf{H}^{\mathrm{T}})\{\mathbf{I}_r + \widehat{\mathbf{Q}}^{\mathrm{T}}(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1}\widehat{\mathbf{Q}}\}^{-1}\widehat{\mathbf{Q}}^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon}^{-1}\right\|_{\mathcal{L}},\\
L_5 &= \left\|\boldsymbol{\Sigma}_{\varepsilon}^{-1}(\widehat{\mathbf{Q}} - \mathbf{Q}\mathbf{H}^{\mathrm{T}})\{\mathbf{I}_r + \widehat{\mathbf{Q}}^{\mathrm{T}}(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1}\widehat{\mathbf{Q}}\}^{-1}\mathbf{H}\mathbf{Q}^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon}^{-1}\right\|_{\mathcal{L}},\\
L_6 &= \left\|\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{Q}\mathbf{H}^{\mathrm{T}}\Big[\{\mathbf{I}_r + \widehat{\mathbf{Q}}^{\mathrm{T}}(\widehat{\mathbf{R}}^{\mathcal{A}})^{-1}\widehat{\mathbf{Q}}\}^{-1} - (\mathbf{I}_r + \mathbf{H}\mathbf{Q}^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon}^{-1}\mathbf{Q}\mathbf{H}^{\mathrm{T}})^{-1}\Big]\mathbf{H}\mathbf{Q}^{\mathrm{T}}\boldsymbol{\Sigma}_{\varepsilon}^{-1}\right\|_{\mathcal{L}}.
\end{aligned}$$

Combining with Lemmas B37 and B39, the desired result follows from a similar argument to the proof of Theorem 5.

B.12 Proof of Theorem 2

(i) By Proposition 1 and Lemma B20, $|\lambda_j - p^2\theta_j| \leqslant \|\boldsymbol{\Omega}_{\mathcal{R}}\| \asymp p$ for $j \in [r]$ which implies that $\lambda_j \asymp p^2$ for $j \in [r]$, and $|\lambda_j| \lesssim p$ for $r+1 \leqslant j \leqslant p$. Applying Lemma A1 and Lemma B22, we have $|\hat{\lambda}_j - \lambda_j| \leqslant \|\widehat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\| = O_p(\mathcal{M}_\varepsilon p^2\sqrt{\log p/n}) = o_p(p^2)$, which indicates that $\hat{\lambda}_j \asymp p^2$ with probability approaching 1 for $j \in [r]$, and $\hat{\lambda}_j = o_p(p^2)$ for $r+1 \leqslant j \leqslant p$ which implies that $\hat{\lambda}_{j+1}/\hat{\lambda}_j = (\hat{\lambda}_{j+1}/p^2)/(\hat{\lambda}_j/p^2) \rightarrow_p 0/0 = 1$ for $r+1 \leqslant j \leqslant p$. Furthermore, $\lambda_{r+1} = O(p)$ and $\hat{\lambda}_{r+1} = O_p(\mathcal{M}_\varepsilon p^2\sqrt{\log p/n} + p)$. The desired result follows immediately since $\hat{\lambda}_{j+1}/\hat{\lambda}_j \asymp 1$ with probability approaching 1 if $j \neq r$, and $\hat{\lambda}_{r+1}/\hat{\lambda}_r = o_p(1)$.

(ii) By Proposition 2, $|\tau_j - p\vartheta_j| \leqslant \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}} = O(1)$ for $j \in [r]$, which implies that $\tau_j \asymp p$ for $j \in [r]$, and $|\tau_j| = O(1)$ for $j \geqslant r+1$. Using Lemma A3 and the proof of Lemma B34, we have $|\hat{\tau}_j - \tau_j| \leqslant \|\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}} - \boldsymbol{\Sigma}_y\|_{\mathcal{S},\mathrm{F}} = O_p(\mathcal{M}_\varepsilon p\sqrt{\log p/n}) = o_p(p)$, which indicates that $\hat{\tau}_j \asymp p$ with probability approaching 1 for $j \in [r]$, and $\hat{\tau}_j = o_p(p)$ for $j \geqslant r+1$. Furthermore, $\tau_{r+1} = O(1)$ and $\hat{\tau}_{r+1} = O_p(\mathcal{M}_\varepsilon p\sqrt{\log p/n} + 1)$. The desired result follows immediately since $\hat{\tau}_{j+1}/\hat{\tau}_j \asymp 1$ with probability approaching 1 if $j \neq r$, and $\hat{\tau}_{r+1}/\hat{\tau}_r = o_p(1)$.

C Proofs of theoretical results in Section 4

For the sake of brevity and readability, in this section, we suppose that the orthogonal matrix $\mathbf{U}$ in Theorem 1 and Lemmas B24–B27 is an identity matrix, which means that, when we perform eigen-decomposition on $\widehat{\boldsymbol{\Omega}}$, we can always select the correct direction of $\widehat{\boldsymbol{\xi}}_j$ to ensure $\widehat{\boldsymbol{\xi}}_j^{\mathrm{T}}\widetilde{\mathbf{b}}_j \geqslant 0$. The proofs in Section 3 verify that the choice of $\mathbf{U}$ does not affect the theoretical results.

C.1 Propositions S.1–S.2 and their proofs

The following two propositions are used in Section 4.1 to quantify the maximum absolute and relative approximation errors of the functional portfolio variance.

Proposition S.1. *Let* $\mathbf{\Sigma} = \{\Sigma_{ij}(\cdot,\cdot)\}_{p\times p}$, *and* $\widehat{\mathbf{\Sigma}} = \{\widehat{\Sigma}_{ij}(\cdot,\cdot)\}_{p\times p}$ *with each* $\Sigma_{ij}, \widehat{\Sigma}_{ij} \in \mathbb{S}$. *For any fixed* $\mathbf{w}(\cdot) \in \mathbb{H}^p$, *we have*

$$\Big|\langle \mathbf{w}, \widehat{\mathbf{\Sigma}}(\mathbf{w})\rangle - \langle \mathbf{w}, \mathbf{\Sigma}(\mathbf{w})\rangle\Big| \leqslant \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\mathcal{S},\max}\Big(\sum_{i\in[p]}\|w_i\|\Big)^2.$$

Proof. Consider that

$$\begin{aligned}\langle \mathbf{w}, \mathbf{\Sigma}(\mathbf{w})\rangle &= \int\int\sum_{i\in[p]}\sum_{j\in[p]} w_i(u)w_j(v)\Sigma_{ij}(u,v)\mathrm{d}u\mathrm{d}v\\ &\leqslant \sum_{i\in[p]}\sum_{j\in[p]}\left\{\int\int \Sigma_{ij}(u,v)^2\mathrm{d}u\mathrm{d}v\right\}^{1/2}\left\{\int\int w_i(u)^2 w_j(v)^2\mathrm{d}u\mathrm{d}v\right\}^{1/2}\\ &\leqslant \max_{i\in[p],j\in[p]}\|\Sigma_{ij}\|_{\mathcal{S}}\cdot\sum_{i\in[p]}\sum_{j\in[p]}\left\{\int w_i(u)^2\mathrm{d}u\right\}^{1/2}\left\{\int w_j(v)^2\mathrm{d}v\right\}^{1/2}\\ &=\|\mathbf{\Sigma}\|_{\mathcal{S},\max}\cdot\sum_{i\in[p]}\sum_{j\in[p]}\|w_i\|\|w_j\| = \|\mathbf{\Sigma}\|_{\mathcal{S},\max}\Big(\sum_{i\in[p]}\|w_i\|\Big)^2.\end{aligned}$$

Thus, $\Big|\langle \mathbf{w}, \widehat{\mathbf{\Sigma}}(\mathbf{w})\rangle - \langle \mathbf{w}, \mathbf{\Sigma}(\mathbf{w})\rangle\Big| = \Big|\langle \mathbf{w}, (\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})\mathbf{w})\rangle\Big| \leqslant \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|_{\mathcal{S},\max}(\sum_{i=1}^p\|w_i\|)^2$. □

Proposition S.2. *Suppose* $\mathbf{\Sigma} = \{\Sigma_{ij}(\cdot,\cdot)\}_{p\times p}$ *has a bounded inverse. For any fixed* $\mathbf{w}(\cdot) \in \mathbb{H}^p$, *we have*

$$\left|\frac{\langle \mathbf{w}, \widehat{\mathbf{\Sigma}}(\mathbf{w})\rangle}{\langle \mathbf{w}, \mathbf{\Sigma}(\mathbf{w})\rangle} - 1\right| \leqslant \|\mathbf{\Sigma}^{-1/2}\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1/2} - \mathbf{I}_p\|_{\mathcal{L}}.$$

Proof. For any given $\mathbf{w} \in \mathbb{H}^p$, we denote $\mathbf{x} = \mathbf{\Sigma}^{1/2}\mathbf{w}$ and $\mathbf{w} = \mathbf{\Sigma}^{-1/2}\mathbf{x}$, provided that $\mathbf{\Sigma}$ has a bounded inverse. Consider that

$$\begin{aligned}\langle \mathbf{w}, \mathbf{\Sigma}(\mathbf{w})\rangle &= \int\int \mathbf{w}(u)^{\mathrm{T}}\mathbf{\Sigma}(u,v)\mathbf{w}(v)\mathrm{d}u\mathrm{d}v = \int\int \mathbf{w}(u)^{\mathrm{T}}\left\{\int \mathbf{\Sigma}^{1/2}(u,w)\mathbf{\Sigma}^{1/2}(w,v)\mathrm{d}w\right\}\mathbf{w}(v)\mathrm{d}u\mathrm{d}v\\ &= \int\left\{\int \mathbf{w}(u)^{\mathrm{T}}\mathbf{\Sigma}^{1/2}(u,w)\mathrm{d}u\right\}\left\{\int \mathbf{\Sigma}^{1/2}(w,v)\mathbf{w}(v)\mathrm{d}v\right\}\mathrm{d}w = \int \mathbf{x}(w)^{\mathrm{T}}\mathbf{x}(w)\mathrm{d}w = \|\mathbf{x}\|^2\end{aligned}$$

The relative error can be bounded by

$$\begin{aligned}\left|\frac{\langle \mathbf{w}, \widehat{\mathbf{\Sigma}}(\mathbf{w})\rangle}{\langle \mathbf{w}, \mathbf{\Sigma}(\mathbf{w})\rangle} - 1\right| &= \left|\frac{\langle \mathbf{w}, \widehat{\mathbf{\Sigma}}(\mathbf{w})\rangle - \langle \mathbf{w}, \mathbf{\Sigma}(\mathbf{w})\rangle}{\langle \mathbf{w}, \mathbf{\Sigma}(\mathbf{w})\rangle}\right|\\ &= \frac{\Big|\langle \mathbf{x}, \mathbf{\Sigma}^{-1/2}(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})\mathbf{\Sigma}^{-1/2}(\mathbf{x})\rangle\Big|}{\|\mathbf{x}\|^2}\\ &\leqslant \|\mathbf{\Sigma}^{-1/2}\widehat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-1/2} - \mathbf{I}_p\|_{\mathcal{L}},\end{aligned}$$

where the last line follows from Lemma B14. □

C.2 Proof of Theorem 6

Since there exist constants c_1 and $c_2 > 0$ such that $\lambda_{\min}(\mathbf{\Sigma}_\varepsilon) > c_1, \lambda_{\min}(\mathbf{\Sigma}_f) > c_2$, we can obtain that all the eigenvalues of $\mathbf{\Sigma}_y$ are bounded away from zero, and thus for any $\mathbf{K} \in \mathbb{H}^p \otimes \mathbb{H}^p$, $\|\mathbf{K}\|^2_{\mathcal{S},\Sigma_y} = O(p^{-1})\|\mathbf{K}\|^2_{\mathcal{S},\mathrm{F}}$ by Lemma A8(iii).

To prove Theorem 6, we first present some technical lemmas with their proofs.

Lemma C40. *Under the assumptions of Theorem 6, we have* $\|\mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_y^{-1}\mathbf{B}\|_{\mathcal{L}} = O(1)$.

Proof. By Theorem 3.5.6 of Hsing and Eubank (2015), we obtain that

$$\mathbf{\Sigma}_y^{-1} = \mathbf{\Sigma}_\varepsilon^{-1} - \mathbf{\Sigma}_\varepsilon^{-1}\mathbf{B}(\mathbf{\Sigma}_f^{-1} + \mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_\varepsilon^{-1}.$$

Then it follows that

$$\begin{aligned}\mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_y^{-1}\mathbf{B} =& \mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{B} - \mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{B}(\mathbf{\Sigma}_f^{-1} + \mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{B})^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{B} \\ =& \mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{B}(\mathbf{\Sigma}_f^{-1} + \mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{B})^{-1}\mathbf{\Sigma}_f^{-1} \\ =& \mathbf{\Sigma}_f^{-1} - \mathbf{\Sigma}_f^{-1}(\mathbf{\Sigma}_f^{-1} + \mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{B})^{-1}\mathbf{\Sigma}_f^{-1},\end{aligned} \tag{S.13}$$

which also implies that $\mathbf{\Sigma}_f^{-1} \succeq \mathbf{\Sigma}_f^{-1}(\mathbf{\Sigma}_f^{-1} + \mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{B})^{-1}\mathbf{\Sigma}_f^{-1}$ since $\mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_y^{-1}\mathbf{B} \succeq 0$. Here, for two Mercer's kernels $\mathbf{K}, \mathbf{G} \in \mathbb{H}^r \otimes \mathbb{H}^r$, we denote $\mathbf{K} \succeq \mathbf{G}$ as the eigenvalues of $\mathbf{K} - \mathbf{G}$ are nonnegative, i.e., $\mathbf{K} - \mathbf{G}$ is still a Mercer's kernel. Similar to the monotonicity of matrix spectral norm, it can be shown that the operator norm is monotone, i.e., $\mathbf{K} \succeq \mathbf{G}$ implies $\|\mathbf{K}\|_{\mathcal{L}} \geqslant \|\mathbf{G}\|_{\mathcal{L}}$. Thus, from (S.13) we have

$$\|\mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_y^{-1}\mathbf{B}\|_{\mathcal{L}} \leqslant \|\mathbf{\Sigma}_f^{-1}\|_{\mathcal{L}} + \|\mathbf{\Sigma}_f^{-1}(\mathbf{\Sigma}_f^{-1} + \mathbf{B}^{\mathrm{T}}\mathbf{\Sigma}_\varepsilon^{-1}\mathbf{B})^{-1}\mathbf{\Sigma}_f^{-1}\|_{\mathcal{L}} \leqslant 2\|\mathbf{\Sigma}_f^{-1}\|_{\mathcal{L}} \leqslant 2c_2^{-1} = O(1),$$

where the second inequality follows from the monotonicity of the operator norm. $\square$

Lemma C41. *Under the assumptions of Theorem 6, it follows that*

(i) $\|\widehat{\mathbf{\Sigma}}_\varepsilon^{\mathcal{A}} - \mathbf{\Sigma}_\varepsilon\|^2_{\mathcal{S},\Sigma_y} = O_p(\varpi_{n,p}^{2-2q}s_p^2)$;

(ii) $\|(\widehat{\mathbf{B}} - \mathbf{B})\widehat{\mathbf{\Sigma}}_f(\widehat{\mathbf{B}} - \mathbf{B})^{\mathrm{T}}\|^2_{\mathcal{S},\Sigma_y} = O_p(\varpi_{n,p}^4 p)$;

(iii) $\|\mathbf{B}\widehat{\mathbf{\Sigma}}_f(\widehat{\mathbf{B}} - \mathbf{B})^{\mathrm{T}}\|^2_{\mathcal{S},\Sigma_y} = O_p(\varpi_{n,p}^2)$;

(iv) $\|\mathbf{B}(\widehat{\mathbf{\Sigma}}_f - \mathbf{\Sigma}_f)\mathbf{B}^{\mathrm{T}}\|^2_{\mathcal{S},\Sigma_y} = O_p(\varpi_{n,p}^2/p)$.

Proof. (i) Since all the eigenvalues of $\boldsymbol{\Sigma}_y$ are bounded away from zero, and by Theorem 3,

$$\|\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}} - \boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\Sigma_y}^2 \asymp p^{-1}\|\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}} - \boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\mathrm{F}}^2 \asymp \|\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}} - \boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}}^2 = O_p(\varpi_{n,p}^{2-2q}s_p^2).$$

(ii) By applying Lemmas A8(iii) and B24, we have

$$\|(\widehat{\mathbf{B}}-\mathbf{B})\widehat{\boldsymbol{\Sigma}}_f(\widehat{\mathbf{B}}-\mathbf{B})^{\mathrm{T}}\|_{\mathcal{S},\Sigma_y}^2 \leqslant p^{-1}\|\widehat{\mathbf{B}}-\mathbf{B}\|_{\mathrm{F}}^4\|\widehat{\boldsymbol{\Sigma}}_f\|_{\mathcal{L}}^2\|\boldsymbol{\Sigma}_y^{-1}\|_{\mathcal{L}}^2 = O_p(\mathcal{M}_\varepsilon^4 p/n^2 + 1/p^3) = O_p(\varpi_{n,p}^4 p).$$

(iii) Consider

$$\begin{aligned}
\|\mathbf{B}\widehat{\boldsymbol{\Sigma}}_f(\widehat{\mathbf{B}}-\mathbf{B})^{\mathrm{T}}\|_{\mathcal{S},\Sigma_y}^2 =& p^{-1}\mathrm{tr}\left\{\int\left[\widehat{\boldsymbol{\Sigma}}_f(\widehat{\mathbf{B}}-\mathbf{B})^{\mathrm{T}}\boldsymbol{\Sigma}_y^{-1}(\widehat{\mathbf{B}}-\mathbf{B})\widehat{\boldsymbol{\Sigma}}_f\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_y^{-1}\mathbf{B}\right](u,u)\mathrm{d}u\right\}\\
\leqslant& p^{-1}\|\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_y^{-1}\mathbf{B}\|_{\mathcal{L}}\|\boldsymbol{\Sigma}_y^{-1}\|_{\mathcal{L}}\|\widehat{\boldsymbol{\Sigma}}_f\|_{\mathcal{L}}^2\|\widehat{\mathbf{B}}-\mathbf{B}\|_{\mathrm{F}}^2\\
=& O_p(\mathcal{M}_\varepsilon^2/n + 1/p^2) = O_p(\varpi_{n,p}^2),
\end{aligned}$$

where the last line follows from Lemmas A8(ii), C40 and B24.

(iv) A similar argument shows that

$$\begin{aligned}
\|\mathbf{B}(\widehat{\boldsymbol{\Sigma}}_f-\boldsymbol{\Sigma}_f)\mathbf{B}^{\mathrm{T}}\|_{\mathcal{S},\Sigma_y}^2 =& p^{-1}\mathrm{tr}\left\{\int\left[(\widehat{\boldsymbol{\Sigma}}_f-\boldsymbol{\Sigma}_f)\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_y^{-1}\mathbf{B}(\widehat{\boldsymbol{\Sigma}}_f-\boldsymbol{\Sigma}_f)\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_y^{-1}\mathbf{B}\right](u,u)\mathrm{d}u\right\}\\
\leqslant& p^{-1}\|\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}_y^{-1}\mathbf{B}\|_{\mathcal{L}}^2\|\widehat{\boldsymbol{\Sigma}}_f-\boldsymbol{\Sigma}_f\|_{\mathcal{L}}\|\widehat{\boldsymbol{\Sigma}}_f-\boldsymbol{\Sigma}_f\|_{\mathcal{N}}\\
=& O_p(\mathcal{M}_\varepsilon^2/np + 1/p^2) = O_p(\varpi_{n,p}^2/p),
\end{aligned}$$

where the last line follows from Lemmas B27(ii), C40 and B24. □

We are now ready to prove Theorem 6. By Lemma C41,

$$\begin{aligned}
\|\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{D}} - \boldsymbol{\Sigma}_y\|_{\mathcal{S},\Sigma_y}^2 \leqslant& 2\|\widehat{\boldsymbol{\Sigma}}_\varepsilon^{\mathcal{A}} - \boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S},\Sigma_y}^2 + 2\|(\widehat{\mathbf{B}}-\mathbf{B})\widehat{\boldsymbol{\Sigma}}_f(\widehat{\mathbf{B}}-\mathbf{B})^{\mathrm{T}}\|_{\mathcal{S},\Sigma_y}^2\\
&+ 4\|\mathbf{B}\widehat{\boldsymbol{\Sigma}}_f(\widehat{\mathbf{B}}-\mathbf{B})^{\mathrm{T}}\|_{\mathcal{S},\Sigma_y}^2 + 2\|\mathbf{B}(\widehat{\boldsymbol{\Sigma}}_f-\boldsymbol{\Sigma}_f)\mathbf{B}^{\mathrm{T}}\|_{\mathcal{S},\Sigma_y}^2\\
=& O_p\left(\frac{\mathcal{M}_\varepsilon^4 p}{n^2} + \varpi_{n,p}^{2-2q}s_p^2\right),
\end{aligned}$$

which then implies that

$$\|\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{D}} - \boldsymbol{\Sigma}_y\|_{\mathcal{S},\Sigma_y} = O_p\left(\frac{\mathcal{M}_\varepsilon^2\sqrt{p}}{n} + \varpi_{n,p}^{1-q}s_p\right).$$

D Proposition S.3 and its proof

The following proposition supporting Section 5 gives the true covariance matrix functions for two DGPs and the functional sparsity condition.

Proposition S.3. *(i) For* $\mathbf{y}_t(\cdot)$ *generated from model* (1),

$$\boldsymbol{\Sigma}_y(u,v) = \mathbf{B}\Big\{\sum_{i=1}^{50} i^{-2}\phi_i(u)\phi_i(v)(\mathbf{I}_r - \mathbf{A}^2)^{-1}\Big\}\mathbf{B}^{\mathrm{T}} + \sum_{l=1}^{25} 2^{-l}\phi_l(u)\phi_l(v)\mathbf{C}_\zeta.$$

(ii) For $\mathbf{y}_t(\cdot)$ *generated from model* (2),

$$\boldsymbol{\Sigma}_y(u,v) = \mathbf{Q}(u)(\mathbf{I}_r - \mathbf{A}^2)^{-1}\mathbf{Q}(v)^{\mathrm{T}} + \sum_{l=1}^{25} 2^{-l}\phi_l(u)\phi_l(v)\mathbf{C}_\zeta.$$

(iii) The functional sparsity condition on $\boldsymbol{\Sigma}_\varepsilon$ *as specified in* (7) *satisfies* $s_p \lesssim p^{1-\alpha}$ *for* $\alpha \in [0,1]$ *and* $q = 0$.

Proof. (i) Note that $\mathbb{E}\{\mathbf{f}_t(\cdot)\} = \mathbf{0}$ and

$$\boldsymbol{\Sigma}_f(u,v) = \mathrm{Cov}\{\boldsymbol{\Xi}_t\boldsymbol{\phi}(u), \boldsymbol{\Xi}_t\boldsymbol{\phi}(v)\} = \sum_{i=1}^{50}\phi_i(u)\phi_i(v)\mathrm{Var}(\boldsymbol{\xi}_{ti}),$$

provided that $\mathrm{Cov}(\boldsymbol{\xi}_{ti}, \boldsymbol{\xi}_{ti'}) = \mathbf{0}_{r\times r}$ for any $i \neq i'$. Let $\mathbf{C}_i = \mathrm{Var}(\boldsymbol{\xi}_{ti})$ and $\mathbf{C}_u = i^{-2}\mathbf{I}_r$ be the covariance matrix of the innovation $\mathbf{u}_{ti}$. For weakly stationary VAR(1), it holds that

$$\begin{aligned}\mathbf{C}_i =& \mathbf{C}_u + \mathbf{A}\mathbf{C}_u\mathbf{A}^{\mathrm{T}} + \mathbf{A}^2\mathbf{C}_u(\mathbf{A}^{\mathrm{T}})^2 + \cdots = \sum_{s=0}^{\infty}\mathbf{A}^s\mathbf{C}_u(\mathbf{A}^{\mathrm{T}})^s \\ =& i^{-2}\sum_{s=0}^{\infty}(\mathbf{A}\mathbf{A}^{\mathrm{T}})^s = i^{-2}\sum_{s=0}^{\infty}\mathbf{A}^{2s} = i^{-2}(\mathbf{I}_r - \mathbf{A}^2)^{-1}.\end{aligned}$$

Similarly, $\boldsymbol{\Sigma}_\varepsilon(u,v) = \sum_{l=1}^{25} 2^{-l}\phi_l(u)\phi_l(v)\mathrm{Var}(\boldsymbol{\psi}_{tl}) = \sum_{l=1}^{25} 2^{-l}\phi_l(u)\phi_l(v)\mathbf{C}_\zeta$. Hence we have

$$\boldsymbol{\Sigma}_y(u,v) = \mathbf{B}\boldsymbol{\Sigma}_f(u,v)\mathbf{B}^{\mathrm{T}} + \boldsymbol{\Sigma}_\varepsilon(u,v) = \mathbf{B}\Big\{\sum_{i=1}^{50}\phi_i(u)\phi_i(v)\mathbf{C}_i\Big\}\mathbf{B}^{\mathrm{T}} + \sum_{l=1}^{25} 2^{-l}\phi_l(u)\phi_l(v)\mathbf{C}_\zeta.$$

(ii) The desired result follows immediately from the proof of part (i).

(iii) To see the functional sparsity condition on $\boldsymbol{\Sigma}_\varepsilon$, notice that

$$\sigma_i(u) = \boldsymbol{\Sigma}_{\varepsilon,ii}(u,u) = \sum_{l=1}^{25} 2^{-l}\phi_l(u)^2 D_i^2(1+\delta) \quad \text{and} \quad \|\sigma_i\|_{\mathcal{N}} = \int \sigma_i(u)\mathrm{d}u = cD_i^2,$$

where $c = (1 - 2^{-25})(1 + \delta)$ is a constant. Then, for $q = 0$ in (7), we have

$$
\begin{aligned}
s_p &= \max_{i \in [p]} \sum_{j=1}^{p} \|\sigma_i\|_{\mathcal{N}}^{(1-q)/2} \|\sigma_j\|_{\mathcal{N}}^{(1-q)/2} \|\Sigma_{\varepsilon,ij}\|_{\mathcal{S}}^{q} = \max_{i \in [p]} \sum_{j=1}^{p} c D_i D_j I\{\|\mathbf{\Sigma}_{\varepsilon,ij}\|_{\mathcal{S}} \neq 0\} \\
&\leqslant \big(c \max_{i \in [p]} D_i^2\big) \max_{i \in [p]} \sum_{j=1}^{p} I(\mathbf{C}_{0,ij} \neq 0) \lesssim \max_{i \in [p]} \sum_{j=1}^{p} I(\breve{\mathbf{C}}_{ij}^{\mathcal{T}} \neq 0) \lesssim p^{1-\alpha}.
\end{aligned}
$$

□

E Further derivations and definitions

This section contains further derivations and definitions supporting the main context of the paper.

E.1 Estimating FFM (1) from a least squares perspective

Similar to Section 2.2, we develop a least squares method to fit model (1) with functional factors. Let $\mathbf{Y}(\cdot) = \{\mathbf{y}_1(\cdot), \ldots, \mathbf{y}_n(\cdot)\} \in \mathbb{R}^{p \times n}$, and $\mathbf{F}(u)^{\mathrm{T}} = \{\mathbf{f}_1(\cdot), \ldots, \mathbf{f}_n(\cdot)\} \in \mathbb{R}^{r \times n}$. Consider solving the least-squares minimization problem

$$
\arg\min_{\mathbf{B}, \mathbf{F}(\cdot)} \int \|\mathbf{Y}(u) - \mathbf{B}\mathbf{F}(u)^{\mathrm{T}}\|_{\mathrm{F}}^2 \mathrm{d}u = \arg\min_{\mathbf{B}, \mathbf{F}(\cdot)} \sum_{t=1}^{n} \|\mathbf{y}_t - \mathbf{B}\mathbf{f}_t\|^2, \tag{S.14}
$$

subject to the normalization $p^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I}_r$. Following the similar procedure in Section 2.2, we obtain that, for each given $\mathbf{B}$, the constrained least squares estimator $\widetilde{\mathbf{F}}(\cdot) = p^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{Y}(\cdot)$. Plugging this into (S.14), objective function becomes $\int \mathrm{tr}[(\mathbf{I}_p - p^{-1}\mathbf{B}\mathbf{B}^{\mathrm{T}})\mathbf{Y}(u)\mathbf{Y}(u)^{\mathrm{T}}]\mathrm{d}u$, whose minimizer is equivalent to the maximizer of $\mathrm{tr}\{\mathbf{B}^{\mathrm{T}}[\int \mathbf{Y}(u)\mathbf{Y}(u)^{\mathrm{T}}\mathrm{d}u]\mathbf{B}\}$. Apparently, $\widehat{\mathbf{B}}/\sqrt{p}$ are the eigenvectors corresponding to the r largest eigenvalues of the $p \times p$ matrix $\int \mathbf{Y}(u)\mathbf{Y}(u)^{\mathrm{T}}\mathrm{d}u = n \int \widehat{\mathbf{\Sigma}}_y^{\mathcal{S}}(u,u)\mathrm{d}u$.

For the DIGIT method, the loading matrix $\mathbf{B}$ is estimated by the eigenanalysis of $\int\int \widehat{\mathbf{\Sigma}}_y^{\mathcal{S}}(u,v)\widehat{\mathbf{\Sigma}}_y^{\mathcal{S}}(u,v)^{\mathrm{T}}\mathrm{d}u\mathrm{d}v$, while the above shows that minimizing the least squares criterion (S.14) is equivalent to performing eigenanalysis of $\int \widehat{\mathbf{\Sigma}}_y^{\mathcal{S}}(u,u)\mathrm{d}u$. By comparison, the DIGIT method contains more covariance information by taking into account not only the diagonal entries $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,u)$ but also the off-diagonal entries $\widehat{\boldsymbol{\Sigma}}_y^{\mathcal{S}}(u,v)$ for $u \neq v$. Although such increased information may not alter the convergence rate of the proposed estimator, it will reduce the variance to improve the estimation efficiency.

E.2 Relationship between two FFMs

We rewrite model (1) as $\mathbf{y}_t(\cdot) = \mathbf{B}\mathbf{f}_t(\cdot) + \boldsymbol{\varepsilon}_t(\cdot) = \boldsymbol{\chi}_t(\cdot) + \boldsymbol{\varepsilon}_t(\cdot)$, and model (2) as $\mathbf{y}_t(\cdot) = \mathbf{Q}(\cdot)\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t(\cdot) = \boldsymbol{\kappa}_t(\cdot) + \boldsymbol{\varepsilon}_t(\cdot)$, where $\boldsymbol{\chi}_t(\cdot)$ and $\boldsymbol{\kappa}_t(\cdot)$ are the common components of the two FFMs. The covariance matrix function of $\boldsymbol{\chi}_t(\cdot)$ is

$$\boldsymbol{\Sigma}_\chi(u,v) = \mathbf{B}\boldsymbol{\Sigma}_f(u,v)\mathbf{B}^{\mathrm{T}} = \mathbf{B}\Big\{\sum_{i=1}^{\infty}\omega_i\boldsymbol{\phi}_i(u)\boldsymbol{\phi}_i(v)^{\mathrm{T}}\Big\}\mathbf{B}^{\mathrm{T}} = \sum_{i=1}^{\infty}p\omega_i\boldsymbol{\psi}_i(u)\boldsymbol{\psi}_i(v)^{\mathrm{T}},$$

where, by Mercer's theorem, $\boldsymbol{\Sigma}_f(u,v) = \sum_{i=1}^{\infty}\omega_i\boldsymbol{\phi}_i(u)\boldsymbol{\phi}_i(v)^{\mathrm{T}}$ and $\boldsymbol{\psi}_i(\cdot) = \mathbf{B}\boldsymbol{\phi}_i(\cdot)/\sqrt{p}$. Suppose that Assumption S.1 is satisfied with $\boldsymbol{\Sigma}_\gamma = \mathrm{diag}(\check{\vartheta}_1,\ldots,\check{\vartheta}_r)$. The covariance matrix function of $\boldsymbol{\kappa}_t(\cdot)$ is

$$\boldsymbol{\Sigma}_\kappa(u,v) = \mathbf{Q}(u)\boldsymbol{\Sigma}_\gamma\mathbf{Q}(v)^{\mathrm{T}} = \sum_{j=1}^{r}\check{\vartheta}_j\check{\mathbf{q}}_j(u)\check{\mathbf{q}}_j(v)^{\mathrm{T}} = \sum_{j=1}^{r}p\check{\vartheta}_j\boldsymbol{\nu}_j(u)\boldsymbol{\nu}_j(v)^{\mathrm{T}},$$

where $\{\check{\mathbf{q}}_j(\cdot)\}_{j=1}^r$ is the set of columns of $\mathbf{Q}(\cdot)$ such that $\{\|\check{\mathbf{q}}_j\|\}_{j=1}^r$ is in a descending order, and $\boldsymbol{\nu}(\cdot) = \check{\mathbf{q}}_j(\cdot)/\sqrt{p}$. Note that

$$\int\boldsymbol{\psi}_i(u)^{\mathrm{T}}\boldsymbol{\psi}_j(u)\mathrm{d}u = \int\boldsymbol{\phi}_i(u)^{\mathrm{T}}p^{-1}\mathbf{B}^{\mathrm{T}}\mathbf{B}\boldsymbol{\phi}_j(u)\mathrm{d}u = \int\boldsymbol{\phi}_i(u)^{\mathrm{T}}\boldsymbol{\phi}_j(u)\mathrm{d}u = I(i=j),\ \text{and}$$
$$\int\boldsymbol{\nu}_i(u)^{\mathrm{T}}\boldsymbol{\nu}_j(u)\mathrm{d}u = p^{-1}\int\check{\mathbf{q}}_i(u)^{\mathrm{T}}\check{\mathbf{q}}_j(u)\mathrm{d}u = I(i=j)\ \text{from Assumption S.1.}$$

Consequently, $\{\boldsymbol{\psi}_i(\cdot)\}_{i=1}^{\infty}$ are the eigenfunctions of $\boldsymbol{\Sigma}_\chi$ with nonnegative eigenvalues $\{p\omega_i\}_{i=1}^{\infty}$, and $\{\boldsymbol{\nu}_j(\cdot)\}_{j=1}^r$, which can be extended to a set of orthonormal basis functions, are the eigenfunctions of $\boldsymbol{\Sigma}_\kappa$ with nonnegative eigenvalues $\{p\check{\vartheta}_j\}_{j=1}^{\infty}$ satisfying $\check{\vartheta}_j = 0$ when $j > r$.

In this point of view, FFM (1) can be converted to FFM (2) if and only if $\omega_i = 0$ when $i > r$. On the contrary, model (2) can be regarded as a special case of model (1) if and only if the solutions of $\{\boldsymbol{\phi}_j(\cdot)\}_{j=1}^r$ to the functional equations $\mathbf{B}\boldsymbol{\phi}_j(\cdot) = \check{\mathbf{q}}_j(\cdot)$ for $j \in [r]$ exist given $\mathbf{B}$ and $\{\check{\mathbf{q}}_j(\cdot)\}_{j=1}^r$. Since the rank of the space spanned by columns of matrix $\mathbf{B}$ is r, the equivalent condition for the existence of the solutions follows that the rank of the space spanned by $\{\check{\mathbf{q}}_j(\cdot)\}_{j=1}^r$ is r.

E.3 Sub-Gaussian (functional) linear process

We first define sub-Gaussian functional process.

Definition S.1. *Let $x_t(\cdot)$ be a mean zero random variable in $\mathbb{H}$ and $\Sigma_0 : \mathbb{H} \to \mathbb{H}$ be a covariance operator. Then $x_t(\cdot)$ is a sub-Gaussian process if there exists a constant $\alpha \geqslant 0$ such that for all $x \in \mathbb{H}$, $\mathbb{E}(\exp\langle x, x_t\rangle) \leqslant \exp\{\alpha^2\langle x, \Sigma_0(x)\rangle/2\}$.*

To develop finite-sample theory for relevant estimators in Section 3, we focus on multivariate functional linear process with sub-Gaussian errors, namely sub-Gaussian functional linear process. Specifically, we assume $\mathbf{z}_t(\cdot) = \{z_{t1}(\cdot), \ldots, z_{tp}(\cdot)\}^{\mathrm{T}} \in \mathbb{H}^p$ admits the representation

$$\mathbf{z}_t(\cdot) = \sum_{l=0}^{\infty} \mathbf{A}_l(\mathbf{x}_{t-l}),\ t \in \mathbb{Z}, \tag{S.15}$$

where $\mathbf{A}_l = (A_{l,ij})_{p\times p}$ with each $A_{l,ij} \in \mathbb{S}$ and $\mathbf{x}_t(\cdot) = \{x_{t1}(\cdot), \ldots, x_{tp}(\cdot)\}^{\mathrm{T}} \in \mathbb{H}^p$, whose components are independent sub-Gaussian processes satisfying Definition S.1, and the coefficient functions satisfy $\sum_{l=0}^{\infty} \|\mathbf{A}_l\|_{\mathcal{S},\infty} = O(1)$. In Section 3, we assume that $\mathbf{f}_t(\cdot)$ in model (1) and $\boldsymbol{\varepsilon}_t(\cdot)$ follow sub-Gaussian functional linear processes, and $\boldsymbol{\gamma}_t$ in model (2) follows sub-Gaussian linear process, which can be correspondingly defined from the non-functional versions of (S.15) and Definition S.1.

E.4 Optimal functional portfolio allocation

In this section, we derive the optimal functional portfolio allocation $\widehat{\mathbf{w}}(\cdot)$ that is required in Section 6.2. Specifically, we aim to solve the following constrained minimization problem:

$$\widehat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathbb{H}^p} \left\langle \mathbf{w}, \widehat{\boldsymbol{\Sigma}}_y(\mathbf{w})\right\rangle \ \text{ subject to }\ \mathbf{w}(u)^{\mathrm{T}}\mathbf{1}_p = 1 \text{ for any } u \in \mathcal{U},$$

where $\mathbf{1}_p = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^p$. To solve this, we apply the method of Lagrange multipliers by defining the Lagrangian function as

$$L(\mathbf{w}, \lambda) = \int\int \mathbf{w}(u)^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}}_y(u, v)\mathbf{w}(v)\mathrm{d}u\mathrm{d}v - \int \lambda(u)\{\mathbf{w}(u)^{\mathrm{T}}\mathbf{1}_p - 1\}\mathrm{d}u,$$

where $\lambda(\cdot) \in \mathbb{H}$. Setting the functional derivative of $L(\mathbf{w}, \lambda)$ with respect to $\mathbf{w}(\cdot)$ to zero, i.e., $\int \widehat{\boldsymbol{\Sigma}}_y(\cdot, v)\mathbf{w}(v)\mathrm{d}v - \lambda(\cdot)\mathbf{1}_p = 0$, we obtain that

$$\widehat{\mathbf{w}}(u) = \int \widehat{\boldsymbol{\Sigma}}_y^{-1}(u, v)\lambda(v)\mathbf{1}_p\mathrm{d}v, u \in \mathcal{U}.$$

Let $H(\cdot, \cdot) = \mathbf{1}_p^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_y^{-1}(\cdot, \cdot)\mathbf{1}_p$. With the constraint $\mathbf{1}_p^{\mathrm{T}}\widehat{\mathbf{w}}(u) = 1$ for any $u \in \mathcal{U}$, we have

$$\mathbf{1}_p^{\mathrm{T}}\widehat{\mathbf{w}}(u) = \int \lambda(v)\mathbf{1}_p^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_y^{-1}(u, v)\mathbf{1}_p\mathrm{d}v = \int \lambda(v)H(u, v)\mathrm{d}v = 1,$$

which indicates that $\lambda(u) = \int H^{-1}(u, v)\mathrm{d}v$ for any $u \in \mathcal{U}$. Combining the above results yield the desired solution

$$\widehat{\mathbf{w}}(u) = \int\int \widehat{\boldsymbol{\Sigma}}_y^{-1}(u, v)\mathrm{diag}\{H^{-1}(v, z), \cdots, H^{-1}(v, z)\}\mathbf{1}_p\mathrm{d}v\mathrm{d}z.$$

F Additional simulation results

This section provides additional results supporting Section 5. Figure S.1 presents boxplots of $\Delta\mathrm{PC}_i$ and $\Delta\mathrm{IC}_i$ ($i \in [3]$) for two DGPs under the setting $p = 200, n = 50, \alpha = 0.5$, and $r = 3, 5, 7$.

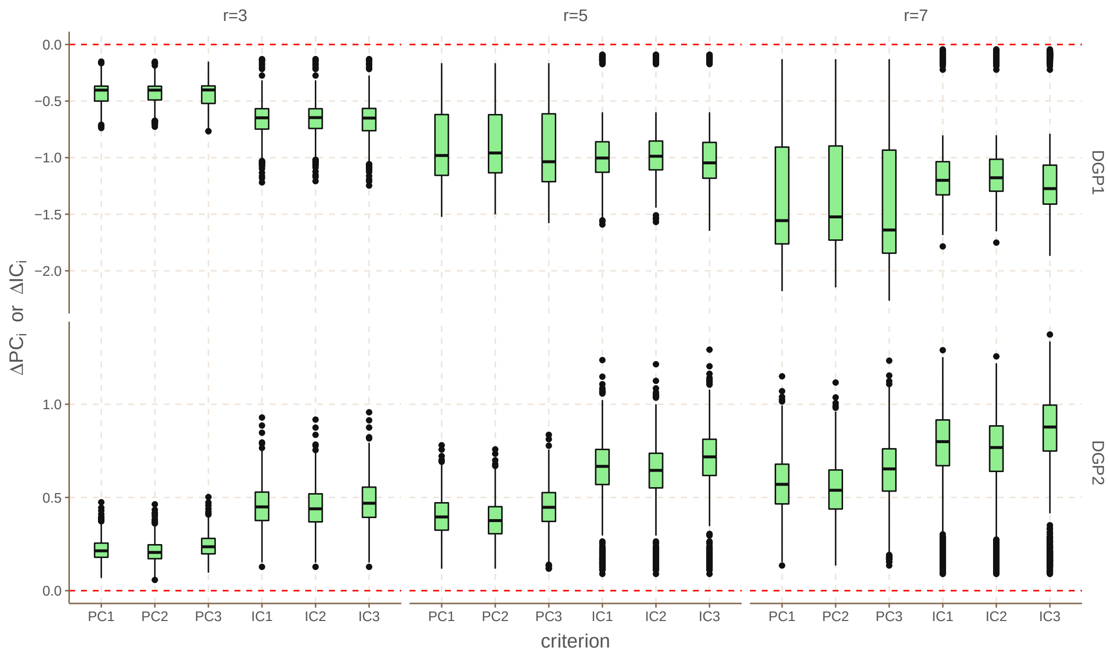

Figure S.1: The boxplots of $\Delta\mathrm{PC}_i$ and $\Delta\mathrm{IC}_i$ ($i \in [3]$) for DGP1 and DGP2 with $p = 200, n = 50, \alpha = 0.5$, and $r = 3, 5, 7$ over 1000 simulation runs.

We compare our AFT estimator in (10) with two related methods for estimating the idiosyncratic covariance $\boldsymbol{\Sigma}_{\varepsilon}$, specifically, the sample covariance estimator defined as $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}^{\mathcal{S}}(u, v) = n^{-1}\sum_{t=1}^{n} \widehat{\boldsymbol{\varepsilon}}_t(u)\widehat{\boldsymbol{\varepsilon}}_t(v)^{\mathrm{T}}$, and Fang et al. (2023)'s AFT estimator in (11). Figures S.2 and S.3 plot average losses of $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}$ measured by functional matrix ℓ_1 norm and operator norm for DGP1 and DGP2, respectively, under the settings $n = p = 60, 80, \ldots, 200$ and $\alpha = 0.25, 0.5, 0.75$. We observe several evident patterns. First, the estimation accuracy measured by both functional matrix norms substantially improves when using the AFT estimators compared to $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}^{\mathcal{S}}$. Second, despite our AFT proposal requiring weaker assumptions compared to Fang et al. (2023)'s method, both AFT estimators exhibit very similar empirical performance. Third, for $\alpha = 0.25$ and 0.5, the performance of the sample and AFT estimators deteriorates as p increases. However, when $\alpha = 0.75$, both losses of two AFT estimators decrease as p increases. This phenomenon can be attributed to the fact that $\{(\log p/n)^{1/2} + p^{-1/2}\}p^{1-\alpha} = o(1)$ as $n, p \to \infty$ if $\alpha > 0.5$, which is implied by Theorems 3 and 3' under the setting $n = p, q =$

$0, \mathcal{M}_\varepsilon = O(1)$. These simulation results nicely validate Theorems 3 and 3'.

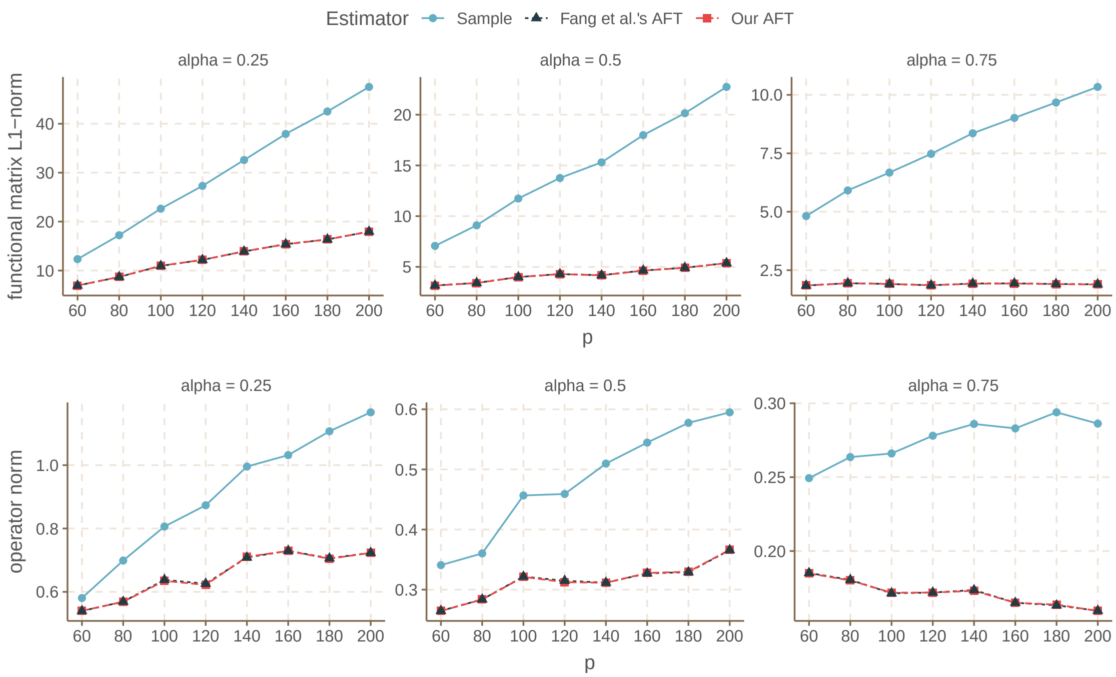

Figure S.2: The average losses of $\widehat{\mathbf{\Sigma}}_\varepsilon$ in functional matrix ℓ_1 norm (top row) and operator norm (bottom row) for DGP1 over 1000 simulation runs with $n = p = 60, 80, \ldots, 200$ and $\alpha = 0.25, 0.5, 0.75$.

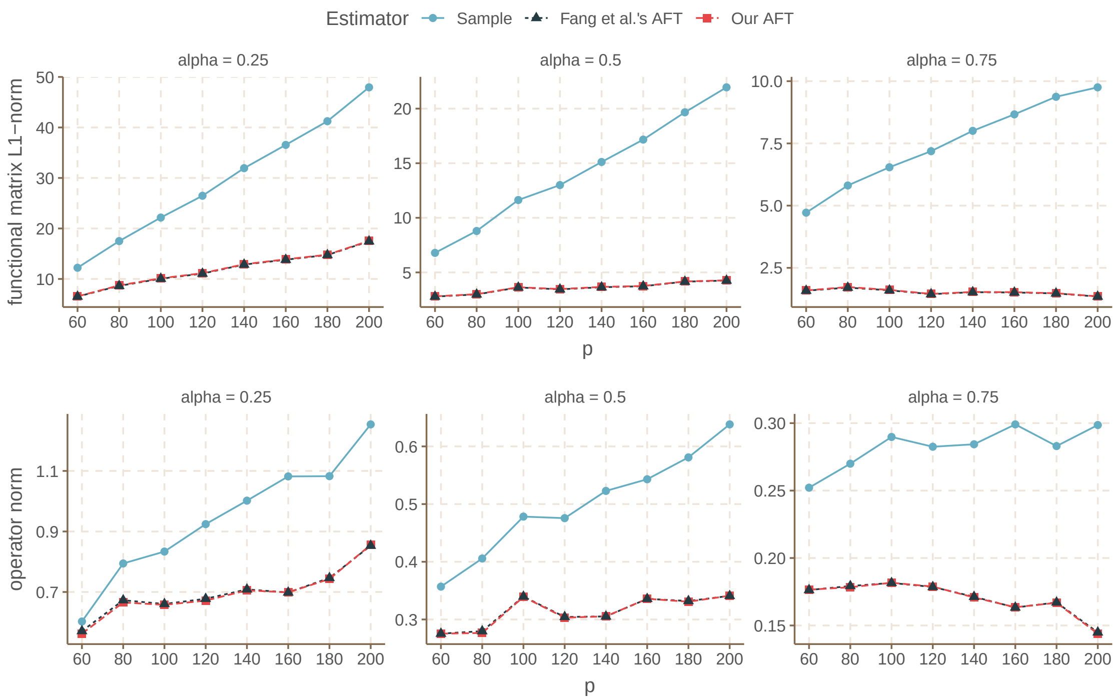

Figure S.3: The average losses of $\widehat{\boldsymbol{\Sigma}}_{\varepsilon}$ in functional matrix ℓ_1 norm (top row) and operator norm (bottom row) for DGP2 over 1000 simulation runs with $n = p = 60, 80, \ldots, 200$ and $\alpha = 0.25, 0.5, 0.75$.

We also give results for $\dot{C} = 1$ in Figures S.4 and S.5 to illustrate the robustness of our threshold choice.

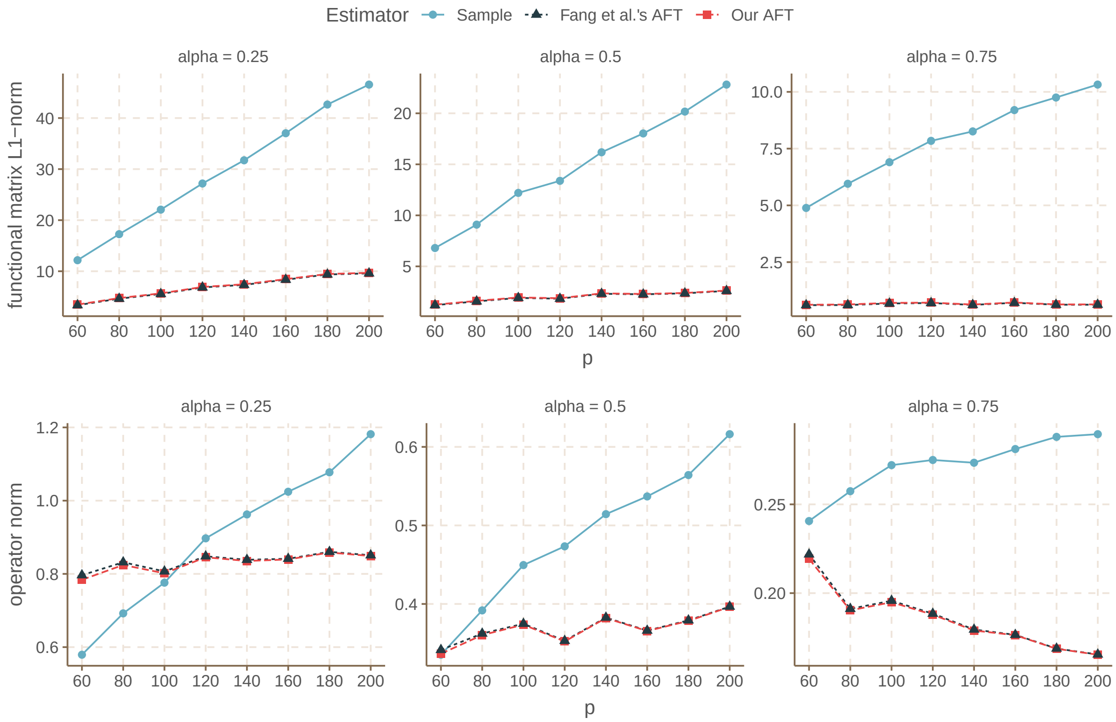

Figure S.4: The average losses of $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ in functional matrix ℓ_1 norm (top row) and operator norm (bottom row) for DGP1 with $\dot{C} = 1$ over 1000 simulation runs with $n = p = 60, 80, \ldots, 200$ and $\alpha = 0.25, 0.5, 0.75$.

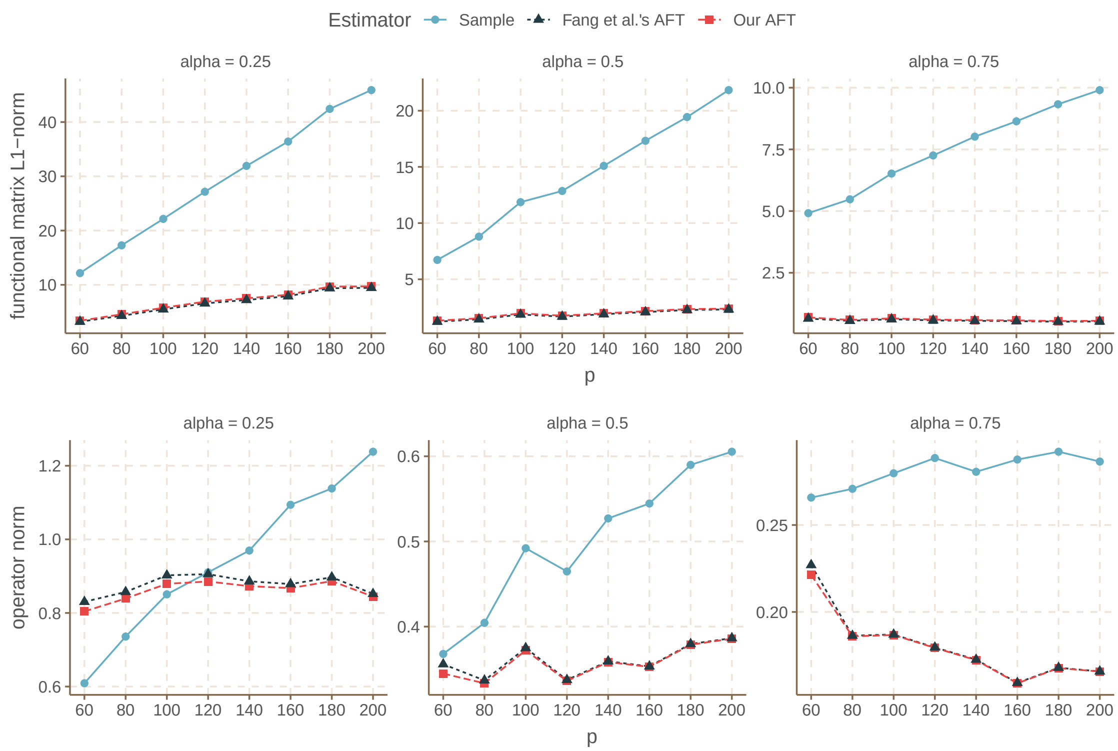

Figure S.5: The average losses of $\widehat{\boldsymbol{\Sigma}}_\varepsilon$ in functional matrix ℓ_1 norm (top row) and operator norm (bottom row) for DGP2 with $\dot{C} = 1$ over 1000 simulation runs with $n = p = 60, 80, \ldots, 200$ and $\alpha = 0.25, 0.5, 0.75$.

Figures S.6 and S.7 plot average losses of $\widehat{\boldsymbol{\Sigma}}_y$ measured by functional versions of elementwise ℓ_∞ norm, Frobenius norm and matrix ℓ_1 norm for DGP1 and DGP2, respectively, when $\dot{C} = 1$.

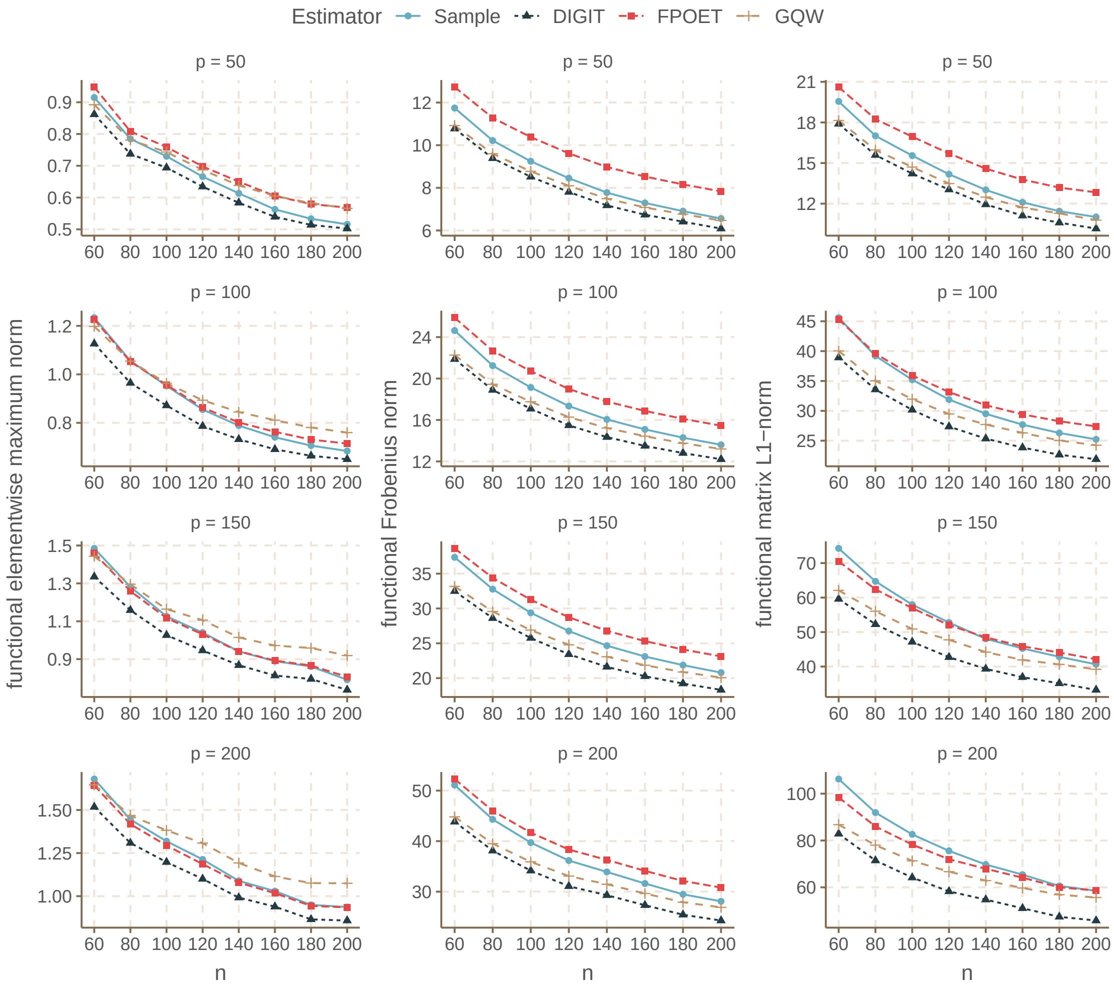

Figure S.6: The average losses of $\widehat{\mathbf{\Sigma}}_y$ in functional versions of elementwise ℓ_∞ norm (left column), Frobenius norm (middle column) and matrix ℓ_1 norm (right column) for DGP1 with $\dot{C} = 1$ over 1000 simulation runs.

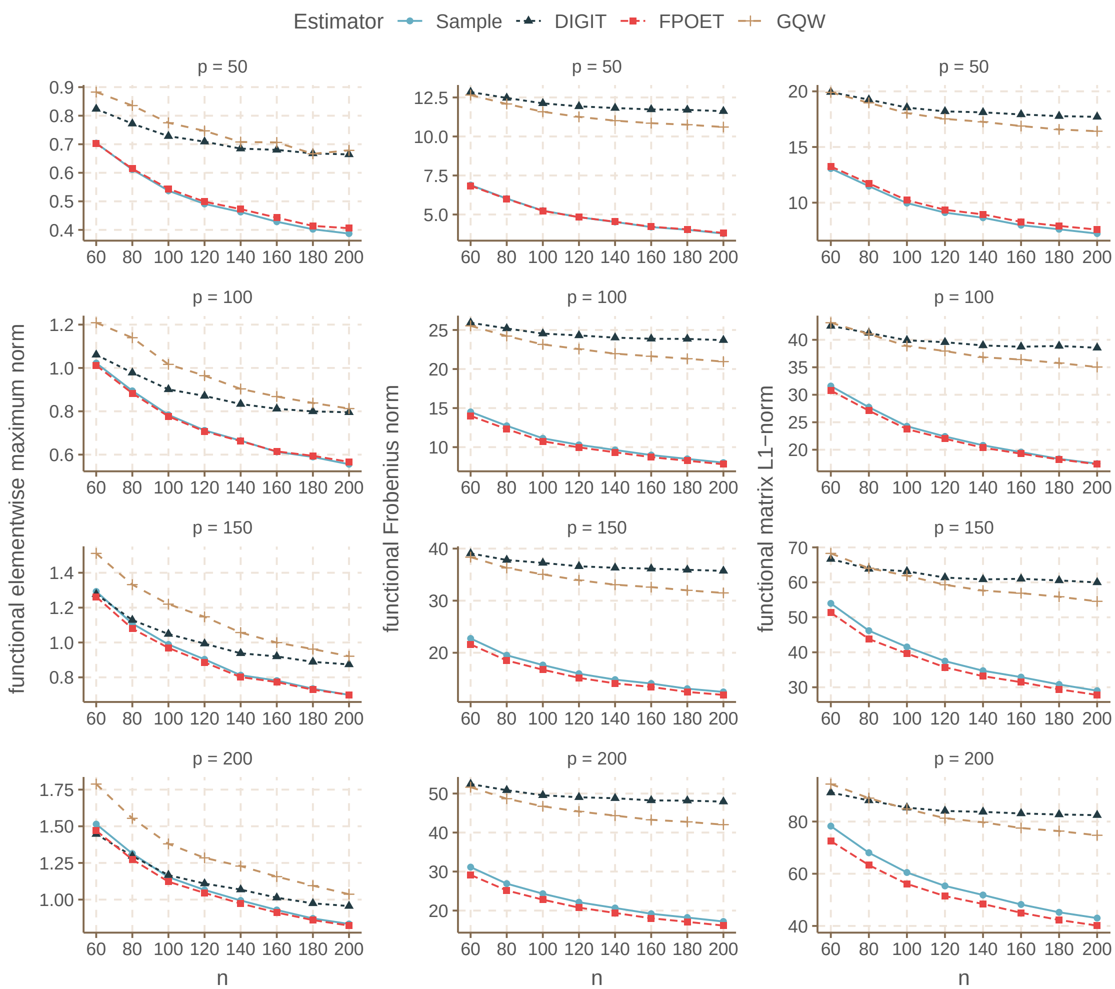

Figure S.7: The average losses of $\widehat{\boldsymbol{\Sigma}}_y$ in functional versions of elementwise ℓ_∞ norm (left column), Frobenius norm (middle column) and matrix ℓ_1 norm (right column) for DGP2 with $\dot{C} = 1$ over 1000 simulation runs.

G Additional real data results

This section presents additional results supporting Section 6. Table S.1 gives a list of inclusive countries with corresponding ISO Alpha-3 codes. Figure S.8 provides the rainbow plots of observed and smoothed age-specific log-mortality rates for females in GBR. Figure S.9 displays smoothed log-mortality rates for females in six randomly selected countries. Figure S.10 presents spatial heatmaps of factor loading of some European countries for males.

Figure S.11 provides the rainbow plots of the estimated age-specific factors for males.

Table S.1: List of inclusive countries with corresponding ISO Alpha-3 codes.

Country	Code	Country	Code	Country	Code	Country	Code
Australia	AUS	Estonia	EST	Lithuania	LTU	Russia	RUS
Austria	AUT	Finland	FIN	Latvia	LVA	Slovakia	SVK
Belgium	BEL	France	FRA	Luxembourg	LUX	Spain	ESP
Belarus	BLR	Hungary	HUN	Norway	NOR	Sweden	SWE
Bulgaria	BGR	Iceland	ISL	Portugal	PRT	Switzerland	CHE
Canada	CAN	Ireland	IRE	Poland	POL	Great Britain	GBR
Denmark	DNK	Italy	ITA	Netherlands	NLD	United States	USA
Czech Republic	CZE	Japan	JPN	New Zealand	NZL	Ukraine	UKR

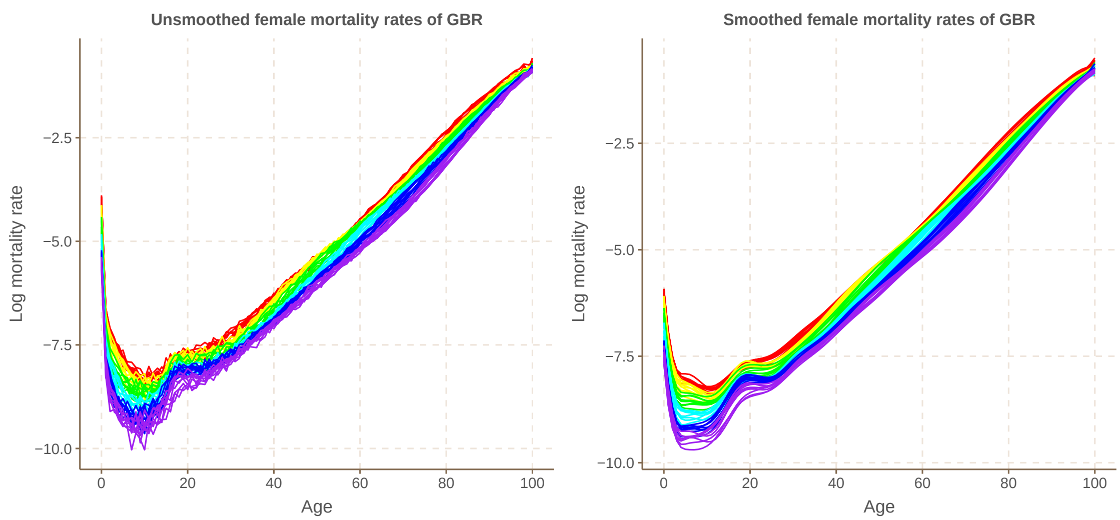

Figure S.8: The observed and smoothed age-specific log-mortality rates for females in GBR from 1960 to 2013.

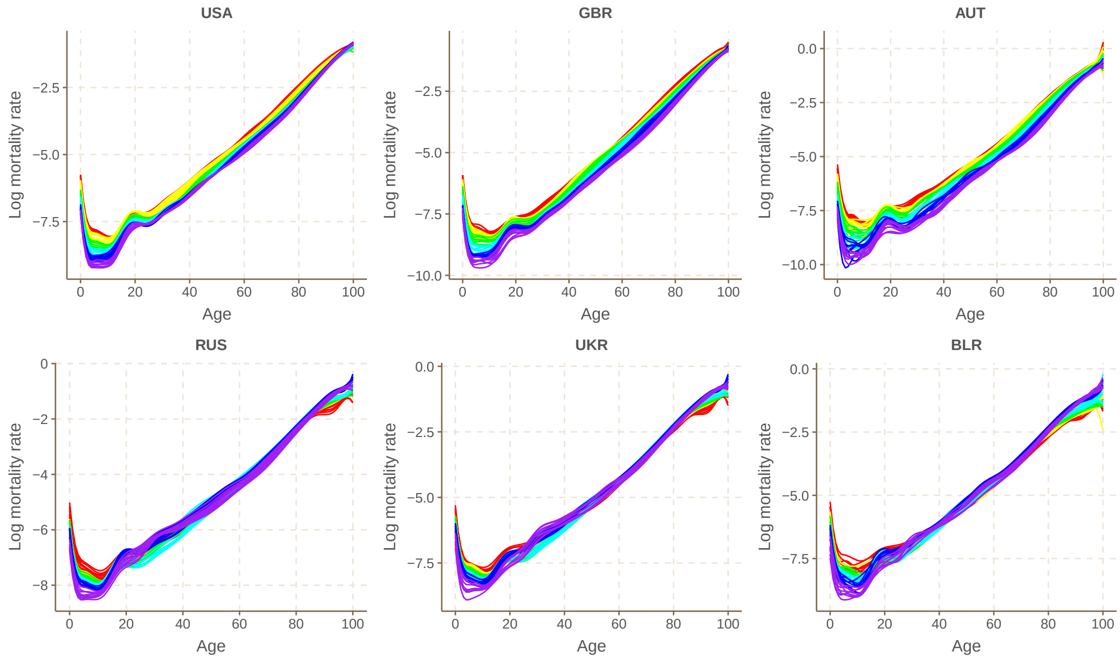

Figure S.9: The smoothed age-specific log-mortality rates for females in six randomly selected countries from 1960 to 2013.

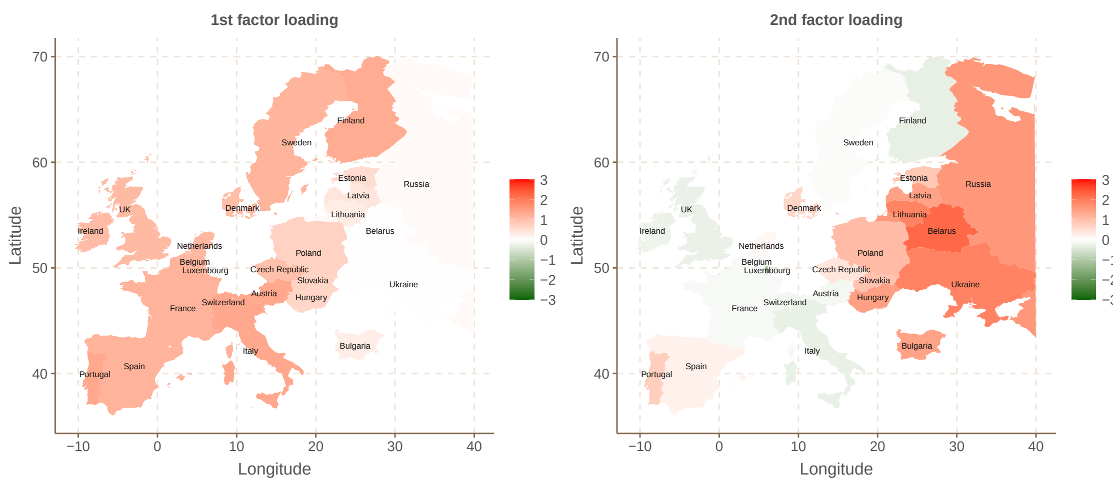

Figure S.10: Spatial heatmaps of factor loadings of some European countries for males.

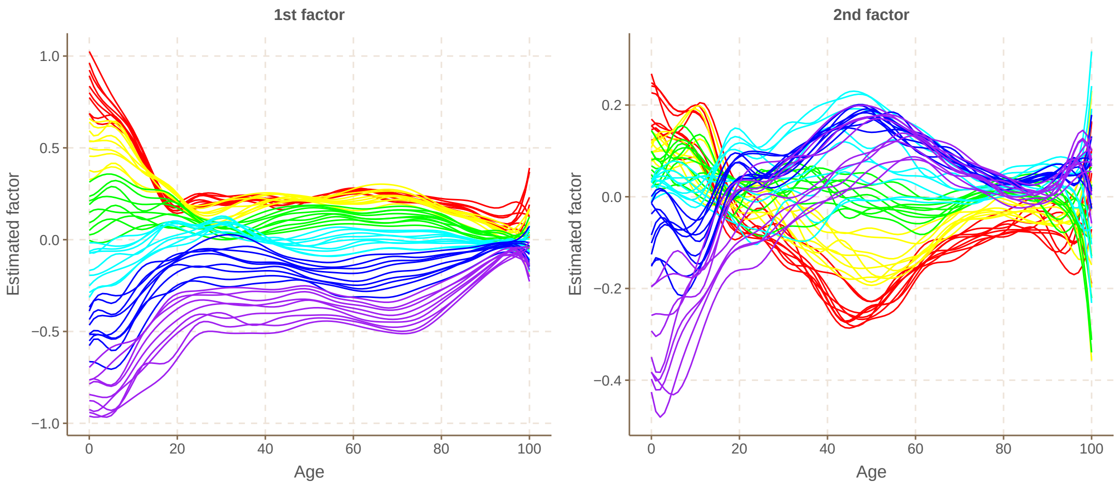

Figure S.11: The estimated age-specific factors for males.

References

Bai, J. (2003). Inferential theory for factor models of large dimensions, *Econometrica* **71**: 135–171.

Bathia, N., Yao, Q. and Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series, *The Annals of Statistics* **38**: 3352 - 3386.

Bosq, D. (2000). *Linear Processes in Function Spaces*, Springer, New York.

Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation, *Journal of the American Statistical Association* **106**: 672–684.

Carlsson, M. (2021). Von Neumann's trace inequality for Hilbert–Schmidt operators, *Expositiones Mathematicae* **39**: 149–157.

Davis, C. and Kahan, W. M. (1970). The rotation of eigenvectors by a perturbation. iii, *SIAM Journal on Numerical Analysis* **7**: 1–46.

Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements, *Journal of the Royal Statistical Society: Series B* **75**: 603–680.

Fan, J., Wang, W. and Zhong, Y. (2018). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation, *Journal of Machine Learning Research* **18**:

1–42.

Fang, Q., Guo, S. and Qiao, X. (2022). Finite sample theory for high-dimensional functional/scalar time series with applications, *Electronic Journal of Statistics* **16**: 527–591.

Fang, Q., Guo, S. and Qiao, X. (2023). Adaptive functional thresholding for sparse covariance function estimation in high dimensions, *Journal of the American Statistical Association*, *in press*.

Guo, S. and Qiao, X. (2023). On consistency and sparsity for high-dimensional functional time series with application to autoregressions, *Bernoulli* **29**: 451–472.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, John Wiley & Sons, Chichester.

Weyl, H. (1912). Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung), *Mathematische Annalen* **71**: 441–479.

Yu, Y., Wang, T. and Samworth, R. J. (2015). A useful variant of the Davis–Kahan theorem for statisticians, *Biometrika* **102**: 315–323.