

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

Dong Li<sup>1</sup>, Xinghao Qiao<sup>\*2</sup>, and Zihan Wang<sup>1</sup>

<sup>1</sup>*Department of Statistics and Data Science, Tsinghua University, Beijing 100084, China*

<sup>2</sup>*Faculty of Business and Economics, The University of Hong Kong, Hong Kong SAR*

## 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 model selection with theoretical guarantees. The convergence rates of involved estimated quantities are respectively established for DIGIT and FPOET estimators. Numerical studies including extensive simulations and a real data application on functional portfolio allocation are conducted to examine the finite-sample performance of the proposed methodology.

*Keywords:* Adaptive functional thresholding; Asymptotic identifiability; Eigenanalysis; Functional factor model; High-dimensional functional time series; Model selection.

*JEL code:* C13, C32, C38, C55

---

<sup>\*</sup>The authors' names are sorted alphabetically, and the corresponding author is Xinghao Qiao (xinghaoq@hku.hk).

# 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  $\mathbf{y}_t(\cdot) = \{y_{t1}(\cdot), \dots, y_{tp}(\cdot)\}^T$  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:

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

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

The estimation of the 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, multivariate functional linear regression (Chiou et al., 2016) and functional linear discriminant analysis (Xue et al., 2024). See Section 4 for details of some 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  $\hat{\boldsymbol{\Sigma}}_y^S$  performs poorly and some regularization is needed. Fang et al. (2024) 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  $\hat{\boldsymbol{\Sigma}}_y^S$  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. (2025), admits the representation with functional factors and scalar loadings:

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

where  $\mathbf{f}_t(\cdot)$  is a latent stationary  $r$ -vector of 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:

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

where  $\gamma_t$  is a latent stationary  $r$ -vector of factors and  $\mathbf{Q}(\cdot)$  is a  $p \times r$  matrix of functional factor loadings. We refer to  $\Sigma_f$ ,  $\Sigma_\chi$ ,  $\Sigma_\kappa$  and  $\Sigma_\varepsilon$  as the covariance matrix functions of  $\mathbf{f}_t$ ,  $\chi_t$ ,  $\kappa_t$  and  $\varepsilon_t$ , respectively. One may consider a more generalized FFM with functional factors and operator-valued loadings, see (13) below. However, estimating such a complex structure will introduce elevated errors when estimating the covariance matrix function. Moreover, by employing a basis expansion approach, the estimation of (13) can be reduced to that of (2), see Remark 2.5 below. Hence, our paper focuses on FFMs (1) and (2).

Within the FFM framework, our goal is to estimate the covariance matrix function  $\Sigma_y = \Sigma_\chi + \Sigma_\varepsilon$  for model (1) (or  $\Sigma_y = \Sigma_\kappa + \Sigma_\varepsilon$  for model (2)). Inspired by Fan et al. (2013), we impose the approximately functional sparsity assumption on  $\Sigma_\varepsilon$  instead of  $\Sigma_y$  directly giving rise to the conditional functional sparsity structure in models (1) and (2). To effectively separate  $\chi_t(\cdot)$  (or  $\kappa_t(\cdot)$ ) from  $\varepsilon_t(\cdot)$ , we rely on the spiked matrix model (Fan et al., 2021) and its functional generalization, i.e., a large nonnegative definite matrix or operator-valued matrix  $\mathbf{\Lambda} = \mathbf{L} + \mathbf{S}$ , 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

$$\underbrace{\Sigma_y(\cdot, *)}_{\mathbf{\Lambda}} = \underbrace{\mathbf{Q}(\cdot)\text{Cov}(\gamma_t)\mathbf{Q}(\cdot)^T}_{\mathbf{L}} + \underbrace{\Sigma_\varepsilon(\cdot, *)}_{\mathbf{S}}, \quad (3)$$

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

$$\underbrace{\int \int \Sigma_y(u, v) \Sigma_y(u, v)^T du dv}_{\mathbf{\Lambda}} = \underbrace{\mathbf{B} \left\{ \int \int \Sigma_f(u, v) \mathbf{B}^T \mathbf{B} \Sigma_f(u, v)^T du dv \right\} \mathbf{B}^T}_{\mathbf{L}} + \underbrace{\text{remaining terms}}_{\mathbf{S}}, \quad (4)$$

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

The new contribution of this paper can be summarized in five key aspects. First, though our model (1) shares the same form as the one in Guo et al. (2025) 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  $\varepsilon_t(\cdot)$  rather than assuming the white noise. (ii) Unlike the autocovariance-based method (Guo et al., 2025) 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. Building upon these findings, we provide mathematical insights when the functional factor analysis for models (1) and (2) and the proposed eigenanalysis of the respective  $\Lambda$ 's in (3) and (4) are approximately the same for high-dimensional functional data. Differing from the least squares method in Hallin et al. (2023), we develop a novel MFPCA method to estimate model (2) and also establish the theoretical equivalence of the covariance matrix function estimators based on two methods.

Third, we develop a new adaptive functional thresholding approach to estimate sparse  $\Sigma_\varepsilon$ . Compared to the competitor in Fang et al. (2024), our approach requires weaker assumptions while achieving similar finite-sample performance. Fourth, with the aid of such thresholding technique combined 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  $\Sigma_\varepsilon$ ,  $\Sigma_y$  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) with theoretical guarantees. Last but not least, we establish a new functional risk management framework to account for uncertainties in intraday returns of financial data, where our proposed estimators can be applied.

The rest of the paper is organized as follows. Section 2 presents the corresponding procedures for estimating  $\Sigma_y$  under two FFMs and the information criteria used for model selection. Section 3 provides the asymptotic theory for the 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 a financial data application in Section 6. Section 7 discusses some future work.

Throughout the paper, for any matrix  $\mathbf{M} = (M_{ij})_{p \times q}$ , we denote its matrices  $\ell_1$  norm,  $\ell_\infty$  norm, operator norm, Frobenius norm and elementwise  $\ell_\infty$  norm by  $\|\mathbf{M}\|_1 = \max_j \sum_i |M_{ij}|$ ,  $\|\mathbf{M}\|_\infty = \max_i \sum_j |M_{ij}|$ ,  $\|\mathbf{M}\| = \lambda_{\max}^{1/2}(\mathbf{M}^\top \mathbf{M})$ ,  $\|\mathbf{M}\|_F = (\sum_{i,j} M_{ij}^2)^{1/2}$  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  $\mathbf{f} = (f_1, \dots, f_p)^\top$ ,  $\mathbf{g} = (g_1, \dots, g_p)^\top \in \mathbb{H}^p$ , we denote the inner product by  $\langle \mathbf{f}, \mathbf{g} \rangle = \int_{\mathcal{U}} \mathbf{f}(u)^\top \mathbf{g}(u) du$  with induced norm  $\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$ . For an integral matrix operator  $\mathbf{K} : \mathbb{H}^p \rightarrow \mathbb{H}^q$  induced from the kernel matrix function  $\mathbf{K} = \{K_{ij}(\cdot, \cdot)\}_{q \times p}$  with each  $K_{ij} \in \mathbb{S}$ ,  $\mathbf{K}(\mathbf{f})(\cdot) = \int_{\mathcal{U}} \mathbf{K}(\cdot, u) \mathbf{f}(u) du \in \mathbb{H}^q$  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  $\ell_1$  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  $\|K_{ij}\|_{\mathcal{S}} = \{\iint K_{ij}(u, v)^2 du dv\}^{1/2}$ . Similarly, we define  $\|\mathbf{K}\|_{\mathcal{S},\infty} = \max_i \sum_j \|K_{ij}\|_{\mathcal{S}}$ ,  $\|\mathbf{K}\|_{\mathcal{S},F} = \{\sum_{i,j} \|K_{ij}\|_{\mathcal{S}}^2\}^{1/2}$  and  $\|\mathbf{K}\|_{\mathcal{S},\max} = \max_{i,j} \|K_{ij}\|_{\mathcal{S}}$  as the functional versions of matrix  $\ell_\infty$ , Frobenius and elementwise  $\ell_\infty$  norms, respectively. We define the operator norm by  $\|\mathbf{K}\|_{\mathcal{L}} = \sup_{\mathbf{x} \in \mathbb{H}^p, \|\mathbf{x}\| \leq 1} \|\mathbf{K}(\mathbf{x})\|$  and the trace norm by  $\|\mathbf{K}\|_{\mathcal{N}} = \text{tr}\{\int \mathbf{K}(u, u) du\}$  for  $p = q$ . For a positive integer  $m$ , write  $[m] = \{1, \dots, m\}$  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  $a_n/b_n \leq c$ , and  $a_n = o(b_n)$  if  $a_n/b_n \rightarrow 0$ . 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  $\mathbf{f}_t(\cdot) = \{f_{t1}(\cdot), \dots, f_{tr}(\cdot)\}^\top$  are uncorrelated with the idiosyncratic errors  $\boldsymbol{\varepsilon}_t(\cdot) = \{\varepsilon_{t1}(\cdot), \dots, \varepsilon_{tp}(\cdot)\}^\top$  and  $r$  is assumed to be fixed. Then we have

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

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:

$$\boldsymbol{\Omega} = \iint \boldsymbol{\Sigma}_y(u, v)\boldsymbol{\Sigma}_y(u, v)^\top dudv \equiv \boldsymbol{\Omega}_\mathcal{L} + \boldsymbol{\Omega}_\mathcal{R}, \quad (6)$$

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

**Assumption 1.**  $p^{-1}\mathbf{B}^\top\mathbf{B} = \mathbf{I}_r$  and  $\iint \boldsymbol{\Sigma}_f(u, v)\boldsymbol{\Sigma}_f(u, v)^\top dudv = \text{diag}(\theta_1, \dots, \theta_r)$ , where there exist some constants  $\bar{\theta} > \underline{\theta} > 0$  such that  $\bar{\theta} > \theta_1 > \theta_2 > \dots > \theta_r > \underline{\theta}$ .

**Remark 2.1.** 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  $p^{-1}\mathbf{B}^\top\mathbf{B} = \mathbf{I}_r$  and the other is  $\text{Cov}(\mathbf{f}_t) = \mathbf{I}_p$ . 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  $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_r)$ . Hence, we introduce an additional constraint based on the diagonalization of  $\iint \boldsymbol{\Sigma}_f(u, v)\boldsymbol{\Sigma}_f(u, v)^\top dudv$ , which is ensured by the fact that any nonnegative-definite matrix can be orthogonally diagonalized. Under Assumption 1, we can express  $\boldsymbol{\Omega}_\mathcal{L} = \sum_{i=1}^r p\theta_i \mathbf{b}_i \mathbf{b}_i^\top$ , 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  $\tilde{\mathbf{B}} = p^{-1/2}\mathbf{B} = (\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_r)$ , which satisfies  $\tilde{\mathbf{B}}^\top\tilde{\mathbf{B}} = \mathbf{I}_r$ . Under Assumption 1, it holds that  $\boldsymbol{\Omega}_\mathcal{L} = p^2 \sum_{i=1}^r \theta_i \tilde{\mathbf{b}}_i \tilde{\mathbf{b}}_i^\top$ , whose eigenvalue/eigenvector pairs are  $\{(p^2\theta_i, \tilde{\mathbf{b}}_i)\}_{i \in [r]}$ . Let  $\lambda_1 \geq \dots \geq \lambda_p$  be the ordered eigenvalues of  $\boldsymbol{\Omega}$  and  $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_p$  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)  $|\lambda_j - p^2\theta_j| \leq \|\boldsymbol{\Omega}_\mathcal{R}\|$  for  $j \in [r]$  and  $|\lambda_j| \leq \|\boldsymbol{\Omega}_\mathcal{R}\|$  for  $j \in [p] \setminus [r]$ ;
- (ii)  $\|\boldsymbol{\xi}_j - \tilde{\mathbf{b}}_j\| = O(p^{-2}\|\boldsymbol{\Omega}_\mathcal{R}\|)$  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  $\{\tilde{\mathbf{b}}_j\}_{j \in [r]}$ , provided that  $\|\boldsymbol{\Omega}_\mathcal{R}\| = o(p^2)$ . Towards this, we impose an approximately

functional sparsity condition on  $\Sigma_\varepsilon$  measured through

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

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 2.2.** (i) 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. (2024) 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  $\|\sigma_i\|_{\infty} = \sup_{u \in \mathcal{U}} \sigma_i(u) \geq \|\sigma_i\|_{\mathcal{N}}$ . As a result, we will use  $s_p$  instead of  $\tilde{s}_p$ .

(ii) With bounded  $\{\|\sigma_i\|_{\mathcal{N}}\}$ , we can easily obtain  $\|\Sigma_\varepsilon\|_{\mathcal{S},1} = \|\Sigma_\varepsilon\|_{\mathcal{S},\infty} = O(s_p)$ , which, along with Lemmas A.6, A.9 in the appendix and Assumption 1, yields that

$$\|\Omega_{\mathcal{R}}\| \leq \|\Sigma_\varepsilon\|_{\mathcal{S},\infty} \|\Sigma_\varepsilon\|_{\mathcal{S},1} + 2(\|\mathbf{B}\Sigma_f\mathbf{B}^T\|_{\mathcal{S},\infty} \|\mathbf{B}\Sigma_f\mathbf{B}^T\|_{\mathcal{S},1})^{1/2} (\|\Sigma_\varepsilon\|_{\mathcal{S},1} \|\Sigma_\varepsilon\|_{\mathcal{S},\infty})^{1/2} = O(s_p^2 + ps_p).$$

Hence, when  $s_p = o(p)$ , Proposition 1 implies that functional factor analysis under model (1) and eigen-analysis of  $\Omega$  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  $\Sigma_y(\cdot, \cdot)$  is given by  $\hat{\Sigma}_y^S(u, v) = n^{-1} \sum_{t=1}^n \mathbf{y}_t(u) \mathbf{y}_t(v)^T$ . Performing eigen-decomposition on the sample version of  $\Omega$ ,

$$\hat{\Omega} = \int \int \hat{\Sigma}_y^S(u, v) \hat{\Sigma}_y^S(u, v)^T du dv, \quad (8)$$

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

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

$$\sum_{t=1}^n \|\mathbf{y}_t - \hat{\mathbf{B}} \mathbf{f}_t\|^2 = \sum_{t=1}^n \int_{\mathcal{U}} \{\mathbf{y}_t(u) - \hat{\mathbf{B}} \mathbf{f}_t(u)\}^T \{\mathbf{y}_t(u) - \hat{\mathbf{B}} \mathbf{f}_t(u)\} du \quad (9)$$

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

Since  $\Sigma_\varepsilon$  is assumed to be functional sparse, we introduce an adaptive functional thresholding (AFT) estimator of  $\Sigma_\varepsilon$ . To this end, we define the functional variance factors  $\Theta_{ij}(u, v) = \text{Var}\{\varepsilon_{ti}(u) \varepsilon_{tj}(v)\}$  for

$i, j \in [p]$ , whose estimators are

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

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

$$\hat{\Sigma}_{\varepsilon, ij}^A = \|\hat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}} \times s_{\lambda} \left( \hat{\Sigma}_{\varepsilon, ij} / \|\hat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}} \right) \text{ with } \lambda = \dot{C}(\sqrt{\log p/n} + 1/\sqrt{p}), \quad (10)$$

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  $\hat{\Sigma}_{\varepsilon, ij} / \|\hat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}$  under functional elementwise  $\ell_{\infty}$  norm. Here  $s_{\lambda}$  is a functional thresholding operator with regularization parameter  $\lambda \geq 0$  (Fang et al., 2024) and belongs to the class  $s_{\lambda} : \mathbb{S} \rightarrow \mathbb{S}$  satisfying: (i)  $\|s_{\lambda}(Z)\|_{\mathcal{S}} \leq c\|Y\|_{\mathcal{S}}$  for all  $Z, Y \in \mathbb{S}$  that satisfy  $\|Z - Y\|_{\mathcal{S}} \leq \lambda$  and some  $c > 0$ ; (ii)  $\|s_{\lambda}(Z)\|_{\mathcal{S}} = 0$  for  $\|Z\|_{\mathcal{S}} \leq \lambda$ ; (iii)  $\|s_{\lambda}(Z) - Z\|_{\mathcal{S}} \leq \lambda$  for all  $Z \in \mathbb{S}$ . This class includes functional versions of commonly-adopted thresholding functions, such as soft thresholding, SCAD (Fan and Li, 2001), and the adaptive lasso (Zou, 2006).

**Remark 2.3.** By comparison, Fang et al. (2024) introduced an alternative AFT estimator

$$\tilde{\Sigma}_{\varepsilon}^A = (\tilde{\Sigma}_{\varepsilon, ij}^A)_{p \times p} \text{ with } \tilde{\Sigma}_{\varepsilon, ij}^A = \hat{\Theta}_{ij}^{1/2} \times s_{\lambda} \left( \hat{\Sigma}_{\varepsilon, ij} / \hat{\Theta}_{ij}^{1/2} \right), \quad (11)$$

which uses a single threshold level to functionally threshold standardized entries  $\hat{\Sigma}_{\varepsilon, ij} / \hat{\Theta}_{ij}^{1/2}$  across all  $(i, j)$ , resulting in entry-dependent functional thresholds for  $\hat{\Sigma}_{\varepsilon, ij}$ . Since  $\tilde{\Sigma}_{\varepsilon, jk}^A$  requires stronger assumptions (see Remark 2.2 above and the remark for Assumption 5 below), we adopt the AFT estimator  $\hat{\Sigma}_{\varepsilon, jk}^A$ , leading to comparable empirical performance (see Section S.6 of the supplementary material).

Finally, we obtain a DIGIT estimator of  $\Sigma_y$  as

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

## 2.2 FFM with functional loadings

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

**Remark 2.4.** (i) While both FFMs assume that different components of  $\mathbf{y}_t(\cdot)$  are defined in a common domain, FFM (2) can be generalized to allow different components to reside in different domains that may differ in dimensions (e.g., curves and images). However, this generalization is not feasible for FFM (1) as it involves integrals beyond the inner product used for FFM (2). In almost all real data applications in the existing literature, the corresponding components all lie in the same domain. Therefore, we focus on a conceptually special yet practical common case as in FFMs (1) and (2).



(ii) Both (1) and (2) yield useful FFM s but are tailored to tackle rather different situations. A crucial aspect of modeling functional time series is to characterize the functional and dynamic structures. FFM (1) assumes that both structures are inherited from functional time series factors  $\{\mathbf{f}_t(\cdot)\}$  with scalar loadings for enhanced interpretability. In contrast, FFM (2) incorporates dynamic information through scalar time series factors  $\{\gamma_t\}$  while preserving infinite-dimensionality in functional loadings of  $\mathbf{Q}(\cdot)$ .

**Remark 2.5.** Consider a more generalized FFM (Leng et al., 2024) with functional factors and operator-valued loadings:

$$\mathbf{y}_t(\cdot) = \tilde{\mathbf{A}}(\tilde{\mathbf{f}}_t)(\cdot) + \varepsilon_t(\cdot) = \int_{\mathcal{U}} \tilde{\mathbf{A}}(\cdot, v) \tilde{\mathbf{f}}_t(v) dv + \varepsilon_t(\cdot), \quad t \in [n], \quad (13)$$

where  $\tilde{\mathbf{f}}_t(\cdot)$  is a latent stationary  $r$ -vector of functional factors and  $\tilde{\mathbf{A}}(\cdot, \cdot)$  is a  $p \times r$  operator-valued matrix. For each  $j \in [r]$ , we apply the basis expansion approach to the  $j$ -th component of  $\tilde{\mathbf{f}}_t(\cdot)$ ,  $\tilde{f}_{tj}(\cdot) = \phi_j(\cdot)^T \tilde{\gamma}_t + \eta_{tj}(\cdot)$ , where  $\phi_j(\cdot)$  is a  $d$ -vector of basis function,  $\tilde{\gamma}_t$  is a  $d$ -vector of basis coefficients, and  $\eta_{tj}(\cdot)$  is the truncation error. Let  $\Phi(\cdot) = \{\phi_1(\cdot), \dots, \phi_r(\cdot)\}^T$ ,  $\boldsymbol{\eta}_t(\cdot) = \{\eta_{t1}(\cdot), \dots, \eta_{tr}(\cdot)\}^T$ ,  $\tilde{\mathbf{Q}}(\cdot) = \int_{\mathcal{U}} \tilde{\mathbf{A}}(\cdot, v) \Phi(v) dv$  and  $\tilde{\varepsilon}_t(\cdot) = \varepsilon_t(\cdot) + \int_{\mathcal{U}} \tilde{\mathbf{A}}(\cdot, v) \boldsymbol{\eta}_t(v) dv$ . Then, FFM (13) can be rewritten as  $\mathbf{y}_t(\cdot) = \tilde{\mathbf{Q}}(\cdot) \tilde{\gamma}_t + \tilde{\varepsilon}_t(\cdot)$ , which mirrors the form of FFM (2) with  $d$  scalar factors, as the truncation errors are incorporated into the idiosyncratic components. However, the estimation procedure demands strong technical conditions, and estimating the covariance matrix function within a more complex structure can lead to increased accumulated errors. Hence, our paper focuses on FFM (2) instead of FFM (13).

Under FFM (2), we have the covariance decomposition

$$\boldsymbol{\Sigma}_y(u, v) = \mathbf{Q}(u) \boldsymbol{\Sigma}_\gamma \mathbf{Q}(v)^T + \boldsymbol{\Sigma}_\varepsilon(u, v), \quad (u, v) \in \mathcal{U}^2. \quad (14)$$

By the multivariate version of Mercer's theorem, which serves as the foundation of MFPCA (Chiou et al., 2014; Happ and Greven, 2018), there exists an orthonormal basis consisting of eigenfunctions  $\{\varphi_i(\cdot)\}_{i=1}^\infty$  of  $\boldsymbol{\Sigma}_y$  and the associated eigenvalues  $\tau_1 \geq \tau_2 \geq \dots \geq 0$  such that

$$\boldsymbol{\Sigma}_y(u, v) = \sum_{i=1}^{\infty} \tau_i \varphi_i(u) \varphi_i(v)^T, \quad (u, v) \in \mathcal{U}^2, \quad (15)$$

where the sum converges absolutely and uniformly in both  $u$  and  $v$ .

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 (14), we impose a normalization-type condition similar to Assumption 1.

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

Suppose Assumption 1' holds, and let  $\tilde{\mathbf{q}}_1(\cdot), \dots, \tilde{\mathbf{q}}_r(\cdot)$  be the normalized columns of  $\mathbf{Q}(\cdot)$  such that  $\|\tilde{\mathbf{q}}_j\| = 1$  for  $j \in [r]$ . By Lemma A.11 in the appendix,  $\{\tilde{\mathbf{q}}_j(\cdot)\}_{j \in [r]}$  are the orthonormal eigenfunctions of the kernel function  $\mathbf{Q}(\cdot) \mathbf{Q}(\cdot)^T$  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  $\|\Sigma_\varepsilon\|_{\mathcal{L}} = o(p)$  hold. Then we have

- (i)  $|\tau_j - p\vartheta_j| \leq \|\Sigma_\varepsilon\|_{\mathcal{L}}$  for  $j \in [r]$  and  $|\tau_j| \leq \|\Sigma_\varepsilon\|_{\mathcal{L}}$  for  $j \in [p] \setminus [r]$ ;
- (ii)  $\|\varphi_j - \tilde{\mathbf{q}}_j\| = O(p^{-1}\|\Sigma_\varepsilon\|_{\mathcal{L}})$  for  $j \in [r]$ .

Proposition 2 implies that, if we can prove  $\|\Sigma_\varepsilon\|_{\mathcal{L}} = o(p)$ , then we can distinguish the principal eigenvalues  $\{\tau_j\}_{j \in [r]}$  from the remaining eigenvalues. Additionally, the first  $r$  eigenfunctions  $\{\varphi_j(\cdot)\}_{j \in [r]}$  are approximately the same as the normalized columns of  $\{\tilde{\mathbf{q}}_j(\cdot)\}_{j \in [r]}$ . To establish this, we impose the same functional sparsity condition on  $\Sigma_\varepsilon$  as measured by  $s_p$  in (7). Applying Lemma A.7(iii) in the appendix, we have  $\|\Sigma_\varepsilon\|_{\mathcal{L}} \leq \|\Sigma_\varepsilon\|_{\mathcal{S},1}^{1/2} \|\Sigma_\varepsilon\|_{\mathcal{S},\infty}^{1/2} = O(s_p)$ . 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  $\hat{\tau}_1 \geq \hat{\tau}_2 \geq \dots \geq 0$  be the eigenvalues of the sample covariance  $\hat{\Sigma}_y^S$  and  $\{\hat{\varphi}_j(\cdot)\}_{j=1}^\infty$  be their corresponding eigenfunctions. Then  $\hat{\Sigma}_y^S$  has the spectral decomposition

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

where  $\hat{\mathbf{R}}(u, v) = \sum_{j=r+1}^\infty \hat{\tau}_j \hat{\varphi}_j(u) \hat{\varphi}_j(v)^\top$  is the functional principal orthogonal complement. Applying AFT as introduced in Section 2.1 to  $\hat{\mathbf{R}}$  yields the estimator  $\hat{\mathbf{R}}^A$ . We finally obtain a FPOET estimator as

$$\hat{\Sigma}_y^F(u, v) = \sum_{j=1}^r \hat{\tau}_j \hat{\varphi}_j(u) \hat{\varphi}_j(v)^\top + \hat{\mathbf{R}}^A(u, v). \quad (16)$$

It is noteworthy that, with  $\Sigma_y$  satisfying decompositions (5) and (14) 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  $\Sigma_\varepsilon$ . 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 S.5.2 of the supplementary material for a discussion about the relationship between two FFMs.

We next present an equivalent representation of FPOET estimator (16) from a least squares perspective. We consider solving a constraint least squares minimization problem:

$$\{\hat{\mathbf{Q}}(\cdot), \hat{\Gamma}\} = \arg \min_{\mathbf{Q}(\cdot), \Gamma} \int \|\mathbf{Y}(u) - \mathbf{Q}(u)\Gamma^\top\|_{\mathbb{F}}^2 du = \arg \min_{\mathbf{Q}(\cdot), \gamma_1, \dots, \gamma_n} \sum_{t=1}^n \|\mathbf{y}_t - \mathbf{Q}\gamma_t\|^2, \quad (17)$$

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

$$\frac{1}{n} \sum_{t=1}^n \gamma_t \gamma_t^\top = \mathbf{I}_r \quad \text{and} \quad \frac{1}{p} \int \mathbf{Q}(u)^\top \mathbf{Q}(u) du \text{ is diagonal,}$$

where  $\mathbf{Y}(\cdot) = \{\mathbf{y}_1(\cdot), \dots, \mathbf{y}_n(\cdot)\}$  and  $\Gamma^\top = (\gamma_1, \dots, \gamma_n)$ . Given each  $\Gamma$ , setting the derivative of the objective in (17) w.r.t.  $\mathbf{Q}(\cdot)$  to zero, we obtain the constrained least squares estimator  $\tilde{\mathbf{Q}}(\cdot) = n^{-1} \mathbf{Y}(\cdot) \Gamma$ . Plugging it into (17), the objective as a function of  $\Gamma$  becomes  $\int \|\mathbf{Y}(u) - n^{-1} \mathbf{Y}(u) \Gamma \Gamma^\top \mathbf{Y}(u)\|_{\mathbb{F}}^2 du = \int \text{tr}\{(\mathbf{I}_n -$

$n^{-1}\mathbf{\Gamma}\mathbf{\Gamma}^T)\mathbf{Y}(u)^T\mathbf{Y}(u)\}du$ , whose minimizer is equivalent to the maximizer of  $\text{tr}[\mathbf{\Gamma}^T\{\int \mathbf{Y}(u)^T\mathbf{Y}(u)du\}\mathbf{\Gamma}]$ . This implies that the columns of  $n^{-1/2}\hat{\mathbf{\Gamma}}$  are the eigenvectors corresponding to the  $r$  largest eigenvalues of  $\int \mathbf{Y}(u)^T\mathbf{Y}(u)du \in \mathbb{R}^{n \times n}$ , and then  $\hat{\mathbf{Q}}(\cdot) = n^{-1}\mathbf{Y}(\cdot)\hat{\mathbf{\Gamma}}$ .

Let  $\tilde{\mathbf{e}}_t(\cdot) = \mathbf{y}_t(\cdot) - \hat{\mathbf{Q}}(\cdot)\hat{\gamma}_t$  and  $\tilde{\mathbf{\Sigma}}_\varepsilon(u, v) = n^{-1} \sum_{t=1}^n \tilde{\mathbf{e}}_t(u)\tilde{\mathbf{e}}_t(v)^T$ . Applying our proposed AFT in (10) to  $\tilde{\mathbf{\Sigma}}_\varepsilon$  yields the estimator  $\tilde{\mathbf{\Sigma}}_\varepsilon^A$ . Analogous to the decomposition (14) under Assumption 1', we propose the following substitution estimator

$$\hat{\mathbf{\Sigma}}_y^C(u, v) = \hat{\mathbf{Q}}(u)\hat{\mathbf{Q}}(v)^T + \tilde{\mathbf{\Sigma}}_\varepsilon^A(u, v). \quad (18)$$

The following proposition reveals the equivalence between the FPOET estimator (16) and the constrained least squares estimator (18).

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

**Remark 2.6.** (i) While our FFM (2) shares the same form as the model studied in [Hallin et al. \(2023\)](#); [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  $\mathbf{\Sigma}_y$ , 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  $\mathbf{\Sigma}_y$  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 S.5.1 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](#); [Ahn and Horenstein, 2013](#)) to determine the value of  $r$ .

For model (1), we let  $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_p$  be the ordered eigenvalues of  $\hat{\mathbf{\Omega}}$  in (8), and propose to estimate  $r$  by

$$\hat{r}^D = \arg \min_{j \in [r_{1,0}]} \frac{\hat{\lambda}_{j+1} + \vartheta_{1n}}{\hat{\lambda}_j + \vartheta_{1n}}, \quad (19)$$

where  $\vartheta_{1n}$  provides a lower bound correction to  $\hat{\lambda}_j$  for  $j > r$  ([Han et al., 2022](#)) and satisfies Assumption 7 below, and  $r_{1,0} > r$  is sufficiently large. For model (2), we employ a similar eigenvalue-ratio estimator:

$$\hat{r}^F = \arg \min_{j \in [r_{2,0}]} \frac{\hat{\tau}_{j+1} + \vartheta_{2n}}{\hat{\tau}_j + \vartheta_{2n}}, \quad (20)$$

where  $\vartheta_{2n}$  provides a lower bound correction to  $\hat{\tau}_j$  for  $j > r$  and satisfies Assumption 7' below,  $\{\hat{\tau}_i\}_{i=1}^\infty$  represents the ordered eigenvalues of the sample covariance  $\hat{\mathbf{\Sigma}}_y^S(\cdot, \cdot)$ , and  $r_{2,0} > r$  is sufficiently large. We set  $\vartheta_{1n} = c_r p^2 n^{-4/5}$  and  $\vartheta_{2n} = c_r p n^{-4/5}$  for some  $c_r > 0$  so that Assumptions 7 and 7' are satisfied. In our empirical analysis, we choose  $c_r = 0.1$ , which consistently yields good finite-sample performance.

## 2.4 Model selection criterion

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

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

$$V^{\mathcal{D}}(k) = (pn)^{-1} \sum_{t=1}^n \|\mathbf{y}_t - p^{-1} \hat{\mathbf{B}}_k \hat{\mathbf{B}}_k^{\top} \mathbf{y}_t\|^2,$$

where  $\hat{\mathbf{B}}_k$  is the estimated factor loading matrix by DIGIT. Analogously, when  $k$  scalar factors are estimated under FFM (2), it follows from the objective function in (17) that the corresponding mean squared residuals is

$$V^{\mathcal{F}}(k) = (pn)^{-1} \sum_{t=1}^n \|\mathbf{y}_t - n^{-1} \mathbf{Y} \hat{\mathbf{\Gamma}}_k \hat{\gamma}_{t,k}\|^2,$$

where  $\hat{\mathbf{\Gamma}}_k^{\top} = (\hat{\gamma}_{1,k}, \dots, \hat{\gamma}_{n,k})$  is formed by estimated factors using FPOET.

For any given  $k$ , we propose the following information criteria:

$$\text{IC}^{\mathcal{D}}(k) = \log\{V^{\mathcal{D}}(k)\} + kg^{\mathcal{D}}(p, n) \quad \text{and} \quad \text{IC}^{\mathcal{F}}(k) = \log\{V^{\mathcal{F}}(k)\} + kg^{\mathcal{F}}(p, n), \quad (21)$$

where  $\text{IC}^{\mathcal{D}}(k)$  and  $\text{IC}^{\mathcal{F}}(k)$  are represented as the sum of the log transformation of the average of squared residuals and the penalty term with  $g^{\mathcal{D}}(p, n)$  and  $g^{\mathcal{F}}(p, n)$  being the corresponding penalty functions of  $(p, n)$  to avoid overparameterization. While there is much existing literature (see e.g., [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 the same penalty functions  $g^{\mathcal{D}}(p, n) = g^{\mathcal{F}}(p, n) = g(p, n)$  in (21), referred to as  $\text{IC}_1$ ,  $\text{IC}_2$ , and  $\text{IC}_3$ , respectively,

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

For model selection, we define the information criterion difference between two FFMs as  $\Delta \text{IC}_i = \text{IC}_i^{\mathcal{D}}(\hat{r}^{\mathcal{D}}) - \text{IC}_i^{\mathcal{F}}(\hat{r}^{\mathcal{F}})$  for  $i = 1, 2, 3$ . The negative (or positive) values of  $\Delta \text{IC}_i$ 's indicate that FFM (1) (or FFM (2)) is more suitable based on the observed data  $\{\mathbf{y}_t(\cdot)\}$ . See the model selection consistency guarantee in Theorem 3 under mild requirements on penalty functions, which three examples of  $g(p, n)$  satisfy.

In general, instead of using the same penalty functions in  $\text{IC}^{\mathcal{D}}$  and  $\text{IC}^{\mathcal{F}}$ , we can employ different penalty functions that satisfy the requirements of Theorem 3 from a model complexity perspective. Specifically, considering the panel nature of the problem with an effective number of observations  $pn$  and assuming each function has a complexity of  $s_n$ , the effective total number of parameters for each model is calculated as the total number of parameters minus the number of constraints in Assumption 1 or 1', which yields  $kp + kns_n - k^2$  for FFM (1) and  $kps_n + kn - k^2$  for FFM (2). Therefore, the feasible options for the penalty functions are  $g^{\mathcal{D}}(p, n) = (p + ns_n - k)/pn$  and  $g^{\mathcal{F}}(p, n) = (ps_n + n - k)/pn$ . However, rigorously determining  $s_n$  remains an open question. Theoretically, by imposing specific requirements on  $s_n$ , Theorem 3 can ensure

the model selection consistency of  $\text{IC}^\mathcal{D}$  and  $\text{IC}^\mathcal{F}$ . In practice, one can select the leading  $s_n$  eigenvalues of the corresponding covariance functions such that the cumulative percentage exceeds a certain threshold or adopt the eigenvalue-ratio-based method.

### 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),  $\{\mathbf{f}_t(\cdot)\}_{t \geq 1}$  and  $\{\boldsymbol{\varepsilon}_t(\cdot)\}_{t \geq 1}$  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),  $\{\boldsymbol{\gamma}_t\}_{t \geq 1}$  and  $\{\boldsymbol{\varepsilon}_t(\cdot)\}_{t \geq 1}$  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)  $\mathbb{E}\|p^{-1/2}\mathbf{B}^\mathbf{T}\boldsymbol{\varepsilon}_t\|^2 < C$ ; (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)  $\mathbb{E}\|p^{-1/2} \int \mathbf{Q}(u)^\mathbf{T} \boldsymbol{\varepsilon}_t(u) du\|^2 < C'$  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). Assumptions 3(ii) and 3'(ii) involve weaker moment constraints compared to Fan et al. (2013) and are needed to estimate factors and loadings consistently. Assumption 3(iii) and 3'(iii) generalize the standard conditions for scalar factor models (Fan et al., 2013; 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 (Chang et al., 2024) 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

$$\mathcal{M}_y = 2\pi \cdot \text{ess sup}_{\theta \in [-\pi, \pi], \phi \in \mathbb{H}_0^p} \frac{\langle \phi, \mathbf{f}_{y, \theta}(\phi) \rangle}{\langle \phi, \boldsymbol{\Sigma}_y(\phi) \rangle}, \quad (22)$$

where  $\boldsymbol{\Sigma}_y(\phi)(\cdot) = \int_{\mathcal{U}} \boldsymbol{\Sigma}_y(\cdot, v) \phi(v) dv$  and  $\mathbb{H}_0^p = \{\phi \in \mathbb{H}^p : \langle \phi, \boldsymbol{\Sigma}_y(\phi) \rangle \in (0, \infty)\}$ . When  $\mathbf{y}_1(\cdot), \dots, \mathbf{y}_n(\cdot)$  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  $\{\boldsymbol{\varepsilon}_t(\cdot)\}$ ,  $\mathcal{M}_f$  of  $\{\mathbf{f}_t(\cdot)\}$  and  $\mathcal{M}_\gamma$  of scalar time series  $\{\boldsymbol{\gamma}_t\}$  (Basu and Michailidis, 2015). To derive 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 S.5.3 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)  $\mathcal{M}_f < \infty, \mathcal{M}_\varepsilon < \infty$  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)  $\mathcal{M}_\gamma < \infty, \mathcal{M}_\varepsilon < \infty$  and  $\mathcal{M}_\varepsilon^2 \log p = o(n)$ .

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

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

**Assumption 7.**  $\vartheta_{1n}p^{-2} \rightarrow 0, \vartheta_{1n}\mathcal{M}_\varepsilon^{-2}p^{-2}n \rightarrow \infty$  and  $\vartheta_{1n} \rightarrow \infty$ , where  $\vartheta_{1n}$  is specified in (19).

**Assumption 7'.**  $\vartheta_{2n}p^{-1} \rightarrow 0, \vartheta_{2n}\mathcal{M}_\varepsilon^{-2}p^{-1}n \rightarrow \infty$  and  $\vartheta_{2n}p \rightarrow \infty$ , where  $\vartheta_{2n}$  is specified in (20).

**Assumption 8.** For model (1), (i) there exists some constant  $\underline{\omega} > 0$  such that  $\omega_1 \geq \omega_2 \geq \dots \geq \omega_{r_{2,0}+1} \geq \underline{\omega}$ , where  $\{\omega_i\}_{i=1}^\infty$  are the ordered eigenvalues of  $\boldsymbol{\Sigma}_f(\cdot, \cdot)$  and  $r_{2,0}$  is given in (20); (ii)  $\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{N}} = o(p)$ .

**Assumption 8'.** For model (2), (i)  $\text{rank}\{\int \mathbf{Q}(u)\mathbf{Q}(u)^\top du\} \geq r_{1,0} + 1$ , where  $r_{1,0}$  is given in (19); (ii)  $\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{N}} = o(p)$ .

Assumption 5 is required when implementing AFT, however, it is weaker than the similar assumption  $\inf_{(u,v) \in \mathcal{U}^2} \min_{i,j \in [p]} \text{Var}[\varepsilon_{ti}(u)\varepsilon_{tj}(v)] \geq \tau$  imposed in Fang et al. (2024). Assumption 6 allows the high-dimensional case, where  $p$  grows exponentially as  $n$  increases. Assumptions 7 and 7' on the lower bound correlations (i.e.,  $\vartheta_{1n}$  and  $\vartheta_{2n}$ ) are imposed to ensure the consistency of ratio-based estimators for the number of factors. Both assumptions are satisfied by setting  $\vartheta_{1n} \asymp p^2 n^{-\rho}$  and  $\vartheta_{2n} \asymp p n^{-\rho}$  when  $\mathcal{M}_\varepsilon = O(1)$  and  $n^\rho = o(p^2)$  for  $\rho \in (0, 1)$ . Assumptions 8 and 8' are needed when establishing the model selection consistency for the proposed information criteria. For model (1), Assumption 8(i) implies that the common covariance  $\boldsymbol{\Sigma}_\chi(\cdot, \cdot) = \mathbf{B}\boldsymbol{\Sigma}_f(\cdot, \cdot)\mathbf{B}^\top$  has at least  $r_{2,0} + 1$  nonzero eigenvalues of order  $p$ , which cannot be recovered by the FPOET estimator with  $\hat{r}^\mathcal{F} \leq r_{2,0}$  factors. Similarly, for model (2), Assumption 8'(i) implies that  $\int \int \boldsymbol{\Sigma}_\kappa(u, v)\boldsymbol{\Sigma}_\kappa(u, v)^\top dudv = \int \mathbf{Q}(u)\mathbf{Q}(u)^\top du$  (which holds under Assumption 1') has at least  $r_{1,0} + 1$  nonzero eigenvalues of order  $p^2$ , which cannot be recovered by the DIGIT estimator with  $\hat{r}^\mathcal{D} \leq r_{1,0}$  factors. To guarantee the asymptotic identifiability of the respective FFMs for model selection, we require Assumptions 8(ii) and 8'(ii) to hold for  $\{\boldsymbol{\varepsilon}_t(\cdot)\}$ .

### 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)  $\|\hat{\mathbf{B}} - \mathbf{B}\mathbf{U}^\top\|_{\max} = O_p(\varpi_{n,p})$ ; (ii)  $n^{-1} \sum_{t=1}^n \|\hat{\mathbf{f}}_t - \mathbf{U}\mathbf{f}_t\|^2 = O_p(\mathcal{M}_\varepsilon^2/n + 1/p)$ ; (iii)  $\max_{t \in [n]} \|\hat{\mathbf{f}}_t - \mathbf{U}\mathbf{f}_t\| = O_p(\mathcal{M}_\varepsilon \sqrt{\log n/n} + \sqrt{\log n/p})$ .

The orthogonal matrix  $\mathbf{U}$  above is needed to ensure that  $\mathbf{b}_j^\top \hat{\mathbf{b}}_j \geq 0$  for each  $j \in [r]$ . Provided that  $\hat{\mathbf{B}}\mathbf{U}\mathbf{U}^\top \hat{\mathbf{f}}_t = \hat{\mathbf{B}}\hat{\mathbf{f}}_t$ , the estimation of the common components and  $\Sigma_y$  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  $\check{\mathbf{b}}_i$  and  $\check{\mathbf{b}}_i$  denote the  $i$ -th rows of  $\mathbf{B}$  and  $\hat{\mathbf{B}}$ , respectively.

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

In the context of FPOET estimation of factors and loadings, we require an additional asymptotically orthogonal matrix  $\mathbf{H}$  such that  $\hat{\gamma}_t$  is a valid estimator of  $\mathbf{H}\gamma_t$ . 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  $\hat{\Sigma}_y^S$  in a decreasing order. Define  $\mathbf{H} = n^{-1} \mathbf{V}^{-1} \hat{\Gamma}^\top \Gamma \int \mathbf{Q}(u)^\top \mathbf{Q}(u) du$ . By Lemma S.6 of the supplementary material,  $\mathbf{H}$  is asymptotically orthogonal such that  $\mathbf{I}_r = \mathbf{H}^\top \mathbf{H} + o_p(1) = \mathbf{H}\mathbf{H}^\top + o_p(1)$ .

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

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

When  $\mathcal{M}_\varepsilon = O(1)$ , the convergence rates presented in Theorem 1(i), (ii) for model (1) are, respectively, consistent to those established in Fan et al. (2013) and Bai and Ng (2002), and the uniform convergence rates presented in Theorem 1(iii) and Corollary 1 are faster than those established in Fan et al. (2013). 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.** (i) *Under Assumptions 1-4 and 7,  $\mathbb{P}(\hat{r}^\mathcal{D} = r) \rightarrow 1$  as  $p, n \rightarrow \infty$ , where  $\hat{r}^\mathcal{D}$  is defined in (19).* (ii) *Under Assumptions 1'-4' and 7',  $\mathbb{P}(\hat{r}^\mathcal{F} = r) \rightarrow 1$  as  $p, n \rightarrow \infty$ , where  $\hat{r}^\mathcal{F}$  is defined in (20).*

**Theorem 3.** *Suppose that  $g(p, n) \rightarrow 0$  and  $(\mathcal{M}_\varepsilon^2/n + 1/p)^{-1}g(p, n) \rightarrow \infty$  as  $p, n \rightarrow \infty$  for both penalty functions  $g^\mathcal{D}(p, n)$  and  $g^\mathcal{F}(p, n)$  in (21). Then, (i) under Assumptions 1-4, 7 and 8,  $\mathbb{P}\{\text{IC}^\mathcal{D}(\hat{r}^\mathcal{D}) < \text{IC}^\mathcal{F}(\hat{r}^\mathcal{F})\} \rightarrow 1$  as  $p, n \rightarrow \infty$ ; (ii) under Assumptions 1'-4', 7' and 8',  $\mathbb{P}\{\text{IC}^\mathcal{D}(\hat{r}^\mathcal{D}) > \text{IC}^\mathcal{F}(\hat{r}^\mathcal{F})\} \rightarrow 1$  as  $p, n \rightarrow \infty$ .*

**Remark 3.1.** *With the aid of Theorems 2 and 3, our estimators explored in Sections 3.2 and 3.3 are asymptotically adaptive to the number of factors and the data-generating model. To see this, consider, e.g., model (2), and let  $\hat{\kappa}_{ti, \hat{r}}(\cdot)$  be the estimated common component, and  $\hat{\gamma}_{t, \hat{r}}$  and  $\hat{\mathbf{q}}_{i, \hat{r}}(\cdot)$  be constructed using  $\hat{r}^\mathcal{F}$  estimated scalar factors and functional loadings. Then, for any constant  $\tilde{c} > 0$ ,  $\mathbb{P}(\varrho^{-1} \max_{i \in [p], t \in [n]} \|\hat{\kappa}_{ti, \hat{r}} - \kappa_{ti}\| > \tilde{c}) \leq \mathbb{P}\{\varrho^{-1} \max_{i \in [p], t \in [n]} \|\hat{\mathbf{q}}_i^\top \hat{\gamma}_t - \mathbf{q}_i^\top \gamma_t\| > \tilde{c} | \hat{r}^\mathcal{F} = r, \text{IC}^\mathcal{D}(\hat{r}^\mathcal{D}) > \text{IC}^\mathcal{F}(\hat{r}^\mathcal{F})\} + \mathbb{P}(\hat{r}^\mathcal{F} \neq r) + \mathbb{P}\{\text{IC}^\mathcal{D}(\hat{r}^\mathcal{D}) \leq \text{IC}^\mathcal{F}(\hat{r}^\mathcal{F})\}$ , which, combined with Corollary 1', implies that  $\max_{i \in [p], t \in [n]} \|\hat{\kappa}_{ti, \hat{r}} - \kappa_{ti}\| = O_p(\varrho)$ . Similar arguments can be applied to other estimated quantities in Sections 3.2 and 3.3. Therefore, we assume that the number of factors and data-generating model are known in our asymptotic results.*

### 3.3 Convergence of estimated covariance matrix functions

Estimating the idiosyncratic covariance matrix function  $\Sigma_\varepsilon$  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  $\Sigma_\varepsilon$  under functional matrix  $\ell_1$  norm  $\|\cdot\|_{S,1}$  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., 2024) when  $\mathcal{M}_\varepsilon = O(1)$  and  $p \log p \gtrsim n$ .

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

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

When assessing the convergence criteria for our DIGIT and FPOET estimators, it is crucial to note that functional matrix norms such as  $\|\cdot\|_{S,1}$  and  $\|\cdot\|_{\mathcal{L}}$  are not suitable choices. This is because  $\hat{\Sigma}_y$  may not converge to  $\Sigma_y$  in these norms for high-dimensional functional data, unless specific structural assumptions are directly imposed on  $\Sigma_y$ . 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  $\ell_\infty$  norm  $\|\cdot\|_{S,\max}$ .

**Theorem 5.** *Under the assumptions of Theorem 4, we have  $\|\hat{\Sigma}_y^D - \Sigma_y\|_{S,\max} = O_p(\varpi_{n,p})$ .*

**Theorem 5'.** *Under the assumptions of Theorem 4', we have  $\|\hat{\Sigma}_y^F - \Sigma_y\|_{S,\max} = O_p(\varpi_{n,p})$ .*

**Remark 3.2.** (i) *The convergence rates of DIGIT and FPOET estimators (we use  $\hat{\Sigma}_y$  to denote both) comprise two terms. The first term  $O_p(\mathcal{M}_\varepsilon \sqrt{\log p/n})$  arises from the rate of  $\hat{\Sigma}_y^S$ , 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  $\hat{\Sigma}_y^S$ , we observe that using a factor-guided approach results in the same rate in  $\|\cdot\|_{S,\max}$  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,  $\hat{\Sigma}_y$  converges to  $\Sigma_y$  in the high-dimensional case (see Theorem 7), while  $\hat{\Sigma}_y^S$  does not achieve this convergence. Second, as evidenced by empirical results in Sections 5 and 6,  $\hat{\Sigma}_y$  significantly outperforms  $\hat{\Sigma}_y^S$  in terms of various functional matrix losses.*

Finally, we explore convergence properties of the inverse covariance matrix function estimation. Based on Section 3.5 of Hsing and Eubank (2015), although the inverse operator  $\Sigma_y^{-1}$  may not be well-defined, we instead use the Moore–Penrose inverse. Denote the null space of  $\Sigma_y$  and its orthogonal complement by  $\text{Ker}(\Sigma_y) = \{\mathbf{x} \in \mathbb{H}^p : \Sigma_y(\mathbf{x}) = \mathbf{0}\}$  and  $\text{Ker}(\Sigma_y)^\perp = \{\mathbf{x} \in \mathbb{H}^p : \langle \mathbf{x}, \mathbf{y} \rangle = 0, \forall \mathbf{y} \in \text{Ker}(\Sigma_y)\}$ , respectively. Let  $\tilde{\Sigma}_y$  be the restriction of  $\Sigma_y$  to  $\text{Ker}(\Sigma_y)^\perp$ . By Definition 3.5.7 of Hsing and Eubank (2015), the Moore–Penrose inverse of  $\Sigma_y$  is defined as  $\Sigma_y^\dagger(\mathbf{x}) = \tilde{\Sigma}_y^{-1}(\mathbf{x})$  for  $\mathbf{x} \in \text{Im}(\Sigma_y)$  and 0 for  $\mathbf{x} \in \text{Im}(\Sigma_y)^\perp$ . The similar definition applies to the Moore–Penrose inverses of other covariance matrix operators. To obtain the inverse DIGIT estimator, we assume  $\text{Ker}(\Sigma_\varepsilon) = \text{Ker}(\mathbf{B}\Sigma_f\mathbf{B}^\top)$ , which implies that  $\text{Ker}(\Sigma_\varepsilon) = \text{Ker}(\Sigma_y)$ , and  $\tilde{\Sigma}_y$ ,  $\tilde{\Sigma}_\varepsilon$  and  $\mathbf{B}\tilde{\Sigma}_f\mathbf{B}^\top$  are all invertible on  $\text{Ker}(\Sigma_y)^\perp$ . We then rely on (5) to apply Sherman–Morrison–Woodbury formula (Theorem 3.5.6 of Hsing and Eubank (2015)) to obtain its inverse  $\Sigma_y^\dagger = \Sigma_\varepsilon^\dagger - \Sigma_\varepsilon^\dagger \mathbf{B}(\Sigma_f^\dagger + \mathbf{B}^\top \Sigma_\varepsilon^\dagger \mathbf{B})^\dagger \mathbf{B}^\top \Sigma_\varepsilon^\dagger$ , and the plug-in inverse DIGIT estimator is  $(\hat{\Sigma}_y^D)^\dagger = \hat{\Sigma}_\varepsilon^\dagger - \hat{\Sigma}_\varepsilon^\dagger \hat{\mathbf{B}}(\hat{\Sigma}_f^\dagger + \hat{\mathbf{B}}^\top \hat{\Sigma}_\varepsilon^\dagger \hat{\mathbf{B}})^\dagger \hat{\mathbf{B}}^\top \hat{\Sigma}_\varepsilon^\dagger$ . The plug-in inverse FPOET estimator  $(\hat{\Sigma}_y^F)^\dagger$  can be defined similarly by assuming that  $\text{Ker}(\Sigma_\varepsilon) = \text{Ker}(\mathbf{Q}\Sigma_\gamma\mathbf{Q}^\top)$ . To make  $(\hat{\Sigma}_y^D)^\dagger$  and  $(\hat{\Sigma}_y^F)^\dagger$  meaningful inverse covariance matrix estimators, we focus on



finite-dimensional functional objects  $\{\mathbf{y}_t(\cdot)\}_{t \in [n]}$ , i.e.,  $\|\Sigma_y^\dagger\|_{\mathcal{L}}$  is bounded. Then, both the inverse DIGIT and FPOET estimators are consistent in the operator norm, as presented in the following theorems.

**Theorem 6.** *Suppose that the assumptions of Theorem 5 hold,  $\varpi_{n,p}^{1-q} s_p = o(1)$ ,  $\text{Ker}(\Sigma_\varepsilon) = \text{Ker}(\mathbf{B}\Sigma_f\mathbf{B}^\top)$ , and both  $\|\Sigma_\varepsilon^\dagger\|_{\mathcal{L}}$  and  $\|\Sigma_f^\dagger\|_{\mathcal{L}}$  are bounded. Then,  $\hat{\Sigma}_y^{\mathcal{D}}$  has a bounded Moore–Penrose inverse with probability approaching one, and  $\|(\hat{\Sigma}_y^{\mathcal{D}})^\dagger - \Sigma_y^\dagger\|_{\mathcal{L}} = O_p(\varpi_{n,p}^{1-q} s_p)$ .*

**Theorem 6'.** *Suppose that the assumptions of Theorem 5' hold,  $\varpi_{n,p}^{1-q} s_p = o(1)$ ,  $\text{Ker}(\Sigma_\varepsilon) = \text{Ker}(\mathbf{Q}\Sigma_\gamma\mathbf{Q}^\top)$ , and  $\|\Sigma_\varepsilon^\dagger\|_{\mathcal{L}}$  is bounded. Then,  $\hat{\Sigma}_y^{\mathcal{F}}$  has a bounded Moore–Penrose inverse with probability approaching one, and  $\|(\hat{\Sigma}_y^{\mathcal{F}})^\dagger - \Sigma_y^\dagger\|_{\mathcal{L}} = O_p(\varpi_{n,p}^{1-q} s_p)$ .*

**Remark 3.3.** (i) The condition that  $\|\Sigma_\varepsilon^\dagger\|_{\mathcal{L}}$  and  $\|\Sigma_f^\dagger\|_{\mathcal{L}}$  are bounded implies that  $\|\Sigma_y^\dagger\|_{\mathcal{L}}$  is bounded, which means that  $\Sigma_y$  has a finite number of nonzero eigenvalues, denoted as  $d < \infty$ . Then  $\Sigma_y(u, v) = \sum_{i=1}^d \tau_i \varphi_i(u) \varphi_i(v)^\top$  with its inverse  $\Sigma_y^\dagger(u, v) = \sum_{i=1}^d \tau_i^{-1} \varphi_i(u) \varphi_i(v)^\top$ . While the inverse of  $\hat{\Sigma}_y^{\mathcal{S}}$  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  $d$  can be made arbitrarily large relative to  $n$ , e.g.,  $d = 2000, n = 200$ . Hence, this finite-dimensional assumption does not place a practical constraint on our method. See applications of inverse covariance matrix function estimation in Sections 4.1 and 4.2.

(ii) An example that satisfies  $\text{Ker}(\Sigma_\varepsilon) = \text{Ker}(\mathbf{B}\Sigma_f\mathbf{B}^\top)$  is  $\Sigma_\varepsilon(u, v) = \sum_{j=1}^d \lambda_j \psi_j(u) \psi_j(v)^\top$  and  $\Sigma_f(u, v) = \sum_{j=1}^d \omega_j \phi_j(u) \phi_j(v)^\top$ , where  $\lambda_j > 0$  and  $\omega_j > 0$  for  $j \in [d]$  and  $\text{span}\{\psi_j(\cdot)\}_{j \in [d]} = \text{span}\{\mathbf{B}\phi_j(\cdot)\}_{j \in [d]}$ . While the existing literature has not explored the convergence properties of high-dimensional inverse covariance operators, our papers makes a first attempt within the functional factor modeling framework by assuming the finite-dimensional functional objects and the same spaces spanned by the corresponding eigenfunctions. We leave the possible relaxation of these conditions as future research.

(iii) Within infinite-dimensional Hilbert space, the inverse operator  $\Sigma_y^\dagger(u, v) = \sum_{i=1}^\infty \tau_i^{-1} \varphi_i(u) \varphi_i(v)^\top$  becomes an unbounded operator, which is discontinuous and cannot be estimated in a meaningful way. However,  $\Sigma_y^\dagger$  is usually associated with another function/operator, and the composite function/operator in  $\text{Ker}(\Sigma_y)^\perp$  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 (15), which is truncated at  $d < \infty$ , i.e.,  $\Sigma_{y,d}(u, v) = \sum_{i=1}^d \tau_i \varphi_i(u) \varphi_i(v)^\top$ . 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  $\sum_{i=d+1}^\infty \tau_i^{-1} \varphi_i(u) \varphi_i(v)^\top$  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  $\Sigma_{y,d}$ , and our results in Theorems 6 and 6' 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 the intraday uncertainties. Consider a portfolio consisting of  $p$  stocks. Let  $\mathbf{P}_t(\cdot) = \{P_{t1}(\cdot), \dots, P_{tp}(\cdot)\}^\top$ ,  $\mathbf{Z}_t(\cdot) = \{Z_{t1}(\cdot), \dots, Z_{tp}(\cdot)\}^\top$  and  $\tilde{\mathbf{w}}_t(\cdot) =$

$\{\tilde{w}_{t1}(\cdot), \dots, \tilde{w}_{tp}(\cdot)\}^T$ , where  $P_{ti}(u)$ ,  $Z_{ti}(u)$  and  $\tilde{w}_{ti}(u) = Z_{ti}(u)P_{ti}(u)$  respectively denote the price per share, the quantity held, and the amount of money held for the  $i$ -th stock at time  $u$  on the  $t$ -th trading day. We assume that  $\mathcal{U} = [0, 1]$  and  $\tilde{\mathbf{w}}_t(0)^T \mathbf{1}_p = 1$ , where  $\mathbf{1}_p = (1, \dots, 1)^T \in \mathbb{R}^p$ , meaning that the initial portfolio amount is normalized to facilitate subsequent analysis. Then, the functional portfolio return of the  $t$ -th day is

$$\begin{aligned} \langle \mathbf{Z}_t, \nabla \mathbf{P}_t \rangle &= \int_0^1 \tilde{\mathbf{w}}_t(u)^T \nabla \{\log \mathbf{P}_t(u)\} du = \tilde{\mathbf{w}}_t(u)^T \log \{\mathbf{P}_t(u)\} \Big|_0^1 - \langle \nabla \tilde{\mathbf{w}}_t, \log(\mathbf{P}_t) \rangle \\ &= \tilde{\mathbf{w}}_t(1)^T \mathbf{y}_t(1) - \langle \nabla \tilde{\mathbf{w}}_t, \mathbf{y}_t \rangle, \end{aligned} \quad (23)$$

where  $\mathbf{y}_t(u) = \log \{\mathbf{P}_t(u)\} - \log \{\mathbf{P}_t(0)\}$ ,  $u \in [0, 1]$  turns to be the cumulative intraday return (CIDR) trajectory as defined in Horváth et al. (2014). As discussed in Lou et al. (2019), the close-to-close return of the stock can be decomposed into overnight and intraday components, and the profits of popular trading strategies are either earned entirely overnight or entirely intraday. Since the CIDR trajectories are only recorded during the trading period, our functional risk management focuses on the intraday strategy, where assets are bought and sold within the same day to profit from short-term price movements, so we close out all positions by the end of the trading day with  $\tilde{\mathbf{w}}_t(1) = \mathbf{0}$ . By (23), the functional portfolio return of the  $t$ -th day is  $r_t = \langle \mathbf{w}_t, \mathbf{y}_t \rangle$  with  $\mathbf{w}_t(\cdot) = -\nabla \tilde{\mathbf{w}}_t(\cdot)$ . If  $w_{ti}(u) = -\nabla \tilde{w}_{ti}(u) > 0$  (or  $< 0$ ), it indicates that an amount  $w_{ti}(u)$  of the  $i$ -th stock is sold (or bought) at time  $u$ . The amount held in each stock at the opening is determined by  $\tilde{\mathbf{w}}_t(0) = \int_0^1 \mathbf{w}_t(u) du$ . The constraint on  $\mathbf{w}_t(\cdot)$  is  $\int_0^1 \mathbf{w}_t(u)^T \mathbf{1}_p du = \tilde{\mathbf{w}}_t(0)^T \mathbf{1}_p - \tilde{\mathbf{w}}_t(1)^T \mathbf{1}_p = 1$ , and hence  $\mathbf{w}_t(\cdot)$  can be viewed as the functional portfolio allocation vector. Despite being formulated within a functional framework, the practical implementation can be achieved through discretization, such as at intervals of every 5 or 10 minutes. The functional portfolio variance of the  $t$ -th day is calculated as  $\text{Var}(r_t) = \langle \mathbf{w}_t, \Sigma_y(\mathbf{w}_t) \rangle$ . For a more general  $\tilde{\mathbf{w}}_t(1)$ , the corresponding functional portfolio variance involves the cross-covariance between two terms in (23), which largely complicates functional risk management and is left for future research.

For a given  $\mathbf{w}(\cdot)$ , the true and perceived variances (i.e., risks) of the functional portfolio are  $\langle \mathbf{w}, \Sigma_y(\mathbf{w}) \rangle$  and  $\langle \mathbf{w}, \hat{\Sigma}_y(\mathbf{w}) \rangle$ , respectively. According to Proposition S.1 of the supplementary material, the estimation error of the functional portfolio variance is bounded by

$$|\langle \mathbf{w}, \hat{\Sigma}_y(\mathbf{w}) \rangle - \langle \mathbf{w}, \Sigma_y(\mathbf{w}) \rangle| \leq \|\hat{\Sigma}_y - \Sigma_y\|_{S, \max} \left( \sum_{i=1}^p \|w_i\| \right)^2,$$

in which Theorems 5 and 5' quantify the maximum approximation error  $\|\hat{\Sigma}_y - \Sigma_y\|_{S, \max}$ .

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  $\|\mathbf{K}\|_{S, \Sigma_y} = p^{-1/2} \|(\Sigma_y^\dagger)^{1/2} \mathbf{K} (\Sigma_y^\dagger)^{1/2}\|_{S, F}$ , where  $\mathbf{K} \in \mathbb{H}^p \otimes \mathbb{H}^p$  and the normalization factor  $p^{-1/2}$  serves the role of  $\|\Sigma_y\|_{S, \Sigma_y} = 1$ . To ensure the validity of this functional norm, we assume that  $\Sigma_y$  has a bounded inverse, which does not place a constraint in practice (see Remark 3.3(i)). With such functional norm, the relative error can be measured by

$$p^{-1/2} \|(\Sigma_y^\dagger)^{1/2} \hat{\Sigma}_y (\Sigma_y^\dagger)^{1/2} - \tilde{\mathbf{I}}_p\|_{S, F} = \|\hat{\Sigma}_y - \Sigma_y\|_{S, \Sigma_y}, \quad (24)$$

where  $\tilde{\mathbf{I}}_p$  is defined as  $\tilde{\mathbf{I}}_p(\mathbf{x}) = \mathbf{x}$  for  $\mathbf{x} \in \text{Im}(\Sigma_y)$  and 0 for  $\mathbf{x} \in \text{Im}(\Sigma_y)^\perp$ . Provided that  $\|\hat{\Sigma}_y^S - \Sigma_y\|_{\mathcal{S}, \Sigma_y} = O_p(\mathcal{M}_\varepsilon \sqrt{p/n})$ , the sample covariance 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  $\hat{\Sigma}_y^D$  converges to  $\Sigma_y$  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 7.** *Under the assumptions of Theorem 6, we have  $\|\hat{\Sigma}_y^D - \Sigma_y\|_{\mathcal{S}, \Sigma_y} = O_p(\mathcal{M}_\varepsilon^2 p^{1/2} n^{-1} + \varpi_{n,p}^{1-q} s_p)$ .*

By Proposition S.2 of the supplementary material, the relative error is bounded by

$$|\langle \mathbf{w}, \hat{\Sigma}_y(\mathbf{w}) \rangle / \langle \mathbf{w}, \Sigma_y(\mathbf{w}) \rangle - 1| \leq \|(\Sigma_y^\dagger)^{1/2} \hat{\Sigma}_y (\Sigma_y^\dagger)^{1/2} - \tilde{\mathbf{I}}_p\|_{\mathcal{L}},$$

which, in conjunction with Theorem 7 and (24), controls the maximum relative error.

## 4.2 Estimation of regression and discriminant direction functions

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

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

where  $\beta(\cdot, \cdot) = \{\beta_1(\cdot, \cdot), \dots, \beta_p(\cdot, \cdot)\}^\top$  is an operator-valued coefficient vector to be estimated. We can impose certain smoothness condition such that  $\beta(u, v) = \sum_{i=1}^\infty \tilde{\tau}_i \varphi_i(u) \varphi_i(v)^\top$  is sufficiently smooth relative to  $\Sigma_y(u, v) = \sum_{i=1}^\infty \tau_i \varphi_i(u) \varphi_i(v)^\top$ , ensuring the boundedness of the regression operator  $\beta(u, v) = \int_{\mathcal{U}} \Sigma_y^\dagger(u, u') \text{Cov}\{\mathbf{y}_t(u'), z_t(v)\} du'$ . Replacing relevant terms by their (truncated) sample versions, we obtain  $\hat{\beta}(u, v) = n^{-1} \sum_{t=1}^n \int_{\mathcal{U}} \hat{\Sigma}_{y,d}^\dagger(u, u') \mathbf{y}_t(u') z_t(v) du'$ . This application highlights the need for estimators  $\hat{\Sigma}_{y,d}^\dagger$ , as studied in Theorems 6 and 6'.

The third application delves into linear discriminant analysis for classifying multivariate functional data (Xue et al., 2024) 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  $\mu_1(\cdot)$  and  $\mu_2(\cdot)$ , respectively, while sharing a common covariance matrix function  $\Sigma_y$ . Our goal is to determine the linear classifier by estimating the discriminant direction function  $\int_{\mathcal{U}} \Sigma_y^\dagger(u, v) \{\mu_1(v) - \mu_2(v)\} dv$ , which takes the same form as the regression function  $\beta(u) = \int_{\mathcal{U}} \Sigma_y^\dagger(u, v) \text{Cov}\{\mathbf{y}_t(v), z_t\} dv$  encountered in the second application with a scalar response  $z_t$ . By similar arguments as above, both applications call for the use of estimators  $\hat{\Sigma}_{y,d}^\dagger$ .

## 4.3 Estimation of correlation matrix function

The fourth application involves estimating the correlation matrix function and its inverse, which are essential in graphical models for truly infinite-dimensional objects, see, e.g., Solea and Li (2022). Our proposed covariance estimators can be employed to estimate the corresponding correlation matrix function and its inverse. Specifically, let  $\mathbf{D}_y(\cdot, \cdot) = \text{diag}\{\Sigma_{y,11}(\cdot, \cdot), \dots, \Sigma_{y,pp}(\cdot, \cdot)\}$  be the  $p \times p$  diagonal matrix function. According to Baker (1973), there exists a correlation matrix function  $\mathbf{C}_y$  with  $\|\mathbf{C}_y\|_{\mathcal{L}} \leq 1$  such that  $\Sigma_y = \mathbf{D}_y^{1/2} \mathbf{C}_y \mathbf{D}_y^{1/2}$ . Under certain compactness and smoothness assumptions,  $\mathbf{C}_y$  has a bounded inverse, denoted by  $\Theta_y$ , 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). Although the inverse of the estimator  $\hat{\mathbf{D}}_y = \text{diag}(\hat{\Sigma}_{y,11}, \dots, \hat{\Sigma}_{y,pp})$  is not well-defined, we adopt the Tikhonov regularization to estimate  $\mathbf{C}_y$  by  $\hat{\mathbf{C}}_y^{(\kappa)} = (\hat{\mathbf{D}}_y + \kappa \mathbf{I}_p)^{-1/2} \hat{\Sigma}_y (\hat{\mathbf{D}}_y + \kappa \mathbf{I}_p)^{-1/2}$  for some regularization parameter  $\kappa > 0$ . The

estimator of  $\Theta_y$  is then given by  $\hat{\Theta}_y^{(\kappa)} = \hat{\mathbf{D}}_y^{1/2}(\hat{\Sigma}_y + \kappa \mathbf{I}_p)^{-1} \hat{\mathbf{D}}_y^{1/2}$ . Consequently, we can plug into the DIGIT or the FPOET estimator for estimating  $\mathbf{C}_y$  and its inverse  $\Theta_y$ .

## 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  $\text{Uniform}[-0.75, 0.75]$ , satisfying Assumption 3(i). To mimic the infinite-dimensionality of functional data, each functional factor is generated from  $f_{tj}(\cdot) = \sum_{i=1}^{50} i^{-1} \xi_{tji} \phi_i(\cdot)$  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  $\xi_{ti} = (\xi_{t1i}, \dots, \xi_{t50i})^T$  are generated from a vector autoregressive model,  $\xi_{ti} = \mathbf{A} \xi_{t-1,i} + \mathbf{u}_{ti}$  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  $\mathcal{N}(\mathbf{0}_r, \mathbf{I}_r)$ . For the second data-generating process (denoted as DGP2), we generate observed data from model (2), where  $r$ -vector of scalar factors  $\gamma_t$  is generated from a vector autoregressive model,  $\gamma_t = \mathbf{A} \gamma_{t-1} + \mathbf{u}_t$  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 from  $\varepsilon_t(\cdot) = \sum_{l=1}^{25} 2^{-l/2} \psi_{tl} \phi_l(\cdot)$ , where each  $\psi_{tl}$  is generated from  $\psi_{tl} = 0.5 \psi_{t-1,l} + \zeta_{tl}$  with  $\zeta_{tl}$  being independently sampled from  $\mathcal{N}(\mathbf{0}_p, \mathbf{C}_\zeta)$  with  $\mathbf{C}_\zeta = \mathbf{D} \mathbf{C}_0 \mathbf{D}$ . Given this autoregressive structure, it can be shown that  $\mathcal{M}_\varepsilon = O(1)$ . Here, we set  $\mathbf{D} = \text{diag}(D_1, \dots, D_p)$ , where each  $D_i$  is generated from  $\text{Gamma}(3, 1)$ . The generation of  $\mathbf{C}_0$  involves the following three steps: (i) we set the diagonal entries of  $\check{\mathbf{C}}$  to 1, and generate the off-diagonal and symmetrical entries from  $\text{Uniform}[0, 0.5]$ ; (ii) we employ hard thresholding (Cai and Liu, 2011) on  $\check{\mathbf{C}}$  to obtain a sparse matrix  $\check{\mathbf{C}}^\tau$ , where the threshold level is found as the smallest value such that  $\max_{i \in [p]} \sum_{j=1}^p I(\check{C}_{ij}^\tau \neq 0) \leq p^{1-\alpha}$  for  $\alpha \in [0, 1]$ ; (iii) we set  $\mathbf{C}_0 = \check{\mathbf{C}}^\tau + \delta \mathbf{I}_p$  where  $\delta = \max\{-\lambda_{\min}(\check{\mathbf{C}}), 0\} + 0.01$  to guarantee the positive-definiteness of  $\mathbf{C}_0$ . The parameter  $\alpha$  controls the sparsity level with larger values yielding sparser structures in  $\mathbf{C}_0$  as well as functional sparser patterns in  $\Sigma_\varepsilon(\cdot, \cdot)$ . 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.

We firstly assess the finite-sample performance of the proposed information criteria in Section 2.4 under different combinations of  $p, n$  and  $\alpha$  for DGP1 and DGP2. Figure 1 presents boxplots of  $\Delta \text{IC}_i$  ( $i = 1, 2, 3$ ) for two DGPs under the setting  $p = 100$ ,  $n = 50$ ,  $\alpha = 0.25, 0.5$  and  $r = 3, 5, 7$ . See also similar results for  $p = 100$ ,  $n = 50$ ,  $\alpha = 0.05, 0.1$  in Figure S.1 and the corresponding model selection accuracies in Table S.1 of the supplementary material. We observe a few trends. First, the proposed criteria can lead to high model selection accuracies in all cases. Second, larger values of  $\alpha$  lead to improved model selection accuracy, as they correspond to a higher relative strength of the common components over idiosyncratic components, quantified by  $\|\Sigma_\chi\|_{\mathcal{S}, \text{F}} / \|\Sigma_\varepsilon\|_{\mathcal{S}, \text{F}} \gtrsim p^{\alpha/2}$  for DGP 1 and  $\|\Sigma_\kappa\|_{\mathcal{S}, \text{F}} / \|\Sigma_\varepsilon\|_{\mathcal{S}, \text{F}} \gtrsim p^{\alpha/2}$  for DGP 2. Third, different penalty functions  $g(n, p)$  have similar impacts on the information criteria when  $p$  and  $n$  are relatively large.

Once the more appropriate FFM is selected based on observed data, our next step adopts the ratio-based estimator (19) (or (20)) 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. When implementing (19) and (20), we choose  $c_r = 0.1$  and  $r_{1,0} = r_{2,0} = 20$ . Additional

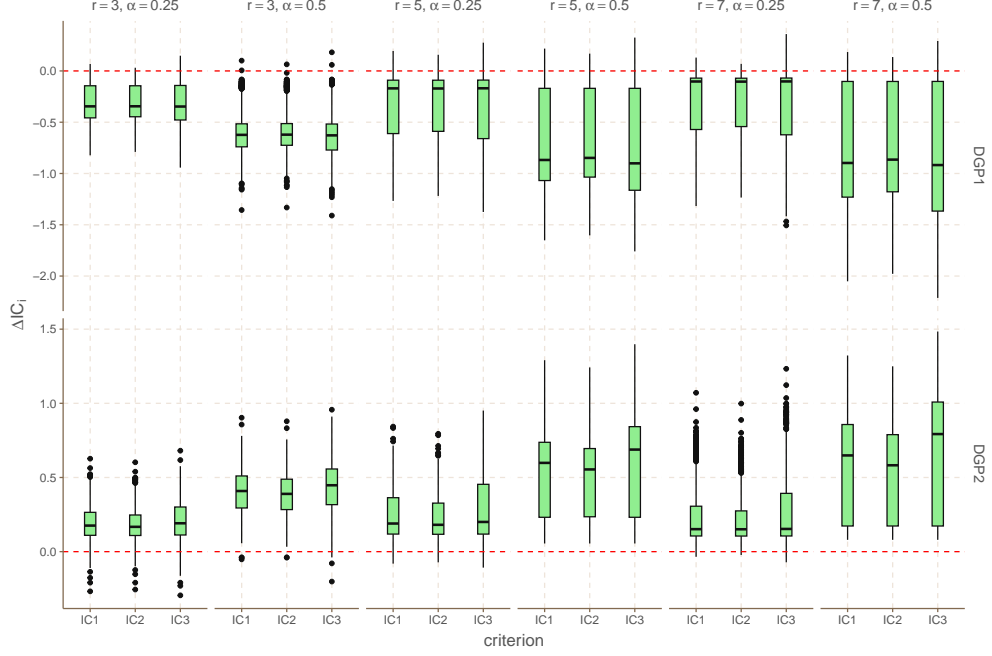


Figure 1: The boxplots of  $\Delta IC_i$  ( $i \in [3]$ ) for DGP1 and DGP2 with  $p = 100, n = 50, \alpha = 0.25, 0.5$  and  $r = 3, 5, 7$  over 1000 simulation runs.

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

$\alpha$	$p$	$n$	$r = 3$		$r = 5$		$r = 7$	
			$\mathbb{P}(\hat{r}^D = r)$	$\mathbb{P}(\hat{r}^F = r)$	$\mathbb{P}(\hat{r}^D = r)$	$\mathbb{P}(\hat{r}^F = r)$	$\mathbb{P}(\hat{r}^D = r)$	$\mathbb{P}(\hat{r}^F = r)$
0.25	100	100	0.674	0.632	0.468	0.449	0.331	0.355
		200	0.757	0.726	0.622	0.571	0.466	0.483
	200	100	0.757	0.647	0.563	0.430	0.345	0.319
		200	0.846	0.797	0.717	0.668	0.560	0.521
0.50	100	100	0.881	0.893	0.815	0.819	0.640	0.743
		200	0.936	0.921	0.891	0.908	0.805	0.828
	200	100	0.977	0.953	0.910	0.911	0.778	0.836
		200	0.970	0.962	0.954	0.958	0.898	0.928
0.75	100	100	0.974	0.983	0.947	0.963	0.899	0.929
		200	0.979	0.985	0.975	0.978	0.951	0.971
	200	100	0.999	0.997	0.989	0.998	0.945	0.980
		200	0.997	0.999	0.998	1.000	0.992	0.997

simulations suggest that the results are not sensitive to the choice of  $r_{1,0}$  and  $r_{2,0}$ , and a small value of  $c_r$  leads to good performance. 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  $\alpha$  enhance the accuracy of identifying  $r$ . 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.

We next compare our proposed AFT estimator in (10) with two related methods for estimating the idiosyncratic covariance  $\Sigma_\varepsilon$ , where the details can be found in Section S.6 of the supplementary material. Following Fan et al. (2013), the threshold level under hard thresholding for AFT is selected as  $\lambda = \dot{C}(\sqrt{\log p/n} + 1/\sqrt{p})$  with  $\dot{C} = 0.5$ . To select the optimal  $\dot{C}$ , we also implemented a cross-validation method over a candidate set, whose lower bound was determined in a way similar to (4.1) of Fan et al. (2013) to ensure the positive definiteness of the AFT estimators. However, such method incurred heavy

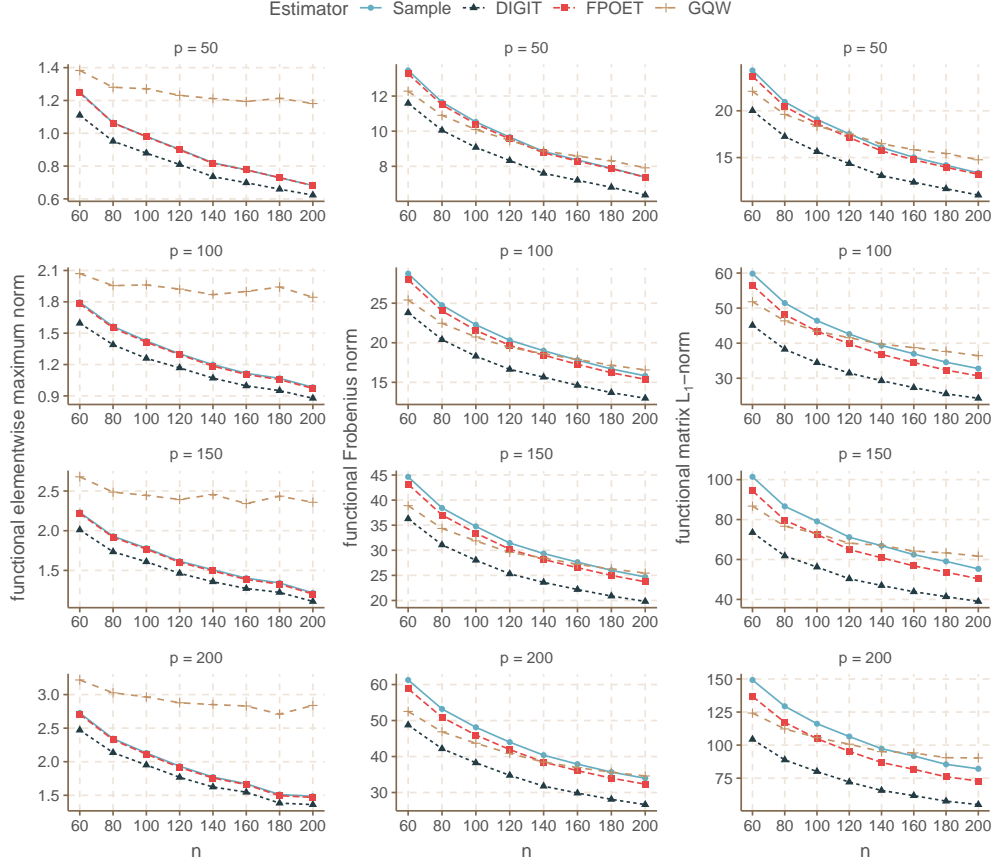


Figure 2: The average losses of  $\hat{\Sigma}_y$  in functional elementwise  $\ell_\infty$  norm (left column), Frobenius norm (middle column) and matrix  $\ell_1$  norm (right column) for DGP1 over 1000 simulation runs.

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  $\Sigma_y$ . The first competitor is the sample covariance estimator  $\hat{\Sigma}_y^S$ . For comparison, we also implement the method of Guo et al. (2025) in conjunction with our AFT (denoted as GQW). This combined method employs autocovariance-based eigenanalysis to estimate  $\mathbf{B}$  and then follows the similar procedure as DIGIT to estimate  $\mathbf{f}_t(\cdot)$  and  $\Sigma_\varepsilon$ . 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  $\Sigma_y$  under DGP2 (or DGP1) to evaluate the robustness of each proposal under model misspecifications. For both DGPs, we set  $\alpha = 0.5$  and generate  $n = 60, 80, \dots, 200$  observations of  $p = 50, 100, 150, 200$  functional variables. Figures 2 and 3 display the numerical summaries of losses measured by functional versions of elementwise  $\ell_\infty$  norm, Frobenius norm, and matrix  $\ell_1$  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. In high-dimensional large  $p$  scenarios, the factor-guided estimators lead to more competitive performance, whereas the results of  $\hat{\Sigma}_y^S$  severely deteriorate especially in terms of functional matrix  $\ell_1$  loss. Second, although both DIGIT and GQW estimators are developed to estimate model (1), our proposed DIGIT estimator is prominently superior to the GQW estimator for DGP1 under all scenarios, as seen in Figure 2. Third, the FPOET estimator exhibits enhanced robustness compared to DIGIT



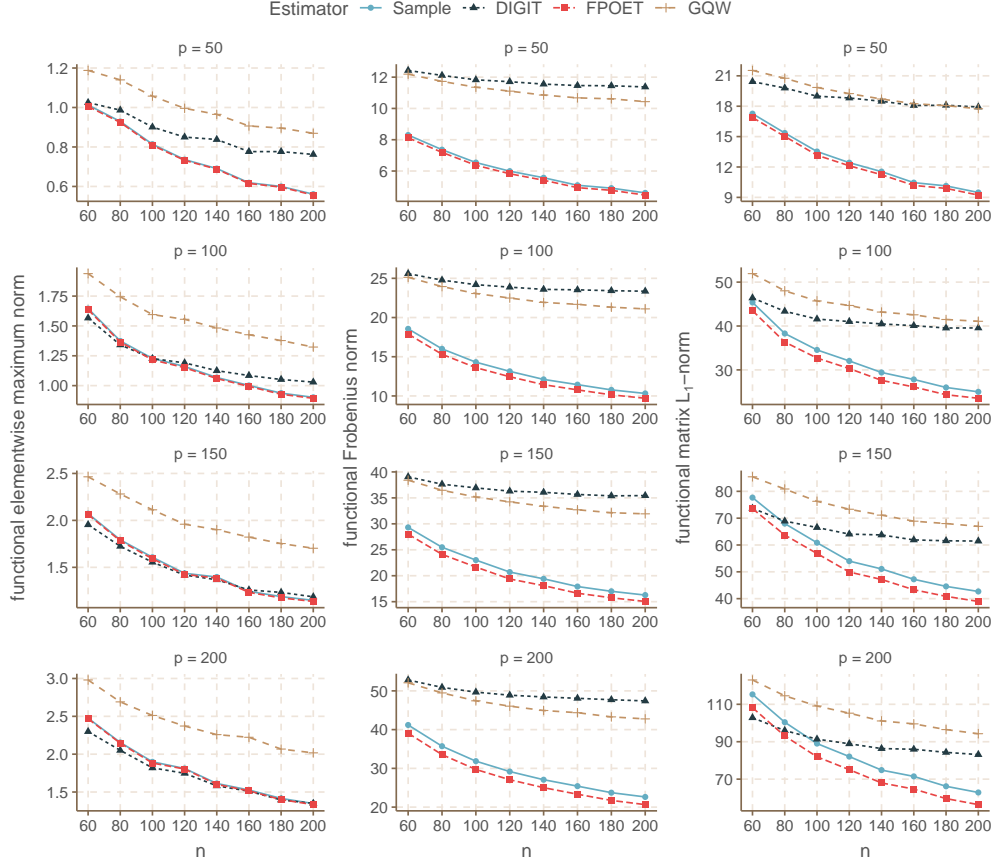


Figure 3: The average losses of  $\hat{\Sigma}_y$  in functional elementwise  $\ell_\infty$  norm (left column), Frobenius norm (middle column) and matrix  $\ell_1$  norm (right column) for DGP2 over 1000 simulation runs.

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  $\ell_1$  losses, while, for DGP1, FPOET still achieves reasonably good performance.

## 6 Real data analysis

Our 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  $\mathcal{U} = [0, 1]$ . We construct CIDR (Horváth et al., 2014) trajectories, in percentage, by  $y_{ti}(u_k) = 100[\log\{P_{ti}(u_k)\} - \log\{P_{ti}(u_0)\}]$ , 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 ( $u_k = k/78$ ) 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,  $IC_1^{\mathcal{P}} = -0.632 < IC_1^{\mathcal{F}} = -0.614$ , suggests that FFM (1)



is slightly more preferable and implies that the latent factors may exhibit intraday varying patterns. We consider the problem of functional risk management in Section 4.1. Our task is to obtain the optimal functional portfolio allocation  $\hat{\mathbf{w}}(\cdot)$  by minimizing the perceived risk of the functional portfolio, specifically,

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{H}^p} \langle \mathbf{w}, \hat{\Sigma}_y(\mathbf{w}) \rangle \quad \text{subject to} \quad \int_{\mathcal{U}} \mathbf{w}(u)^T \mathbf{1}_p du = 1.$$

Following the derivations in Section S.5.4 of the supplementary material, we obtain the solution:

$$\hat{\mathbf{w}}(u) = \frac{\int_{\mathcal{U}} \hat{\Sigma}_y^\dagger(u, v) \mathbf{1}_p dv}{\int_{\mathcal{U}} \int_{\mathcal{U}} \mathbf{1}_p^T \hat{\Sigma}_y^\dagger(z, v) \mathbf{1}_p dz dv}, \quad u \in \mathcal{U}, \quad (25)$$

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  $\hat{\Sigma}_y^\dagger$  using the leading eigenpairs of  $\hat{\Sigma}_y$  such that the corresponding cumulative percentage of selected eigenvalues exceeds 95%.

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

Estimator	$\hat{r}$	July	August	September	October	November	December	Average
DIGIT	1	0.0064	0.0127	0.0050	0.0154	0.0081	0.0994	0.0245
	3	0.0029	0.0082	0.0032	0.0054	0.0094	0.0080	0.0062
	5	0.0057	0.0136	0.0077	0.0075	0.0137	0.0149	0.0105
	7	0.0081	0.0105	0.0058	0.0112	0.0096	0.0142	0.0099
	9	0.0105	0.0124	0.0067	0.0095	0.0070	0.0201	0.0110
	11	0.0130	0.0129	0.0069	0.0077	0.0085	0.0257	0.0125
FPOET	1	0.0350	0.0136	0.0109	0.0441	0.0378	0.0174	0.0265
	3	0.0227	0.0207	0.0150	0.0224	0.0280	0.0524	0.0269
	5	0.0154	0.0139	0.0222	0.0349	0.0275	0.0199	0.0223
	7	0.0142	0.0162	0.0142	0.0108	0.0283	0.0224	0.0177
	9	0.0428	0.0180	0.0215	0.0201	0.0306	0.0194	0.0254
	11	0.0562	0.0224	0.0129	0.0294	0.0342	0.0348	0.0316
GQW	1	0.0063	0.0249	0.2104	1.6762	0.0047	0.4441	0.3944
	3	0.0036	0.0092	0.0045	0.0231	0.0093	0.0105	0.0100
	5	0.0062	0.0114	0.0061	0.0063	0.0117	0.0074	0.0082
	7	0.0069	0.0152	0.0064	0.0089	0.0081	0.0155	0.0102
	9	0.0122	0.0111	0.0063	0.0090	0.0114	0.0120	0.0103
	11	0.0140	0.0157	0.0087	0.0072	0.0103	0.0251	0.0135
POET	1	0.0271	0.0272	0.0352	0.0197	0.0242	0.0389	0.0287
	3	0.0270	0.0342	0.0463	0.0267	0.0257	0.0405	0.0334
	5	0.0185	0.0269	0.0346	0.0229	0.0273	0.0430	0.0289
	7	0.0214	0.0275	0.0373	0.0226	0.0243	0.0375	0.0284
	9	0.0206	0.0250	0.0411	0.0241	0.0209	0.0342	0.0277
	11	0.0178	0.0271	0.0407	0.0223	0.0229	0.0366	0.0279
Sample		0.0203	0.0290	0.0372	0.0254	0.0267	0.0310	0.0282

Following the procedure in Fan et al. (2013), on the 1st trading day of each month from July to December, we estimate  $\hat{\Sigma}_y$  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  $\hat{\mathbf{w}}(u_k)$  for  $k \in [78]$ . To illustrate the superiority of functional analytic methods, we also introduce a non-functional competing method based on the POET estimator, whose portfolio construction procedure is detailed in Section S.5.5 of the supplementary material. At the end of the month after 21 trading days, we compare actual risks calculated

by  $78^{-2} \sum_{k,k' \in [78]} \hat{\mathbf{w}}(u_k)^\top \{21^{-1} \sum_{t=1}^{21} \mathbf{y}_t(u_k) \mathbf{y}_t(u_{k'})^\top\} \hat{\mathbf{w}}(u_{k'})$ . Following [Fan et al. \(2013\)](#) and [Wang et al. \(2021\)](#), we try  $\hat{r} = 1, 3, 5, 7, 9$  and 11 to check the effect of  $r$  in out-of-sample performance. The numerical results are summarized in Table 2. Among the functional analytic methods, we observe that the optimal functional portfolio allocation created by DIGIT, FPOET, and GQW result in minimum averaged risks over six months as 0.0062, 0.0177, and 0.0082, respectively, while the sample covariance estimator gives 0.0282. The risk has been significantly reduced by at least 37% using our factor-guided approach. Additionally, the POET-based method yields a minimum averaged risk of 0.0277, providing empirical evidence for the advantage of functional analytic methods.

## 7 Discussions

Our theoretical results are established under a sub-Gaussian condition, which is imposed to facilitate the use of Hanson–Wright-type concentration inequalities for time series within Hilbert space ([Chang et al., 2024](#)) in our non-asymptotic analysis. To the best of our knowledge, the existing literature on concentration inequalities for high-dimensional functional time series are all of Hanson–Wright-type. It is thus of interest to relax such sub-Gaussian condition to a weaker finite moment condition beyond functional linear process, and based on which developing more generalized Nagaev-type concentration inequalities for high-dimensional time series ([Zhang and Wu, 2021](#)) within Hilbert space to aid our theoretical analysis. This relaxation will result in allowing the dimension  $p$  to grow polynomially rather than exponentially with  $n$ . To achieve this, we also need to propose a new functional dependence measure instead of our functional stability measure defined in (22) to characterize how the established concentration results are affected by the complex serial dependence structure.

It is interesting to conduct specification tests such as testing one FFM against another and testing the constancy of the factor or loading functions. Both tests rely on the inferential theory for FFMs, which requires to explore the limiting distributions of estimated quantities. The existing literature on FFMs only studies the estimation of factors, loadings and number of factors, without delving into the corresponding limiting distributions. Note that the inferential theory for FFM based on MFPCA can be developed for FFM (2) given its equivalence to the least squares method. By comparison, the inferential theory for the eigenanalysis of doubly integrated Gram covariance specific to FFM (1) presents significant challenges. Moreover, deriving the limiting distributions under functional domain involves characterizing the magnitude of functional quantities using a suitable functional norm such as  $L_2$  or supremum norm, presenting additional complexities compared to scalar time series.

While our paper focuses on fully observed functional time series, it is also interesting to consider the common practical scenario, where each curve  $y_{tj}(\cdot)$  is only partially observed, with errors, at  $T_{tj}$  random time points. For densely observed functional time series with  $T_{tj}$ 's being larger than some order of  $n$ , it is customary to apply nonparametric smoothing to the observations from each curve ([Zhang and Chen, 2007](#)), which results in reconstructed curves  $\hat{y}_{tj}(\cdot)$ 's serving as new inputs for subsequent analysis. When  $T_{tj}$ 's are bounded, referred to as sparsely observed functional time series, the pre-smoothing step becomes inapplicable. An alternative approach involves local linear surface smoothing by pooling observations together from all curves ([Chen et al., 2022](#)). This method yields smoothed estimates of  $\Sigma_{y,jk}(u, v)$ 's that can be utilized in our methodological development.

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

## Acknowledgements

We are grateful to the editor, the associate editor and two referees for their insightful and valuable comments and suggestions, which have led to significant improvement of our paper. Li's research is partially supported by the NSFC (no.72471127). Qiao's research is partially supported by the Seed Fund for Basic Research for New Staff at the University of Hong Kong.

## Appendix

This appendix contains the technical proofs of the results in Section 2 and the results for FFM (1) in Section 3, i.e., Theorems 1–6 and Corollary 1. For the technical proofs of the results for FFM (2) in Section 3 and the results in Section 4, see the supplementary material, which also provides the proofs of some technical lemmas in the appendix, some further derivations, additional simulation results and real data results. Throughout, we denote the multiplications of matrix kernel functions as  $\mathbf{M} = \mathbf{K}\mathbf{G} \in \mathbb{H}^p \otimes \mathbb{H}^p$  for  $\mathbf{K}, \mathbf{G} \in \mathbb{H}^p \otimes \mathbb{H}^p$ , where  $\mathbf{M}(u, v) = \int_{\mathcal{U}} \mathbf{K}(u, w)\mathbf{G}(w, v)dw$ .

## A Proofs of theoretical results in Section 2

### A.1 Technical lemmas

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

**Lemma A.1.** (Weyl's theorem; [Weyl \(1912\)](#)).  $|\tilde{\lambda}_j - \lambda_j| \leq \|\tilde{\Sigma} - \Sigma\|$  for  $j \in [p]$ .

**Lemma A.2.** (A useful variant of  $\sin(\theta)$  theorem; [Yu et al. \(2015\)](#)). If  $\tilde{\xi}_j^T \xi_j \geq 0$  for  $j \in [p]$ , then

$$\|\tilde{\xi}_j - \xi_j\| \leq \frac{\|\tilde{\Sigma} - \Sigma\|/\sqrt{2}}{\min(|\tilde{\lambda}_{j-1} - \lambda_j|, |\lambda_j - \tilde{\lambda}_{j+1}|)}.$$

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  $\Sigma(\cdot, \cdot)$  in a descending order and  $\{\varphi_i(\cdot)\}_{i=1}^\infty$  are the corresponding eigenfunctions. Similarly,  $\{\tilde{\tau}_i\}_{i=1}^\infty$  and  $\{\tilde{\varphi}_i(\cdot)\}_{i=1}^\infty$  are the corresponding eigenvalues and eigenfunctions of  $\tilde{\Sigma}(\cdot, \cdot)$ , respectively.

**Lemma A.3.** (Lemma 4.2 in [Bosq \(2000\)](#)).  $|\tilde{\tau}_i - \tau_i| \leq \|\tilde{\Sigma} - \Sigma\|_{\mathcal{L}}$  for all  $i$ .

**Lemma A.4.** (Lemma 4.3 in [Bosq \(2000\)](#)). If  $\langle \tilde{\varphi}_i, \varphi_i \rangle \geq 0$ , then

$$\|\tilde{\varphi}_i - \varphi_i\| \leq \frac{2\sqrt{2}\|\tilde{\Sigma} - \Sigma\|_{\mathcal{L}}}{\min(|\tilde{\tau}_{i-1} - \tau_i|, |\tau_i - \tilde{\tau}_{i+1}|)}.$$

The following lemmas introduce some useful functional norm inequalities. Their proofs are relegated to the supplementary material.

**Lemma A.5.** Suppose that  $\mathbf{K} \in \mathbb{H}^p \otimes \mathbb{H}^p$  is a Mercer's kernel with the spectral decomposition  $\mathbf{K}(u, v) = \sum_{i=1}^\infty \lambda_i \phi_i(u)\phi_i(v)^T$ , where  $\{\lambda_i\}_{i=1}^\infty$  are the eigenvalues of  $\mathbf{K}$  in a descending order and  $\{\phi_i(\cdot)\}$  are the

corresponding eigenfunctions. Then, we have

- (i)  $\text{tr} \left\{ \iint \mathbf{K}(u, v) \mathbf{K}(u, v)^T \text{d}u \text{d}v \right\} = \text{tr} \left\{ \iint \mathbf{K} \mathbf{K}^T(u, u) \text{d}u \right\} = \|\mathbf{K}\|_{\mathcal{S}, \mathcal{F}}^2 = \sum_{i=1}^{\infty} \lambda_i^2;$
- (ii)  $\left\| \iint \mathbf{K}(u, v) \mathbf{K}(u, v)^T \text{d}u \text{d}v \right\| = \left\| \iint \mathbf{K} \mathbf{K}^T(u, u) \text{d}u \right\| = \|\mathbf{K}\|_{\mathcal{L}}^2 = \lambda_1^2;$
- (iii)  $\text{tr} \left\{ \iint \mathbf{K}(u, u) \text{d}u \right\} = \|\mathbf{K}\|_{\mathcal{N}} = \sum_{i=1}^{\infty} \lambda_i.$

**Lemma A.6.** Suppose that  $\mathbf{K}, \mathbf{G} \in \mathbb{H}^p \otimes \mathbb{H}^p$  are Mercer's kernels, then we have

- (i)  $\left\| \iint \mathbf{K}(u, v) \mathbf{G}(u, v)^T \text{d}u \text{d}v \right\|_1 \leq \|\mathbf{K}\|_{\mathcal{S}, 1} \|\mathbf{G}\|_{\mathcal{S}, \infty};$
- (ii)  $\left\| \iint \mathbf{K}(u, v) \mathbf{G}(u, v)^T \text{d}u \text{d}v \right\|_{\infty} \leq \|\mathbf{K}\|_{\mathcal{S}, \infty} \|\mathbf{G}\|_{\mathcal{S}, 1};$
- (iii)  $\left\| \iint \mathbf{K}(u, v) \mathbf{G}(u, v)^T \text{d}u \text{d}v \right\| \leq (\|\mathbf{K}\|_{\mathcal{S}, \infty} \|\mathbf{K}\|_{\mathcal{S}, 1})^{1/2} (\|\mathbf{G}\|_{\mathcal{S}, \infty} \|\mathbf{G}\|_{\mathcal{S}, 1})^{1/2};$
- (iv)  $\left\| \iint \mathbf{K}(u, v) \mathbf{G}(u, v)^T \text{d}u \text{d}v \right\| \leq \left\{ \left\| \iint \mathbf{K}(u, v) \mathbf{K}(u, v)^T \text{d}u \text{d}v \right\| \right\}^{1/2} \left\{ \left\| \iint \mathbf{G}(u, v) \mathbf{G}(u, v)^T \text{d}u \text{d}v \right\| \right\}^{1/2}.$

**Lemma A.7.** Suppose that  $\Sigma = \{\Sigma_{ij}(\cdot, \cdot)\}_{p \times p}$  with  $\Sigma_{ij} \in \mathbb{S}$  and  $\tilde{\Sigma} \in \mathbb{H}^p \otimes \mathbb{H}^p$  are Mercer's kernels. Then we have (i)  $\|\Sigma \tilde{\Sigma}\|_{\mathcal{L}} \leq \|\Sigma\|_{\mathcal{L}} \cdot \|\tilde{\Sigma}\|_{\mathcal{L}},$  (ii)  $\|\Sigma\|_{\mathcal{L}} \leq \|\Sigma\|_{\mathcal{S}, \mathcal{F}},$  and (iii)  $\|\Sigma\|_{\mathcal{L}} \leq \|\Sigma\|_{\mathcal{S}, 1}^{1/2} \|\Sigma\|_{\mathcal{S}, \infty}^{1/2}.$  Furthermore, if  $\|\Sigma_{ij}\|_{\mathcal{S}} = \|\Sigma_{ji}\|_{\mathcal{S}}$  for all  $i, j \in [p],$  then  $\|\Sigma\|_{\mathcal{L}} \leq \|\Sigma\|_{\mathcal{S}, 1}.$

**Lemma A.8.** Suppose that  $\mathbf{K}, \mathbf{G} \in \mathbb{H}^p \otimes \mathbb{H}^p$  are Mercer's kernels, then we have

- (i)  $\text{tr} \left\{ \iint \mathbf{K} \mathbf{G}(u, u) \text{d}u \right\} = \text{tr} \left\{ \iint \mathbf{G} \mathbf{K}(u, u) \text{d}u \right\},$  i.e.,  $\|\mathbf{K} \mathbf{G}\|_{\mathcal{N}} = \|\mathbf{G} \mathbf{K}\|_{\mathcal{N}};$
- (ii)  $\text{tr} \left\{ \iint \mathbf{K} \mathbf{G}(u, u) \text{d}u \right\} \leq \|\mathbf{K}\|_{\mathcal{L}} \text{tr} \left\{ \iint \mathbf{G}(u, u) \text{d}u \right\},$  i.e.,  $\|\mathbf{K} \mathbf{G}\|_{\mathcal{N}} \leq \|\mathbf{K}\|_{\mathcal{L}} \|\mathbf{G}\|_{\mathcal{N}};$
- (iii)  $\|\mathbf{K} \mathbf{G}\|_{\mathcal{S}, \mathcal{F}} \leq \|\mathbf{K}\|_{\mathcal{L}} \|\mathbf{G}\|_{\mathcal{S}, \mathcal{F}}.$

**Lemma A.9.** 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)  $\|\mathbf{A} \mathbf{K}\|_{\mathcal{S}, \max} \leq \|\mathbf{A}\|_{\infty} \|\mathbf{K}\|_{\mathcal{S}, \max},$  and  $\|\mathbf{K} \mathbf{A}^T\|_{\mathcal{S}, \max} \leq \|\mathbf{K}\|_{\mathcal{S}, \max} \|\mathbf{A}^T\|_1 = \|\mathbf{A}\|_{\infty} \|\mathbf{K}\|_{\mathcal{S}, \max};$
- (ii)  $\|\mathbf{A} \mathbf{K}\|_{\mathcal{S}, \mathcal{F}} \leq \|\mathbf{A}\|_{\mathcal{F}} \|\mathbf{K}\|_{\mathcal{S}, \mathcal{F}},$  and  $\|\mathbf{K} \mathbf{A}^T\|_{\mathcal{S}, \mathcal{F}} \leq \|\mathbf{K}\|_{\mathcal{S}, \mathcal{F}} \|\mathbf{A}^T\|_{\mathcal{F}} = \|\mathbf{A}\|_{\mathcal{F}} \|\mathbf{K}\|_{\mathcal{S}, \mathcal{F}};$
- (iii)  $\|\mathbf{A} \mathbf{K}\|_{\mathcal{S}, \infty} \leq \|\mathbf{A}\|_{\infty} \|\mathbf{K}\|_{\mathcal{S}, \infty},$  and  $\|\mathbf{K} \mathbf{A}^T\|_{\mathcal{S}, \infty} \leq \|\mathbf{K}\|_{\mathcal{S}, \infty} \|\mathbf{A}^T\|_{\infty} = \|\mathbf{A}\|_1 \|\mathbf{K}\|_{\mathcal{S}, \infty};$
- (iv)  $\|\mathbf{A} \mathbf{K}\|_{\mathcal{S}, 1} \leq \|\mathbf{A}\|_1 \|\mathbf{K}\|_{\mathcal{S}, 1},$  and  $\|\mathbf{K} \mathbf{A}^T\|_{\mathcal{S}, 1} \leq \|\mathbf{K}\|_{\mathcal{S}, 1} \|\mathbf{A}^T\|_1 = \|\mathbf{A}\|_{\infty} \|\mathbf{K}\|_{\mathcal{S}, 1}.$

**Lemma A.10.** For  $\mathbf{f}, \mathbf{g} \in \mathbb{H}^r,$  and  $\mathbf{A} \in \mathbb{R}^{p \times r},$  we have

- (i)  $\|\mathbf{A} \mathbf{f}\| \leq \|\mathbf{A}\| \cdot \|\mathbf{f}\|;$
- (ii)  $\|K\|_{\mathcal{S}} \leq \|\mathbf{f}\| \cdot \|\mathbf{g}\|$  where  $K(\cdot, \cdot) \in \mathbb{S}$  is defined as  $K(u, v) = \mathbf{f}(u)^T \mathbf{g}(v).$

## A.2 Proof of Proposition 1

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

$$|\lambda_j - p^2 \theta_j| \leq \|\Omega - \Omega_{\mathcal{L}}\| = \|\Omega_{\mathcal{R}}\|,$$

and for  $r + 1 \leq j \leq p,$   $|\lambda_j| = |\lambda_j - 0| \leq \|\Omega_{\mathcal{R}}\|.$

(ii) By Lemma A.2, for  $j \in [r]$  and  $\xi_j^T \tilde{\mathbf{b}}_j \geq 0,$

$$\|\xi_j - \tilde{\mathbf{b}}_j\| \leq \frac{\|\Omega_{\mathcal{R}}\|/\sqrt{2}}{\min(|\lambda_{j-1} - p^2 \theta_j|, |p^2 \theta_j - \lambda_{j+1}|)}.$$

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  $|\lambda_{j-1} - p^2 \theta_{j-1}| \leq \|\Omega_{\mathcal{R}}\| = o(p^2)$  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  $|\lambda_{r+1}| \leq \|\Omega_{\mathcal{R}}\| = o(p^2)$  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  $\xi_j^\top \tilde{\mathbf{b}}_j \geq 0$ , we have

$$\|\xi_j - \tilde{\mathbf{b}}_j\| = O(p^{-2}\|\Omega_{\mathcal{R}}\|), \text{ for } j \in [r].$$

### A.3 Proof of Proposition 2

To prove Proposition 2, we first present a technical lemma with proof in the supplementary material.

**Lemma A.11.** *Suppose that Assumption 1' holds. Then,  $\{p\vartheta_j\}_{j \in [r]}$  are the non-vanishing eigenvalues of  $\mathbf{Q}(\cdot)\mathbf{Q}(\cdot)^\top \in \mathbb{H}^p \otimes \mathbb{H}^p$  with the corresponding eigenfunctions  $\{\tilde{\mathbf{q}}_j(\cdot)\}_{j \in [r]}$ .*

We are now ready to prove Proposition 2.

(i) Note that  $\{\tau_i\}_{i=1}^\infty$  are the eigenvalues of  $\Sigma_y(\cdot, \cdot)$ , and  $\{p\vartheta_j\}_{j=1}^r$  are the  $r$  non-vanishing eigenvalues of  $\mathbf{Q}\mathbf{Q}^\top(\cdot, \cdot)$  by Lemma A.11. Applying Lemma A.3, we have, for  $j \in [r]$ ,

$$|\tau_j - p\vartheta_j| \leq \|\Sigma_y - \mathbf{Q}\mathbf{Q}^\top\|_{\mathcal{L}} = \|\Sigma_\varepsilon\|_{\mathcal{L}},$$

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

(ii) By Lemma A.7(iii), we have  $\|\Sigma_\varepsilon\|_{\mathcal{L}} \leq \|\Sigma_\varepsilon\|_{S,1}^{1/2} \|\Sigma_\varepsilon\|_{S,\infty}^{1/2} = O(s_p) = o(p)$ , which yields that  $\tau_j \asymp p\vartheta_j \asymp p$  for  $j \in [r]$ . Under Assumption 1',  $\vartheta_j$  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 A.4 that  $\|\varphi_j - \tilde{\mathbf{q}}_j\| = O(p^{-1}\|\Sigma_\varepsilon\|_{\mathcal{L}})$  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

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

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

$$\hat{\Sigma}_y^s(u, v) = \frac{1}{n} \mathbf{Y}(u)\mathbf{Y}(v)^\top = \sum_{j=1}^r \hat{\tau}_j \hat{\varphi}_j(u) \hat{\varphi}_j(v)^\top + \hat{\mathbf{R}}(u, v) = \hat{\mathbf{Q}}(u)\hat{\mathbf{Q}}(v)^\top + \tilde{\Sigma}_\varepsilon(u, v),$$

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

We next show that  $\hat{\mathbf{Q}}(u)\hat{\mathbf{Q}}(v)^\top = \sum_{j=1}^r \hat{\tau}_j \hat{\varphi}_j(u) \hat{\varphi}_j(v)^\top$  holds. To do this, we impose another identifiability condition that can serve as an alternative (see also Remark 2.1) to Assumption 1'.

**Assumption A.1.**  $p^{-1} \int \mathbf{Q}(u)^\top \mathbf{Q}(u) du = \mathbf{I}_r$  and  $\Sigma_\gamma$  is diagonal with distinct diagonal elements being bounded away from both 0 and  $\infty$  as  $p \rightarrow \infty$ .

Note that Assumptions 1' and A.1 can be converted to each other by orthogonal transformation. Thus,

for the minimization problem (17), we can use the following two equivalent normalization constraints:

$$\begin{aligned}
(i) \quad & n^{-1} \sum_{t=1}^n \gamma_t \gamma_t^T = \mathbf{I}_r, \text{ and } p^{-1} \int \mathbf{Q}(u)^T \mathbf{Q}(u) du \text{ is diagonal,} \\
(ii) \quad & n^{-1} \sum_{t=1}^n \gamma_t \gamma_t^T \text{ is diagonal, and } p^{-1} \int \mathbf{Q}(u)^T \mathbf{Q}(u) du = \mathbf{I}_r.
\end{aligned} \tag{A.1}$$

Note that (A.1)(i) is used in Section 2.2 to obtain FPOET estimator. Following the similar procedure, we obtain that  $\check{\mathbf{Q}}(\cdot) = \sqrt{p}(\hat{\varphi}_1(\cdot), \dots, \hat{\varphi}_r(\cdot))$  and  $\check{\mathbf{\Gamma}} = p^{-1} \int \mathbf{Y}(u)^T \check{\mathbf{Q}}(u) du$  is the solution to (17) under (A.1)(ii). One can show that the two solutions under normalization constraints (A.1)(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  $\check{\mathbf{Q}}(\cdot) = \hat{\mathbf{Q}}(\cdot)\mathbf{H}$  and  $\check{\mathbf{\Gamma}} = \hat{\mathbf{\Gamma}}\mathbf{H}$ . Notice that  $\check{\mathbf{Q}}(\cdot) = \sqrt{p}\{\hat{\varphi}_1(\cdot), \dots, \hat{\varphi}_r(\cdot)\}$  and  $\check{\mathbf{\Gamma}} = p^{-1} \int \mathbf{Y}(u)^T \check{\mathbf{Q}}(u) du$ , then we have

$$\begin{aligned}
n^{-1} \check{\mathbf{\Gamma}}^T \check{\mathbf{\Gamma}} &= p^{-2} n^{-1} \int \check{\mathbf{Q}}(u)^T \mathbf{Y}(u) du \int \mathbf{Y}(v)^T \check{\mathbf{Q}}(v) dv = p^{-2} \int \int \check{\mathbf{Q}}(u)^T \{n^{-1} \mathbf{Y}(u) \mathbf{Y}(v)^T\} \check{\mathbf{Q}}(v) dudv \\
&= p^{-1} \int \{\hat{\varphi}_1(u)^T, \dots, \hat{\varphi}_r(u)^T\}^T \left[ \int \hat{\Sigma}_y^S(u, v) \{\hat{\varphi}_1(v), \dots, \hat{\varphi}_r(v)\} dv \right] du \\
&= p^{-1} \int \{\hat{\varphi}_1(u)^T, \dots, \hat{\varphi}_r(u)^T\}^T \{\hat{\tau}_1 \hat{\varphi}_1(u), \dots, \hat{\tau}_r \hat{\varphi}_r(u)\} du = p^{-1} \text{diag}(\hat{\tau}_1, \dots, \hat{\tau}_r).
\end{aligned}$$

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

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

## B Proofs of theoretical results in Section 3

### B.1 Technical lemmas

**Lemma B.12.** *Under Assumptions 3(iv) and 4, we have that,*

- (i) *for any  $i, j \in [r]$ ,  $\|n^{-1} \sum_{t=1}^n f_{ti} f_{tj} - \Sigma_{f,ij}\|_S = O_p(1/\sqrt{n})$ ,  $\|n^{-1} \sum_{t=1}^n \mathbf{f}_t \mathbf{f}_t^T - \Sigma_f\|_{S, \max} = O_p(1/\sqrt{n})$ ;*
- (ii) *for any  $i, j \in [p]$ ,  $\|n^{-1} \sum_{t=1}^n \varepsilon_{ti} \varepsilon_{tj} - \Sigma_{\varepsilon,ij}\|_S = O_p(\mathcal{M}_\varepsilon/\sqrt{n})$ ,  $\|n^{-1} \sum_{t=1}^n \varepsilon_t \varepsilon_t^T - \Sigma_\varepsilon\|_{S, \max} = O_p(\mathcal{M}_\varepsilon \sqrt{\log p/n})$ ;*
- (iii) *for any  $i, j \in [p]$ ,  $\|n^{-1} \sum_{t=1}^n y_{ti} y_{tj} - \Sigma_{y,ij}\|_S = O_p(\mathcal{M}_\varepsilon/\sqrt{n})$ ,  $\|n^{-1} \sum_{t=1}^n \mathbf{y}_t \mathbf{y}_t^T - \Sigma_y\|_{S, \max} = O_p(\mathcal{M}_\varepsilon \sqrt{\log p/n})$ .*

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

**Lemma B.13.** *Suppose  $\iota$  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  $\lambda_i$ . Then, there exists an orthogonal matrix  $\mathbf{U} \in \mathbb{R}^{r \times r}$  such that

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

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

The proof of Lemma B.13 can be found in Fan et al. (2018).

**Lemma B.14.** (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

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

**Lemma B.15.** Suppose that  $\mathbf{K}, \mathbf{G} \in \mathbb{H}^p \otimes \mathbb{H}^p$  are Mercer's kernels,  $\|\mathbf{G}^{\dagger}\|_{\mathcal{L}} < c_n$  and  $\|\mathbf{K} - \mathbf{G}\|_{\mathcal{L}} = o_p(c_n^{-1})$  for a sequence  $c_n > 0$ . Let  $\tilde{\mathbf{K}}$  be the restriction of  $\mathbf{K}$  to  $\text{Ker}(\mathbf{G})^{\perp}$ . Define  $\mathbf{K}^{\dagger}$  as  $\mathbf{K}^{\dagger}(\mathbf{x}) = \tilde{\mathbf{K}}^{-1}(\mathbf{x})$  for  $\mathbf{x} \in \text{Im}(\mathbf{G})$  and  $\mathbf{K}^{\dagger}(\mathbf{x}) = \mathbf{0}$  for  $\mathbf{x} \in \text{Im}(\mathbf{G})^{\perp}$ . Then, we have (i)  $\|\mathbf{K}^{\dagger}\|_{\mathcal{L}} < 2c_n$  with probability approaching one, and (ii)  $\|\mathbf{K}^{\dagger} - \mathbf{G}^{\dagger}\|_{\mathcal{L}} = O_p(c_n^2)\|\mathbf{K} - \mathbf{G}\|_{\mathcal{L}}$ .

**Lemma B.16.** (i) Under Assumptions 3(ii)(iv) and 4,  $\max_{t \in [n]} \|\mathbf{f}_t\| = O_p(\sqrt{\log n})$ ,  $\max_{t \in [n]} \|p^{-1/2}\boldsymbol{\varepsilon}_t\| = O_p(\sqrt{\log n})$  and  $\max_{t \in [n]} \|p^{-1/2}\mathbf{B}^T \boldsymbol{\varepsilon}_t\| = O_p(\sqrt{\log n})$ .  
(ii) Under Assumptions 3'(ii)(iv) and 4',  $\max_{t \in [n]} \|\gamma_t\| = O_p(\sqrt{\log n})$ ,  $\max_{t \in [n]} \|p^{-1/2} \int \mathbf{Q}(u)^T \boldsymbol{\varepsilon}_t(u) du\| = O_p(\sqrt{\log n})$  and  $\max_{t' \in [n]} |p^{-1/2} \{\langle \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t'} \rangle - \mathbb{E}\langle \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t'} \rangle\}|^2 = O_p(\log n)$  for each  $t \in [n]$ .

## B.2 Proof of Theorem 1

The proof of part (i) of Theorem 1 mainly relies on Lemma B.13. To prove Theorem 1, we first present some technical lemmas. The proofs of Lemmas B.17–B.20 are provided in the supplementary material.

**Lemma B.17.** Suppose that Assumption 1 holds. Then there exist some constants  $C_{\max}, C_{\infty} > 0$  such that (i)  $\|\boldsymbol{\Sigma}_f\|_{\mathcal{S}, \max} \leq C_{\max}$ , (ii)  $\max(\|\boldsymbol{\Sigma}_f\|_{\mathcal{S}, \infty}, \|\boldsymbol{\Sigma}_f\|_{\mathcal{S}, 1}, \|\boldsymbol{\Sigma}_f\|_{\mathcal{S}, F}) \leq C_{\infty}$ .

**Lemma B.18.** 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)  $\|\boldsymbol{\Sigma}_y\|_{\mathcal{S}, 1} \lesssim p$ .

**Lemma B.19.** Supposing that Assumptions 1–3 hold, we have  $\|\boldsymbol{\Omega}_{\mathcal{L}}\| \asymp p^2$  and  $\|\boldsymbol{\Omega}_{\mathcal{R}}\| \lesssim p$ .

**Lemma B.20.** Under the assumptions of Theorem 1, we have  $\|\boldsymbol{\Omega}_{\mathcal{R}}\|_{\infty} \lesssim ps_p = o(p^2)$ .

**Lemma B.21.** Under the assumptions of Theorem 1, we have (i)  $\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\| = O_p(\mathcal{M}_{\varepsilon} p^2 \sqrt{1/n}) = o_p(p^2)$ , and (ii)  $\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\|_{\infty} = O_p(\mathcal{M}_{\varepsilon} p^2 \sqrt{\log p/n}) = o_p(p^2)$ .



*Proof.* (i) Note that

$$\begin{aligned}
\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}\| &= \left\| \int \int \hat{\mathbf{\Sigma}}_y^S(u, v) \hat{\mathbf{\Sigma}}_y^S(u, v)^T du dv - \int \int \mathbf{\Sigma}_y(u, v) \mathbf{\Sigma}_y(u, v)^T du dv \right\| \\
&= \left\| \int \int \left\{ \hat{\mathbf{\Sigma}}_y^S(u, v) - \mathbf{\Sigma}_y(u, v) \right\} \hat{\mathbf{\Sigma}}_y^S(u, v)^T + \mathbf{\Sigma}_y(u, v) \left\{ \hat{\mathbf{\Sigma}}_y^S(u, v)^T - \mathbf{\Sigma}_y(u, v)^T \right\} du dv \right\| \\
&= \left\| \int \int \left\{ \hat{\mathbf{\Sigma}}_y^S(u, v) - \mathbf{\Sigma}_y(u, v) \right\} \left\{ \hat{\mathbf{\Sigma}}_y^S(u, v)^T - \mathbf{\Sigma}_y(u, v)^T + \mathbf{\Sigma}_y(u, v)^T \right\} \right. \\
&\quad \left. + \mathbf{\Sigma}_y(u, v) \left\{ \hat{\mathbf{\Sigma}}_y^S(u, v)^T - \mathbf{\Sigma}_y(u, v)^T \right\} du dv \right\| \\
&\leq \left\| \int \int \left\{ \hat{\mathbf{\Sigma}}_y^S(u, v) - \mathbf{\Sigma}_y(u, v) \right\} \left\{ \hat{\mathbf{\Sigma}}_y^S(u, v) - \mathbf{\Sigma}_y(u, v) \right\}^T du dv \right\| \tag{B.2} \\
&\quad + 2 \left\| \int \int \left\{ \hat{\mathbf{\Sigma}}_y^S(u, v) - \mathbf{\Sigma}_y(u, v) \right\} \mathbf{\Sigma}_y(u, v)^T du dv \right\| \\
&\leq \|\hat{\mathbf{\Sigma}}_y^S - \mathbf{\Sigma}_y\|_{\mathcal{L}}^2 + 2\|\hat{\mathbf{\Sigma}}_y^S - \mathbf{\Sigma}_y\|_{\mathcal{L}} \|\mathbf{\Sigma}_y\|_{\mathcal{L}} \leq \|\hat{\mathbf{\Sigma}}_y^S - \mathbf{\Sigma}_y\|_{S, F}^2 + 2\|\hat{\mathbf{\Sigma}}_y^S - \mathbf{\Sigma}_y\|_{S, F} \|\mathbf{\Sigma}_y\|_{S, F} \\
&= \sum_{i=1}^p \sum_{j=1}^p \|n^{-1} \sum_{t=1}^n y_{ti} y_{tj} - \Sigma_{y, ij}\|_S^2 + 2 \left( \sum_{i=1}^p \sum_{j=1}^p \|n^{-1} \sum_{t=1}^n y_{ti} y_{tj} - \Sigma_{y, ij}\|_S^2 \right)^{1/2} \|\mathbf{\Sigma}_y\|_{S, F} \\
&= O_p \left( \mathcal{M}_\varepsilon p^2 \sqrt{1/n} \right) = o_p(p^2),
\end{aligned}$$

where the second inequality follows from Lemmas A.5(ii) and A.6(iv), the third inequality follows from Lemma A.7(ii), and the last line follows from Lemma B.12(iii), the fact that  $\|\mathbf{K}\|_{S, F} \leq p \|\mathbf{K}\|_{S, \max}$  and Assumption 4(ii).

(ii) The argument can be proved in matrix  $\ell_\infty$  norm following the similar procedure. Specifically,

$$\begin{aligned}
\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}\|_\infty &\leq \left\| \int \int \left\{ \hat{\mathbf{\Sigma}}_y^S(u, v) - \mathbf{\Sigma}_y(u, v) \right\} \left\{ \hat{\mathbf{\Sigma}}_y^S(u, v) - \mathbf{\Sigma}_y(u, v) \right\}^T du dv \right\|_\infty \\
&\quad + 2 \left\| \int \int \left\{ \hat{\mathbf{\Sigma}}_y^S(u, v) - \mathbf{\Sigma}_y(u, v) \right\} \mathbf{\Sigma}_y(u, v)^T du dv \right\|_\infty \\
&\leq \|\hat{\mathbf{\Sigma}}_y^S - \mathbf{\Sigma}_y\|_{S, \infty} \|\hat{\mathbf{\Sigma}}_y^S - \mathbf{\Sigma}_y\|_{S, 1} + 2\|\hat{\mathbf{\Sigma}}_y^S - \mathbf{\Sigma}_y\|_{S, \infty} \|\mathbf{\Sigma}_y\|_{S, 1} \\
&\asymp p^2 \|\hat{\mathbf{\Sigma}}_y^S - \mathbf{\Sigma}_y\|_{S, \max} = O_p(\mathcal{M}_\varepsilon p^2 \sqrt{\log p/n}) = o_p(p^2),
\end{aligned}$$

where the first inequality can be obtained in a way similar to (B.2), the second inequality follows from Lemma A.6(ii), and the last line follows from Lemma B.12(iii) and Assumption 4(ii).  $\square$

**Lemma B.22.** Let  $\{\hat{\lambda}_j\}_{j=1}^p$  be the eigenvalues of  $\hat{\mathbf{\Omega}}$  in a descending order. Under the assumptions of Theorem 1, it holds that  $\hat{\lambda}_r \gtrsim p^2$  with probability approaching one. Furthermore,  $\hat{\lambda}_i - \hat{\lambda}_j \gtrsim p^2$  for all  $1 \leq i < j \leq r$  with probability approaching one.

*Proof.* By Proposition 1 and Lemma B.19, the  $r$ -th largest eigenvalue  $\lambda_r$  of  $\mathbf{\Omega}$  satisfies  $\lambda_r \geq p^2 \theta_r - |\lambda_r - p^2 \theta_r| \geq p^2 \theta_r - \|\mathbf{\Omega}_{\mathcal{R}}\| \gtrsim p^2$ . Applying Lemma A.1 yields that  $|\hat{\lambda}_j - \lambda_j| \leq \|\hat{\mathbf{\Omega}} - \mathbf{\Omega}\|$ , for  $j \in [p]$ . From Lemma B.21(i), we have  $\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}\| = o_p(p^2)$  and hence  $\hat{\lambda}_r \gtrsim p^2$  with probability approaching one. Furthermore, for all  $1 \leq i < j \leq r$ , with probability approaching one,

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

□

The following lemma is used to prove Theorems 3 and 6, and its proof is provided in the supplementary material.

**Lemma B.23.** *Under the assumptions of Theorem 1, we have  $\|\hat{\mathbf{B}} - \mathbf{B}\mathbf{U}^T\|_F = O_p(\mathcal{M}_\varepsilon\sqrt{p/n} + 1/\sqrt{p})$ .*

We are now ready to prove Theorem 1.

(i) Let  $\mathbf{E} = \hat{\mathbf{\Omega}} - \mathbf{\Omega}$  be the  $p \times p$  perturbation matrix. By Lemma B.21, we have

$$\|\mathbf{E}\|_\infty \leq \|\hat{\mathbf{\Omega}} - \mathbf{\Omega}\|_\infty = O_p(\mathcal{M}_\varepsilon p^2 \sqrt{\log p/n}) = o_p(p^2).$$

Corresponding to Lemma B.13, here  $\mathbf{A} = \mathbf{\Omega}$ ,  $\tilde{\mathbf{A}} = \hat{\mathbf{\Omega}}$ , and the  $r$ -th eigenvalue of  $\mathbf{A}$  satisfies  $\lambda_r \asymp p^2$  by Proposition 1 and Lemma B.19. Then,  $\|\mathbf{A} - \mathbf{A}_r\|_\infty \leq \|\mathbf{\Omega} - \mathbf{\Omega}_\mathcal{L}\|_\infty = \|\mathbf{\Omega}_\mathcal{R}\|_\infty = ps_p = o(p^2)$  from Lemma B.20. Note that  $\mathbf{V} = (\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_r) \in \mathbb{R}^{p \times r}$ , and denote  $\boldsymbol{\xi}_j = (\xi_{1j}, \dots, \xi_{pj})^T$ . The coherence of  $\mathbf{V}$  is given by

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

since  $\max_{i \in [p]} \tilde{b}_{ij} = \max_{i \in [p]} p^{-1/2} b_{ij} \leq p^{-1/2} \|\mathbf{B}\|_{\max} = O(p^{-1/2})$  and  $\|\boldsymbol{\xi}_j - \tilde{\mathbf{b}}_j\| = O(p^{-1})$  if  $\boldsymbol{\xi}_j^T \tilde{\mathbf{b}}_j \geq 0$  by Proposition 1(ii) and Lemma B.19. In addition, supposing that  $\iota \approx \|\mathbf{E}\| = o_p(p^2)$  but  $\iota > \|\mathbf{E}\|$ , we can show that for any  $j \in [r]$ , the interval  $(\lambda_j - \iota, \lambda_j + \iota)$  does not contain any eigenvalues of  $\mathbf{\Omega}$  other than  $\lambda_j$  with probability approaching one. Thus by Lemma B.13, we have for  $j \in [r]$ , if  $\hat{\boldsymbol{\xi}}_j^T \boldsymbol{\xi}_j \geq 0$ ,

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

For  $j \in [r]$ , if  $\boldsymbol{\xi}_j^T \tilde{\mathbf{b}}_j \geq 0$ , we have  $\|\boldsymbol{\xi}_j - \tilde{\mathbf{b}}_j\|_{\max} \leq \|\boldsymbol{\xi}_j - \tilde{\mathbf{b}}_j\| = O(p^{-1})$ , which implies that  $\|\hat{\boldsymbol{\xi}}_j - \tilde{\mathbf{b}}_j\|_{\max} = O_p(\mathcal{M}_\varepsilon\sqrt{\log p/pn} + 1/p)$ . Since  $\hat{\mathbf{B}} = \sqrt{p}(\hat{\boldsymbol{\xi}}_1, \dots, \hat{\boldsymbol{\xi}}_r)$  and  $\mathbf{B} = \sqrt{p}(\tilde{\mathbf{b}}_1, \dots, \tilde{\mathbf{b}}_r)$ , one can obtain that there exists an orthogonal matrix  $\mathbf{U} \in \mathbb{R}^{r \times r}$  (the same as that in Lemma B.23) such that

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

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

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

$$\hat{\mathbf{f}}_t(\cdot) - \mathbf{U}\mathbf{f}_t(\cdot) = p^{-1}(\hat{\mathbf{B}}^T \mathbf{B} - \mathbf{U}\mathbf{B}^T \mathbf{B})\mathbf{f}_t(\cdot) + p^{-1}\hat{\mathbf{B}}^T \boldsymbol{\varepsilon}_t(\cdot). \quad (\text{B.3})$$

For the first term of (B.3), applying Lemmas A.10 and B.23 yields that

$$\|p^{-1}(\hat{\mathbf{B}}^T \mathbf{B} - \mathbf{U}\mathbf{B}^T \mathbf{B})\mathbf{f}_t\| \leq p^{-1}\|\hat{\mathbf{B}}^T - \mathbf{U}\mathbf{B}^T\| \|\mathbf{B}\| \|\mathbf{f}_t\| = O_p(\mathcal{M}_\varepsilon\sqrt{1/n} + 1/p),$$

since  $\|\hat{\mathbf{B}}^T - \mathbf{U}\mathbf{B}^T\| \leq \|\hat{\mathbf{B}} - \mathbf{B}\mathbf{U}^T\|_F = O_p(\mathcal{M}_\varepsilon\sqrt{p/n} + 1/\sqrt{p})$ ,  $\|\mathbf{B}\| = \lambda_{\max}^{1/2}(\mathbf{B}^T \mathbf{B}) = \sqrt{p}$  and  $\|\mathbf{f}_t\| = O_p(1)$ .

For the second term of (B.3), notice that

$$\|p^{-1}\hat{\mathbf{B}}^T \boldsymbol{\varepsilon}_t\| = \|p^{-1}\mathbf{U}\mathbf{B}^T \boldsymbol{\varepsilon}_t\| + \|p^{-1}(\hat{\mathbf{B}}^T - \mathbf{U}\mathbf{B}^T)\boldsymbol{\varepsilon}_t\| = O_p(\mathcal{M}_\varepsilon\sqrt{1/n} + 1/\sqrt{p}),$$

which follows from  $\|p^{-1}\mathbf{B}^T\boldsymbol{\varepsilon}_t\| = O_p(1/\sqrt{p})$  by Assumption 3(ii),  $\|\hat{\mathbf{B}}^T - \mathbf{U}\mathbf{B}^T\| = O_p(\mathcal{M}_\varepsilon\sqrt{p/n} + 1/\sqrt{p})$  by Lemma B.23, and  $\|\boldsymbol{\varepsilon}_t\| = O_p(\sqrt{p})$  since  $\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^2 \leq p \max_{i \in [p]} \mathbb{E}\|\varepsilon_{ti}\|^2 = p \max_{i \in [p]} \|\Sigma_{\varepsilon, ii}\|_{\mathcal{N}} = O(p)$  by Assumption 3(iv). The result follows immediately that for  $t \in [n]$ ,

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

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

(iii) The proof procedure is similar to part (ii). We only need to notice that by Lemma B.16(i), we have  $\max_{t \in [n]} \|\mathbf{f}_t\| = O_p(\sqrt{\log n})$ ,  $\max_{t \in [n]} \|p^{-1}\boldsymbol{\varepsilon}_t\| = O_p(\sqrt{\log n/p})$  and  $\max_{t \in [n]} \|p^{-1}\mathbf{B}\boldsymbol{\varepsilon}_t\| = O_p(\sqrt{\log n/p})$ .

### B.3 Proof of Corollary 1

By Theorem 1(i)(iii), Lemma B.16(i), and Assumption 3(i), we have

$$\begin{aligned} \max_{i \in [p], t \in [n]} \|\check{\mathbf{b}}_i^T \hat{\mathbf{f}}_t - \check{\mathbf{b}}_i^T \mathbf{f}_t\| &\leq \max_{i \in [p]} \|\check{\mathbf{b}}_i - \mathbf{U}\check{\mathbf{b}}_i\| \cdot \max_{t \in [n]} \|\hat{\mathbf{f}}_t\| + \max_{i \in [p]} \|\check{\mathbf{b}}_i\| \cdot \max_{t \in [n]} \|\mathbf{U}^T \hat{\mathbf{f}}_t - \mathbf{f}_t\| \\ &= O_p(\varpi_{n,p} \cdot \sqrt{\log n}) + O_p(\mathcal{M}_\varepsilon \sqrt{\log n/n} + \sqrt{\log n/p}) \\ &= O_p(\mathcal{M}_\varepsilon \sqrt{\log n \log p/n} + \sqrt{\log n/p}). \end{aligned}$$

### B.4 Proof of Theorem 2

(i) By Proposition 1 and Lemma B.19,  $|\lambda_j - p^2\theta_j| \leq \|\boldsymbol{\Omega}_{\mathcal{R}}\| \lesssim p$  for  $j \in [r]$ , which implies that  $\lambda_j \asymp p^2$  for  $j \in [r]$ , and  $|\lambda_j| \lesssim p$  for  $r+1 \leq j \leq r_{1,0}$ . By Lemmas A.1 and B.21, it follows that  $|\hat{\lambda}_j - \lambda_j| \leq \|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\| = O_p(\mathcal{M}_\varepsilon p^2 n^{-1/2})$  for  $j \in [p]$ . Let  $\beta_{1n} = \mathcal{M}_\varepsilon p^2 n^{-1/2}$  and  $\tilde{\beta}_{1n} = p + \mathcal{M}_\varepsilon p^2 n^{-1/2}$ . Then, we have  $|\hat{\lambda}_j - \lambda_j| = O_p(\beta_{1n})$  for  $j \in [r]$  and  $|\hat{\lambda}_j| = O_p(\tilde{\beta}_{1n})$  for  $r+1 \leq j \leq r_{1,0}$ . We next verify the following conditions (a), (b) and (c) by Assumptions 4(ii) and 7:

- (a)  $(\vartheta_{1n} + \tilde{\beta}_{1n})/(\lambda_r^2/\lambda_1) \asymp (\vartheta_{1n} + p + \mathcal{M}_\varepsilon p^2 n^{-1/2})/p^2 \rightarrow 0$ ;
- (b)  $\beta_{1n}/\lambda_r \asymp \mathcal{M}_\varepsilon p^2 n^{-1/2}/p^2 = \mathcal{M}_\varepsilon n^{-1/2} \rightarrow 0$ ;
- (c)  $\tilde{\beta}_{1n}^2/(\vartheta_{1n}\lambda_r) \asymp (p + \mathcal{M}_\varepsilon p^2 n^{-1/2})^2/(\vartheta_{1n}p^2) \asymp \vartheta_{1n}^{-1} + \vartheta_{1n}^{-1}\mathcal{M}_\varepsilon^2 p^2 n^{-1} \rightarrow 0$ .

Under (a), (b) and (c), we apply Proposition 1 of Han et al. (2022) and obtain  $\mathbb{P}(\hat{r}^{\mathcal{D}} = r) \rightarrow 1$  with  $\hat{r}^{\mathcal{D}}$  defined in (19), which completes the proof of Theorem 2(i).

(ii) By Proposition 2,  $|\tau_j - p\vartheta_j| \leq \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}} = O(1)$  for  $j \in [r]$ , which implies that  $\tau_j \asymp p$  for  $j \in [r]$ , and  $|\tau_j| = O(1)$  for  $r+1 \leq j \leq r_{2,0}$ . By Lemma A.3 and the proof of Lemma S.5 of the supplementary material, it follows that  $|\hat{\tau}_j - \tau_j| \leq \|\hat{\boldsymbol{\Sigma}}_y^{\mathcal{S}} - \boldsymbol{\Sigma}_y\|_{\mathcal{S}, \mathcal{F}} = O_p(\mathcal{M}_\varepsilon p n^{-1/2})$  for  $j \in [p]$ . Let  $\beta_{2n} = \mathcal{M}_\varepsilon p n^{-1/2}$  and  $\tilde{\beta}_{2n} = 1 + \mathcal{M}_\varepsilon p n^{-1/2}$ . Then, we have  $|\hat{\tau}_j - \tau_j| = O_p(\beta_{2n})$  for  $j \in [r]$  and  $|\hat{\tau}_j| = O_p(\tilde{\beta}_{2n})$  for  $r+1 \leq j \leq r_{2,0}$ . We next verify the following conditions (d), (e) and (f) by Assumptions 4'(ii) and 7':

- (d)  $(\vartheta_{2n} + \tilde{\beta}_{2n})/(\tau_r^2/\tau_1) \asymp (\vartheta_{2n} + 1 + \mathcal{M}_\varepsilon p n^{-1/2})/p \rightarrow 0$ ;
- (e)  $\beta_{2n}/\tau_r \asymp \mathcal{M}_\varepsilon p n^{-1/2}/p = \mathcal{M}_\varepsilon n^{-1/2} \rightarrow 0$ ;
- (f)  $\tilde{\beta}_{2n}^2/(\vartheta_{2n}\tau_r) \asymp (1 + \mathcal{M}_\varepsilon p n^{-1/2})^2/(\vartheta_{2n}p) \asymp \vartheta_{2n}^{-1}p^{-1} + \vartheta_{2n}^{-1}\mathcal{M}_\varepsilon^2 p n^{-1} \rightarrow 0$ .

Under (d), (e) and (f), we apply Proposition 1 of Han et al. (2022) and obtain  $\mathbb{P}(\hat{r}^{\mathcal{F}} = r) \rightarrow 1$  with  $\hat{r}^{\mathcal{F}}$  defined in (20), which completes the proof of Theorem 2(ii).

## B.5 Proof of Theorem 3

We denote  $r_0 = r_{1,0} \wedge r_{2,0}$  for simplicity.

**Lemma B.24.** (i) Under Assumptions 1–4 and 8, there exists some constant  $c_1 > 0$  such that  $\mathbb{P}\{\log V^{\mathcal{D}}(r) + c_1 \leq \min_{k \in [r-1]} \log V^{\mathcal{F}}(k)\} \rightarrow 1$  and  $\mathbb{P}\{\log V^{\mathcal{D}}(k) + c_1 \leq \log V^{\mathcal{F}}(k)\} \rightarrow 1$  for  $r \leq k \leq r_0$ .

(ii) Under Assumptions 1'–4' and 8', there exists some constant  $c_2 > 0$  such that  $\mathbb{P}\{\min_{k \in [r-1]} \log V^{\mathcal{D}}(k) \geq \log V^{\mathcal{F}}(r) + c_2\} \rightarrow 1$  and  $\mathbb{P}\{\log V^{\mathcal{D}}(k) \geq \log V^{\mathcal{F}}(k) + c_2\} \rightarrow 1$  for  $r \leq k \leq r_0$ .

*Proof.* Let  $\mathbf{e}_{1kt}(\cdot) = \mathbf{y}_t(\cdot) - p^{-1} \hat{\mathbf{B}}_k \hat{\mathbf{B}}_k^T \mathbf{y}_t(\cdot)$  and  $\mathbf{e}_{2kt}(\cdot) = \mathbf{y}_t(\cdot) - n^{-1} \mathbf{Y}_t(\cdot) \hat{\mathbf{\Gamma}}_k \hat{\mathbf{\gamma}}_{t,k}$  for  $t \in [n]$  and  $k \in [r_0]$ , where  $\hat{\mathbf{B}}_k$  is the estimated loadings by DIGIT estimator with  $k$  functional factors, and  $\hat{\mathbf{\Gamma}}_k$  is the estimated factors by FPOET estimator with  $k$  scalar factors. By definitions, it can be seen that

$$V^{\mathcal{D}}(k) = \frac{1}{pn} \sum_{t=1}^n \|\mathbf{e}_{1kt}\|^2 = \frac{1}{pn} \sum_{t=1}^n \int \mathbf{e}_{1kt}(u)^T \mathbf{e}_{1kt}(u) du = \frac{1}{p} \text{tr} \left\{ \int \frac{1}{n} \sum_{t=1}^n \mathbf{e}_{1kt}(u) \mathbf{e}_{1kt}(u)^T du \right\} = p^{-1} \|\hat{\Sigma}_{e,1k}^S\|_{\mathcal{N}},$$

and similarly it follows that  $V^{\mathcal{F}}(k) = p^{-1} \|\hat{\Sigma}_{e,2k}^S\|_{\mathcal{N}}$ , where  $\hat{\Sigma}_{e,1k}^S$  and  $\hat{\Sigma}_{e,2k}^S$  are the sample covariance of  $\{\mathbf{e}_{1kt}(\cdot)\}$  and  $\{\mathbf{e}_{2kt}(\cdot)\}$ , respectively. Notice that  $\chi_t(\cdot) = \mathbf{B} \mathbf{f}_t(\cdot)$  and  $\kappa_t(\cdot) = \mathbf{Q}(\cdot) \gamma_t$  are the common components of the two respective FFMs. By Section S.5.2 of the supplementary material, we have

$$\Sigma_{\chi}(u, v) = \sum_{i=1}^{\infty} p \omega_i \psi_i(u) \psi_i(v)^T \quad \text{and} \quad \Sigma_{\kappa}(u, v) = \sum_{j=1}^r p \check{\gamma}_j \nu_j(u) \nu_j(v)^T \quad \text{for } (u, v) \in \mathcal{U}^2.$$

(i) For the first argument, note that  $\mathbf{e}_{1rt} - \varepsilon_t = \mathbf{B} \mathbf{f}_t - \hat{\mathbf{B}} \hat{\mathbf{f}}_t$  for  $t \in [n]$ . On the one hand, by Theorem 1 and Lemma B.23, it can be shown that  $n^{-1} \sum_{t=1}^n \|\mathbf{B} \mathbf{f}_t - \hat{\mathbf{B}} \hat{\mathbf{f}}_t\|^2 = O_p(\mathcal{M}_{\varepsilon}^2 p/n + 1) = o_p(p)$ . Since  $\mathbb{E}(n^{-1} \sum_{t=1}^n \|\varepsilon_t\|^2) = n^{-1} \sum_{t=1}^n \mathbb{E} \|\varepsilon_t\|^2 = o(p)$  by Assumption 8(ii), we have  $n^{-1} \sum_{t=1}^n \|\varepsilon_t\|^2 = o_p(p)$ . By using the inequality  $(a + b)^2 \leq 2(a^2 + b^2)$ , it follows that  $\|\hat{\Sigma}_{e,1r}^S\|_{\mathcal{N}} = o_p(p)$  and thus  $V^{\mathcal{D}}(r) = o_p(1)$ . On the other hand, by Assumption 8(i) and Lemma A.5(iii), it can be shown that for  $1 \leq k < r$ ,  $n^{-1} \sum_{t=1}^n \|\mathbf{e}_{2kt} - \varepsilon_t\|^2 \gtrsim p$  with probability approaching one, since  $\sum_{i=k+1}^{\infty} p \omega_i \asymp p$  and the leading  $k$  eigenfunctions of  $\hat{\Sigma}_y^S$  cannot recover the space spanned by the eigenfunctions of  $\Sigma_{\chi}$  corresponding to the eigenvalues with order  $p$ . By using the inequality  $(a - b)^2 \geq a^2/2 - b^2$  and  $n^{-1} \sum_{t=1}^n \|\varepsilon_t\|^2 = o_p(p)$ , we have  $\|\hat{\Sigma}_{e,2k}^S\|_{\mathcal{N}} \gtrsim p$  and  $V^{\mathcal{F}}(k) \gtrsim 1$  with probability approaching one for  $1 \leq k < r$ . Hence, there exists some small constant  $0 < c_1 < 1/2$  such that  $\min_{k \in [r-1]} V^{\mathcal{F}}(k)/V^{\mathcal{D}}(r) > 1 + 2c_1$  with probability approaching one for all large  $p$  and  $n$ , which implies that  $\log\{\min_{k \in [r-1]} V^{\mathcal{F}}(k)/V^{\mathcal{D}}(r)\} \geq c_1$  with probability approaching one by similar argument to the proof of Corollary 1 of Bai and Ng (2002).

For the second argument, following the similar procedures to the proof of Lemma 4 in Bai and Ng (2002), it can be shown that for any fixed  $k \geq r$ ,  $V^{\mathcal{D}}(r) - V^{\mathcal{D}}(k) = O_p(\mathcal{M}_{\varepsilon}^2/n + 1/p) = o_p(1)$ , which implies that  $V^{\mathcal{D}}(k) = o_p(1)$  for  $k \geq r$ . By Assumption 8(i) and using the similar procedures to the first argument, we have  $V^{\mathcal{F}}(k) \gtrsim 1$  with probability approaching one. The desired results hold accordingly.

(ii) The proofs are similar to part (i). To show  $V^{\mathcal{D}}(k) \gtrsim 1$  with probability approaching one, it relies on Assumption 8'(i) that  $\text{rank} \left\{ \int \int \Sigma_{\kappa}(u, v) \Sigma_{\kappa}(u, v)^T du dv \right\} = \text{rank} \left\{ \int \mathbf{Q}(u) \mathbf{Q}(u)^T du \right\} \geq k + 1$  for  $k \in [r_0]$ .  $\square$

**Lemma B.25.** Suppose that  $g(p, n) \rightarrow 0$  and  $(\mathcal{M}_{\varepsilon}^2/n + 1/p)^{-1} g(p, n) \rightarrow \infty$  as  $p, n \rightarrow \infty$  for the penalty functions  $g^{\mathcal{D}}(p, n)$  and  $g^{\mathcal{F}}(p, n)$  in (21). Then, (i) under Assumptions 1–4,  $\mathbb{P}\{\text{IC}^{\mathcal{D}}(r) \leq \text{IC}^{\mathcal{D}}(k)\} \rightarrow 1$  for  $k \in [r_0]$ ; (ii) under Assumptions 1'–4',  $\mathbb{P}\{\text{IC}^{\mathcal{F}}(r) \leq \text{IC}^{\mathcal{F}}(k)\} \rightarrow 1$  for  $k \in [r_0]$ .

*Proof.* (ii) Let  $\mathbf{\Gamma}_k = (\gamma_{1,k}, \dots, \gamma_{n,k})^\top$  be a  $n \times k$  matrix. Define the average of squared residuals when using  $\{\gamma_{t,k}\}_{t \in [n]}$  as  $k$  known factors for the estimation of FFM (2):

$$\tilde{V}^\mathcal{F}(k, \mathbf{\Gamma}_k) = \min_{\mathbf{Q}(\cdot)} \frac{1}{pn} \sum_{t=1}^n \|\mathbf{y}_t - \mathbf{Q}\gamma_{t,k}\|^2 = \frac{1}{pn} \sum_{t=1}^n \|\mathbf{y}_t - n^{-1}\mathbf{Y}\mathbf{\Gamma}_k\gamma_{t,k}\|^2.$$

Note that  $\tilde{V}^\mathcal{F}(k, \hat{\mathbf{\Gamma}}_k) = V^\mathcal{F}(k)$ . The proof can be organized in the following three steps.

*Step 1.* For any fixed  $k \in [r_0]$ , define  $\mathbf{H}_k = n^{-1}\mathbf{V}_k^{-1}\hat{\mathbf{\Gamma}}_k^\top \mathbf{\Gamma} \int \mathbf{Q}(u)^\top \mathbf{Q}(u) du$ , where  $\mathbf{V}_k$  is the diagonal matrix of the first  $k$  largest eigenvalues of  $\hat{\mathbf{\Sigma}}_y^S$  in a decreasing order,  $\mathbf{\Gamma} = (\gamma_1, \dots, \gamma_n)^\top$  is the true factors and  $\hat{\mathbf{\Gamma}}_k = (\hat{\gamma}_{1,k}, \dots, \hat{\gamma}_{n,k})^\top$  is formed by  $k$  estimated factors using FPOET. Similar to (S.6) of the supplementary material, we have

$$\hat{\gamma}_{t,k} - \mathbf{H}_k \gamma_t = \left(\frac{\mathbf{V}_k}{p}\right)^{-1} \left\{ \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t',k} \frac{\mathbb{E}\langle \varepsilon_{t'}, \varepsilon_t \rangle}{p} + \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t',k} \zeta_{t't} + \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t',k} \eta_{t't} + \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t',k} \xi_{t't} \right\},$$

where  $\zeta_{t't}$ ,  $\eta_{t't}$  and  $\xi_{t't}$  take the same definitions as (S.6). Analogous to the proofs of Lemma S.4 and Theorem 1'(i), it can be shown that  $n^{-1} \sum_{t=1}^n \|\hat{\gamma}_{t,k} - \gamma_t\|^2 = O_p(\mathcal{M}_\varepsilon^2/n + 1/p)$  for any fixed  $k \in [r_0]$ .

*Step 2.* Following the similar procedures to the proofs of Lemmas 2–4 in Bai and Ng (2002) and combining the result in Step 1, we obtain that: (i) for  $k \in [r]$ ,  $|\tilde{V}^\mathcal{F}(k, \hat{\mathbf{\Gamma}}_k) - \tilde{V}^\mathcal{F}(k, \mathbf{H}_k \mathbf{\Gamma})| = O_p(\mathcal{M}_\varepsilon/\sqrt{n} + 1/\sqrt{p})$ ; (ii) for  $k \in [r]$ , there exists some constant  $c'_k > 0$  such that  $\tilde{V}^\mathcal{F}(k, \mathbf{H}_k \mathbf{\Gamma}) - \tilde{V}^\mathcal{F}(r, \mathbf{\Gamma}) \geq c'_k$  with probability approaching one; (iii) for  $r+1 \leq k \leq r_0$ ,  $|\tilde{V}^\mathcal{F}(k, \hat{\mathbf{\Gamma}}_k) - \tilde{V}^\mathcal{F}(r, \hat{\mathbf{\Gamma}}_r)| = O_p(\mathcal{M}_\varepsilon^2/n + 1/p)$ .

*Step 3.* For  $k \in [r]$ , consider that

$$\begin{aligned} \tilde{V}^\mathcal{F}(k, \hat{\mathbf{\Gamma}}_k) - \tilde{V}^\mathcal{F}(r, \hat{\mathbf{\Gamma}}_r) &= \{\tilde{V}^\mathcal{F}(k, \hat{\mathbf{\Gamma}}_k) - \tilde{V}^\mathcal{F}(k, \mathbf{H}_k \mathbf{\Gamma})\} + \{\tilde{V}^\mathcal{F}(k, \mathbf{H}_k \mathbf{\Gamma}) - \tilde{V}^\mathcal{F}(r, \mathbf{H}_r \mathbf{\Gamma})\} \\ &\quad + \{\tilde{V}^\mathcal{F}(r, \mathbf{H}_r \mathbf{\Gamma}) - \tilde{V}^\mathcal{F}(r, \hat{\mathbf{\Gamma}}_r)\} \geq c'_k \end{aligned}$$

with probability approaching one, where the first and third terms are both  $O_p(\mathcal{M}_\varepsilon/\sqrt{n} + 1/\sqrt{p}) = o_p(1)$  by the first argument of Step 2 and the second term is large than  $c'_k$  with probability approaching one by the second argument of Step 2 since  $\tilde{V}^\mathcal{F}(r, \mathbf{H}_r \mathbf{\Gamma}) = \tilde{V}^\mathcal{F}(r, \mathbf{\Gamma})$ . Thus, there exists some small constant  $0 < \epsilon_1 < 1/2$  such that  $V^\mathcal{F}(k)/V^\mathcal{F}(r) > 1 + 2\epsilon_1$  with probability approaching one, which implies that  $\log\{V^\mathcal{F}(k)/V^\mathcal{F}(r)\} > \epsilon_1$  with probability approaching one. Since  $g^\mathcal{F}(p, n) \rightarrow 0$  as  $p, n \rightarrow \infty$ , we have  $\mathbb{P}\{\text{IC}^\mathcal{F}(k) > \text{IC}^\mathcal{F}(r)\} \rightarrow 1$  for  $k \in [r]$ . For  $r+1 \leq k \leq r_0$ , the third argument of Step 2 shows that  $V^\mathcal{F}(k)/V^\mathcal{F}(r) = 1 + O_p(\mathcal{M}_\varepsilon^2/n + 1/p)$ , which implies that  $\log\{V^\mathcal{F}(k)/V^\mathcal{F}(r)\} = O_p(\mathcal{M}_\varepsilon^2/n + 1/p)$ . As  $(\mathcal{M}_\varepsilon^2/n + 1/p)^{-1} g^\mathcal{F}(p, n) \rightarrow \infty$ , it follows that  $\mathbb{P}\{\text{IC}^\mathcal{F}(k) > \text{IC}^\mathcal{F}(r)\} \rightarrow 1$  for  $r+1 \leq k \leq r_0$ , which completes the proof of part (ii).

(i) Let  $\mathbf{B}_k$  be a given  $p \times k$  loading matrix of  $k$  functional factors. Define

$$\tilde{V}^\mathcal{D}(k, \mathbf{B}_k) = \min_{\mathbf{f}_{t,k}(\cdot) \in \mathbb{H}^k, t \in [n]} \frac{1}{pn} \sum_{t=1}^n \|\mathbf{y}_t - \mathbf{B}_k \mathbf{f}_{t,k}\|^2 = \frac{1}{pn} \sum_{t=1}^n \|\mathbf{y}_t - p^{-1} \mathbf{B}_k \mathbf{B}_k^\top \mathbf{y}_t\|^2.$$

Note that  $\tilde{V}^\mathcal{D}(k, \hat{\mathbf{B}}_k) = V^\mathcal{D}(k)$ . The remaining proof procedures are similar to part (ii) and omitted.  $\square$

We are now ready to prove Theorem 3.

(i) Define the following events:  $E_1 = \{\text{IC}^\mathcal{D}(\hat{r}^\mathcal{D}) = \text{IC}^\mathcal{D}(r)\}$ ,  $E_{2k} = \{\text{IC}^\mathcal{D}(r) \leq \text{IC}^\mathcal{D}(k)\}$  for  $k \in [r_0]$ ,  $E_3 =$

$\{\text{IC}^\mathcal{D}(r) < \min_{k \in [r-1]} \text{IC}^\mathcal{F}(k)\}$  and  $E_{4k} = \{\text{IC}^\mathcal{D}(k) < \text{IC}^\mathcal{F}(k)\}$  for  $r \leq k \leq r_0$ . As  $p, n \rightarrow \infty$ , we have  $\mathbb{P}(E_1) \rightarrow 1$  by Theorem 2(i),  $\mathbb{P}(E_{2k}) \rightarrow 1$  by Lemma B.25(i),  $\mathbb{P}(E_3) \rightarrow 1$  by the first argument of Lemma B.24(i) and  $g^\mathcal{D}(p, n) = o(1), g^\mathcal{F}(p, n) = o(1)$ , and  $\mathbb{P}(E_{4k}) \rightarrow 1$  by the second argument of Lemma B.24(i) and  $g^\mathcal{D}(p, n) = o(1), g^\mathcal{F}(p, n) = o(1)$ . Let  $E_2 = \bigcap_{k=1}^{r_0} E_{2k}$  and  $E_4 = \bigcap_{k=r}^{r_0} E_{4k}$ . Then, under the event  $E = E_1 \cap E_2 \cap E_3 \cap E_4$  with  $\mathbb{P}(E) \rightarrow 1$ , it follows that

$$\text{IC}^\mathcal{D}(\hat{r}^\mathcal{D}) = \text{IC}^\mathcal{D}(r) = \min_{k \in [r_0]} \text{IC}^\mathcal{D}(k) < \min_{k \in [r_0]} \text{IC}^\mathcal{F}(k) \leq \text{IC}^\mathcal{F}(\hat{r}^\mathcal{F}),$$

where the first equality holds under  $E_1$ , the second equality holds under  $E_2$ , the first inequality holds under  $E_3 \cap E_4$  since  $\min_{k \in [r_0]} \text{IC}^\mathcal{F}(k) = \min_{k \in [r-1]} \text{IC}^\mathcal{F}(k) \wedge \text{IC}^\mathcal{F}(r) \wedge \text{IC}^\mathcal{F}(r+1) \wedge \cdots \wedge \text{IC}^\mathcal{F}(r_0) > \text{IC}^\mathcal{D}(r) \wedge \text{IC}^\mathcal{D}(r+1) \wedge \cdots \wedge \text{IC}^\mathcal{D}(r_0) \geq \min_{k \in [r_0]} \text{IC}^\mathcal{D}(k)$ , and the second inequality holds automatically, which together show the desired result.

(ii) The proof is similar to part (i). Define the following events:  $\tilde{E}_1 = \{\text{IC}^\mathcal{F}(\hat{r}^\mathcal{F}) = \text{IC}^\mathcal{F}(r)\}$ ,  $\tilde{E}_{2k} = \{\text{IC}^\mathcal{F}(r) \leq \text{IC}^\mathcal{F}(k)\}$  for  $k \in [r_0]$ ,  $\tilde{E}_3 = \{\text{IC}^\mathcal{F}(r) < \min_{k \in [r-1]} \text{IC}^\mathcal{D}(k)\}$  and  $\tilde{E}_{4k} = \{\text{IC}^\mathcal{F}(k) < \text{IC}^\mathcal{D}(k)\}$  for  $r \leq k \leq r_0$ . As  $p, n \rightarrow \infty$ , we have  $\mathbb{P}(\tilde{E}_1) \rightarrow 1$  by Theorem 2(ii),  $\mathbb{P}(\tilde{E}_{2k}) \rightarrow 1$  by Lemma B.25(ii),  $\mathbb{P}(\tilde{E}_3) \rightarrow 1$  by the first argument of Lemma B.24(ii) and  $g^\mathcal{D}(p, n) = o(1), g^\mathcal{F}(p, n) = o(1)$ , and  $\mathbb{P}(\tilde{E}_{4k}) \rightarrow 1$  by the second argument of Lemma B.24(ii) and  $g^\mathcal{D}(p, n) = o(1), g^\mathcal{F}(p, n) = o(1)$ . Let  $\tilde{E}_2 = \bigcap_{k=1}^{r_0} \tilde{E}_{2k}$  and  $\tilde{E}_4 = \bigcap_{k=r}^{r_0} \tilde{E}_{4k}$ . Then, under the event  $\tilde{E} = \tilde{E}_1 \cap \tilde{E}_2 \cap \tilde{E}_3 \cap \tilde{E}_4$  with  $\mathbb{P}(\tilde{E}) \rightarrow 1$ , it follows that

$$\text{IC}^\mathcal{F}(\hat{r}^\mathcal{F}) = \text{IC}^\mathcal{F}(r) = \min_{k \in [r_0]} \text{IC}^\mathcal{F}(k) < \min_{k \in [r_0]} \text{IC}^\mathcal{D}(k) \leq \text{IC}^\mathcal{D}(\hat{r}^\mathcal{D}),$$

where the first equality holds under  $\tilde{E}_1$ , the second equality holds under  $\tilde{E}_2$ , the first inequality holds under  $\tilde{E}_3 \cap \tilde{E}_4$  since  $\min_{k \in [r_0]} \text{IC}^\mathcal{D}(k) = \min_{k \in [r-1]} \text{IC}^\mathcal{D}(k) \wedge \text{IC}^\mathcal{D}(r) \wedge \text{IC}^\mathcal{D}(r+1) \wedge \cdots \wedge \text{IC}^\mathcal{D}(r_0) > \text{IC}^\mathcal{F}(r) \wedge \text{IC}^\mathcal{F}(r+1) \wedge \cdots \wedge \text{IC}^\mathcal{F}(r_0) \geq \min_{k \in [r_0]} \text{IC}^\mathcal{F}(k)$ , and the second inequality holds automatically, which together show the desired result.

## B.6 Proof of Theorem 4

To prove Theorem 4, we first present some technical lemmas with their proofs.

**Lemma B.26.** *Under the assumptions of Theorem 4, it holds that*

- (i)  $\max_{i \in [p]} n^{-1} \sum_{t=1}^n \|\hat{\varepsilon}_{ti} - \varepsilon_{ti}\|^2 = O_p(\varpi_{n,p}^2)$ ;
- (ii)  $\max_{i,j \in [p]} \|n^{-1} \sum_{t=1}^n \hat{\varepsilon}_{ti} \hat{\varepsilon}_{tj} - n^{-1} \sum_{t=1}^n \varepsilon_{ti} \varepsilon_{tj}\|_S = O_p(\varpi_{n,p})$ ;
- (iii)  $\|\hat{\Sigma}_\varepsilon - \Sigma_\varepsilon\|_{S, \max} = O_p(\varpi_{n,p})$ .

*Proof.* (i) Notice that  $\hat{\varepsilon}_{ti}(\cdot) - \varepsilon_{ti}(\cdot) = \{y_{ti}(\cdot) - \check{\mathbf{b}}_i^\top \mathbf{f}_t(\cdot)\} - \{y_{ti}(\cdot) - \check{\mathbf{b}}_i^\top \hat{\mathbf{f}}_t(\cdot)\} = (\check{\mathbf{b}}_i - \mathbf{U} \check{\mathbf{b}}_i)^\top \hat{\mathbf{f}}_t(\cdot) - \check{\mathbf{b}}_i^\top (\mathbf{U}^\top \hat{\mathbf{f}}_t - \mathbf{f}_t)(\cdot)$ , where  $\check{\mathbf{b}}_i$  and  $\check{\mathbf{b}}_i$  are the  $i$ -th rows of  $\mathbf{B}$  and  $\hat{\mathbf{B}}$ , respectively. Applying the inequality  $(a+b)^2 \leq 2(a^2 + b^2)$  and the Cauchy-Schwarz inequality yields that

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

(ii) Notice that  $\max_{i \in [p]} \mathbb{E} \|\varepsilon_{ti}\|^2 = \max_{i \in [p]} \mathbb{E} \int \varepsilon_{ti}(u)^2 du = \max_{i \in [p]} \int \Sigma_{\varepsilon, ii}(u, u) du = O(1)$  from Assump-

tion 3(iv), thus we have  $\max_{i \in [p]} n^{-1} \sum_{t=1}^n \|\varepsilon_{ti}\|^2 = O_p(1)$ . By the Cauchy-Schwarz inequality,

$$\begin{aligned} \max_{i,j \in [p]} \left\| \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_{ti} \hat{\varepsilon}_{tj} - \frac{1}{n} \sum_{t=1}^n \varepsilon_{ti} \varepsilon_{tj} \right\|_S &= \max_{i,j \in [p]} \left\| \frac{1}{n} \sum_{t=1}^n (\hat{\varepsilon}_{ti} - \varepsilon_{ti}) \hat{\varepsilon}_{tj} + \varepsilon_{ti} (\hat{\varepsilon}_{tj} - \varepsilon_{tj}) \right\|_S \\ &\leq \max_{i \in [p]} \frac{1}{n} \sum_{t=1}^n \|\hat{\varepsilon}_{ti} - \varepsilon_{ti}\|^2 + 2 \left( \max_{i \in [p]} \frac{1}{n} \sum_{t=1}^n \|\varepsilon_{ti}\|^2 \right)^{1/2} \left( \max_{j \in [p]} \frac{1}{n} \sum_{t=1}^n \|\hat{\varepsilon}_{tj} - \varepsilon_{tj}\|^2 \right)^{1/2} \\ &= O_p(\varpi_{n,p}^2) + O_p(\varpi_{n,p}) = O_p(\varpi_{n,p}). \end{aligned}$$

(iii) The result is immediately implied by part (ii) above and Lemma B.12(ii).  $\square$

**Lemma B.27.** *Under the assumptions of Theorem 4, there exist some constants  $\Theta_1, \Theta_2 > 0$  such that with probability approaching one,*

$$\Theta_1 \leq \min_{i \in [p], j \in [p]} \|\hat{\Theta}_{ij}^{1/2}\|_S \leq \max_{i \in [p], j \in [p]} \|\hat{\Theta}_{ij}^{1/2}\|_S \leq \Theta_2.$$

*Proof.* See the supplementary material.  $\square$

We are now ready to prove Theorem 4. By Lemmas B.26(iii) and B.27, we have  $\|\hat{\Sigma}_\varepsilon - \Sigma_\varepsilon\|_{S, \max} = O_p(\varpi_{n,p})$  and  $\max_{i,j \in [p]} \|\Theta_{ij}\|_S = O_p(1)$ . Consequently, for any  $\epsilon > 0$ , there exist some positive constants  $N, \Theta_1$  and  $\Theta_2$  such that each of events

$$\Upsilon_1 = \left\{ \max_{i \in [p], j \in [p]} \|\hat{\Sigma}_{\varepsilon,ij} - \Sigma_{\varepsilon,ij}\|_S < N\varpi_{n,p} \right\}, \quad \Upsilon_2 = \left\{ \Theta_1 \leq \|\hat{\Theta}_{ij}^{1/2}\|_S \leq \Theta_2, \text{ all } i, j \in [p] \right\}$$

hold with probability at least  $1 - \epsilon$ . The thresholding in (10) is equivalent to  $\hat{\Sigma}_{\varepsilon,ij}^A = s_{ij}(\hat{\Sigma}_{\varepsilon,ij})$ , where  $s_{ij}(\cdot) \equiv s_{\lambda_{ij}}(\cdot)$  with  $\lambda_{ij} = \dot{C}\omega_{n,p}\|\hat{\Theta}_{ij}^{1/2}\|_S$  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} \|\hat{\Sigma}_\varepsilon^A - \Sigma_\varepsilon\|_{S,1} &= \max_{i \in [p]} \sum_{j=1}^p \|\hat{\Sigma}_{\varepsilon,ij}^A - \Sigma_{\varepsilon,ij}\|_S = \max_{i \in [p]} \sum_{j=1}^p \|s_{ij}(\hat{\Sigma}_{\varepsilon,ij}) - \Sigma_{\varepsilon,ij}\|_S \\ &\leq \max_{i \in [p]} \sum_{j=1}^p \|s_{ij}(\hat{\Sigma}_{\varepsilon,ij}) - \hat{\Sigma}_{\varepsilon,ij}\|_S I(\|\hat{\Sigma}_{\varepsilon,ij}\|_S > \dot{C}\omega_{n,p}\|\hat{\Theta}_{ij}^{1/2}\|_S) + \max_{i \in [p]} \sum_{j=1}^p \|\hat{\Sigma}_{\varepsilon,ij} - \Sigma_{\varepsilon,ij}\|_S I(\|\hat{\Sigma}_{\varepsilon,ij}\|_S > \dot{C}\omega_{n,p}\|\hat{\Theta}_{ij}^{1/2}\|_S) \\ &\quad + \max_{i \in [p]} \sum_{j=1}^p \|\Sigma_{\varepsilon,ij}\|_S I(\|\hat{\Sigma}_{\varepsilon,ij}\|_S \leq \dot{C}\omega_{n,p}\|\hat{\Theta}_{ij}^{1/2}\|_S) \\ &\leq \max_{i \in [p]} \sum_{j=1}^p \lambda_{ij} I(\|\hat{\Sigma}_{\varepsilon,ij}\|_S > \dot{C}\omega_{n,p}\Theta_1) + \max_{i \in [p]} \sum_{j=1}^p N\varpi_{n,p} I(\|\hat{\Sigma}_{\varepsilon,ij}\|_S > \dot{C}\omega_{n,p}\Theta_1) + \max_{i \in [p]} \sum_{j=1}^p \|\Sigma_{\varepsilon,ij}\|_S I(\|\hat{\Sigma}_{\varepsilon,ij}\|_S \leq \dot{C}\omega_{n,p}\Theta_2) \\ &\leq (\dot{C}\Theta_2 + N)\varpi_{n,p} \max_{i \in [p]} \sum_{j=1}^p I(\|\Sigma_{\varepsilon,ij}\|_S > N\varpi_{n,p}) + \max_{i \in [p]} \sum_{j=1}^p \|\Sigma_{\varepsilon,ij}\|_S I(\|\Sigma_{\varepsilon,ij}\|_S \leq (\dot{C}\Theta_2 + N)\varpi_{n,p}) \\ &\leq (\dot{C}\Theta_2 + N)\varpi_{n,p} \max_{i \in [p]} \sum_{j=1}^p \frac{\|\Sigma_{\varepsilon,ij}\|_S^q}{N^q \varpi_{n,p}^q} I(\|\Sigma_{\varepsilon,ij}\|_S > N\varpi_{n,p}) \\ &\quad + \max_{i \in [p]} \sum_{j=1}^p \|\Sigma_{\varepsilon,ij}\|_S \frac{(\dot{C}\Theta_2 + N)^{1-q} \varpi_{n,p}^{1-q}}{\|\Sigma_{\varepsilon,ij}\|_S^{1-q}} I(\|\Sigma_{\varepsilon,ij}\|_S \leq (\dot{C}\Theta_2 + N)\varpi_{n,p}) \\ &\leq (\dot{C}\Theta_2 + N) \{N^{-q} + (\dot{C}\Theta_2 + N)^{-q}\} \varpi_{n,p}^{1-q} \max_{i \in [p]} \sum_{j=1}^p \|\Sigma_{\varepsilon,ij}\|_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  $\max_{i \in [p]} \|\sigma_i\|_{\mathcal{N}} = \max_{i \in [p]} \int \Sigma_{\varepsilon}(u, u) du = O(1)$  by Assumption 3(iv). Therefore, with probability at least  $1 - 2\epsilon$ ,  $\|\hat{\Sigma}_{\varepsilon}^A - \Sigma_{\varepsilon}\|_{\mathcal{S},1} \lesssim \varpi_{n,p}^{1-q} s_p$ . Considering that  $\epsilon > 0$  can be arbitrarily small, we have the desired result

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

## B.7 Proof of Theorem 5

To prove Theorem 5, we first present a technical lemma with its proof.

**Lemma B.28.** *Suppose that the assumptions of Theorem 5 hold. For the sample covariance of  $\hat{\mathbf{f}}_t$ , i.e.,  $\hat{\Sigma}_f(u, v) = n^{-1} \sum_{t=1}^n \hat{\mathbf{f}}_t(u) \hat{\mathbf{f}}_t(v)^T$ , we have*

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

*Proof.* Consider  $\hat{\mathbf{f}}_t(u) \hat{\mathbf{f}}_t(v)^T - \mathbf{U} \mathbf{f}_t(u) \mathbf{f}_t(v)^T \mathbf{U}^T = \{\hat{\mathbf{f}}_t(u) - \mathbf{U} \mathbf{f}_t(u)\} \hat{\mathbf{f}}_t(v)^T + \mathbf{U} \mathbf{f}_t(u) \{\hat{\mathbf{f}}_t(v) - \mathbf{U} \mathbf{f}_t(v)\}^T$ . Then

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

where the second inequality follows from the Cauchy–Schwarz inequality, and the last line follows from  $n^{-1} \sum_{t=1}^n \|\hat{\mathbf{f}}_t - \mathbf{U} \mathbf{f}_t\|^2 = O_p(\mathcal{M}_{\varepsilon}^2/n + 1/p)$  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 B.12(i), the desired result follows immediately.  $\square$

We are now ready to prove Theorem 5. Consider that

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

Then we have

$$\begin{aligned} &\|\mathbf{B} \Sigma_f \mathbf{B}^T - \hat{\mathbf{B}} \hat{\Sigma}_f \hat{\mathbf{B}}^T\|_{\mathcal{S}, \max} \\ &\leq \|\mathbf{B} \mathbf{U}^T\|_{\infty} \|\mathbf{U} \Sigma_f \mathbf{U}^T - \hat{\Sigma}_f\|_{\mathcal{S}, \max} \|\mathbf{U} \mathbf{B}^T\|_1 + 2\|\mathbf{B} \mathbf{U}^T - \hat{\mathbf{B}}\|_{\infty} (\|\mathbf{U} \Sigma_f \mathbf{U}^T\|_{\mathcal{S}, \max} + \|\mathbf{U} \Sigma_f \mathbf{U}^T - \hat{\Sigma}_f\|_{\mathcal{S}, \max}) \|\mathbf{B} \mathbf{U}^T\|_{\infty} \\ &\leq r^3 C^2 \|\mathbf{U} \Sigma_f \mathbf{U}^T - \hat{\Sigma}_f\|_{\mathcal{S}, \max} + 2r^{5/2} C (r C_{\max} + \|\mathbf{U} \Sigma_f \mathbf{U}^T - \hat{\Sigma}_f\|_{\mathcal{S}, \max}) \|\mathbf{B} \mathbf{U}^T - \hat{\mathbf{B}}\|_{\max} \\ &= O_p(\mathcal{M}_{\varepsilon}/\sqrt{n} + 1/\sqrt{p}) + O_p(\varpi_{n,p}) = O_p(\varpi_{n,p}), \end{aligned} \tag{B.4}$$

where the first inequality follows from Lemma A.9(i), the second inequality follows from  $\|\mathbf{B} \mathbf{U}^T\|_{\infty} \leq r \|\mathbf{B}\|_{\max} \|\mathbf{U}\|_{\infty} \leq r^{3/2} C$  provided that  $\|\mathbf{U}\|_{\infty} \leq \sqrt{r} \|\mathbf{U}\| = \sqrt{r}$ ,  $\|\mathbf{B} \mathbf{U}^T - \hat{\mathbf{B}}\|_{\infty} \leq r \|\mathbf{B} \mathbf{U}^T - \hat{\mathbf{B}}\|_{\max}$  and  $\|\mathbf{U} \Sigma_f \mathbf{U}^T\|_{\mathcal{S}, \max} \leq C_{\max} \|\mathbf{U}\|_{\infty}^2 \leq r C_{\max}$  in Lemma B.17(i), and the last line follows from Lemma

B.28 and  $\|\mathbf{B}\mathbf{U}^\top - \hat{\mathbf{B}}\|_{\max} = O_p(\varpi_{n,p})$  in Theorem 1(i). Then note that

$$\|\hat{\Sigma}_\varepsilon^{\mathcal{A}} - \Sigma_\varepsilon\|_{\mathcal{S},\max} \leq \|\hat{\Sigma}_\varepsilon^{\mathcal{A}} - \hat{\Sigma}_\varepsilon\|_{\mathcal{S},\max} + \|\hat{\Sigma}_\varepsilon - \Sigma_\varepsilon\|_{\mathcal{S},\max} \leq \max_{i,j \in [p]} (\|\hat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}} \lambda) + O_p(\varpi_{n,p}) = O_p(\varpi_{n,p}),$$

where the last line follows from Lemma B.26(iii), the choice of  $\lambda = \dot{C}(\sqrt{\log p/n} + \sqrt{1/p}) \lesssim \varpi_{n,p}$ , and the fact  $\max_{i,j \in [p]} \|\hat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}} = O_p(1)$  by Lemma B.27. By combining (5), (12), and (B.4), we obtain the desired result.

## B.8 Proof of Theorem 6

For the sake of brevity, in this section, we suppose that the orthogonal matrix  $\mathbf{U}$  in Theorem 1 and Lemmas B.23–B.28 is an identity matrix, which means, when we perform eigen-decomposition on  $\hat{\Omega}$ , we can always select the correct direction of  $\hat{\xi}_j$  to ensure  $\hat{\xi}_j^\top \tilde{\mathbf{b}}_j \geq 0$ . The proofs of Theorems 4 and 5 verify that the choice of  $\mathbf{U}$  does not affect the theoretical results. In this section,  $(\hat{\Sigma}_\varepsilon^{\mathcal{A}})^\dagger$ ,  $\hat{\Sigma}_f^\dagger$  and  $\{\hat{\Sigma}_f^\dagger + \hat{\mathbf{B}}^\top (\hat{\Sigma}_\varepsilon^{\mathcal{A}})^\dagger \hat{\mathbf{B}}\}^\dagger$  are defined in a similar way to the inverse in Lemma B.15 associated with  $\Sigma_\varepsilon$ ,  $\Sigma_f$  and  $\Sigma_f^\dagger + \mathbf{B}^\top \Sigma_\varepsilon^\dagger \mathbf{B}$ , respectively. To prove Theorem 6, we first present some technical lemmas with their proofs.

**Lemma B.29.** *Under the assumptions of Theorem 6, then,  $\hat{\Sigma}_\varepsilon^{\mathcal{A}}$  has a bounded Moore–Penrose inverse with probability approaching one, and  $\|(\hat{\Sigma}_\varepsilon^{\mathcal{A}})^\dagger - \Sigma_\varepsilon^\dagger\|_{\mathcal{L}} = O_p(\varpi_{n,p}^{1-q} s_p)$ .*

*Proof.* Provided that  $\varpi_{n,p}^{1-q} s_p = o(1)$  and  $\|\Sigma_\varepsilon^\dagger\|_{\mathcal{L}} < c_3$  for some constant  $c_3 > 0$ , we combine Lemma B.15 and Theorem 4 to yield that  $\|(\hat{\Sigma}_\varepsilon^{\mathcal{A}})^\dagger\|_{\mathcal{L}} < 2c_3$  with probability approaching one, and thus  $\hat{\Sigma}_\varepsilon^{\mathcal{A}}$  has a bounded Moore–Penrose inverse with probability approaching one together with the desired result  $\|(\hat{\Sigma}_\varepsilon^{\mathcal{A}})^\dagger - \Sigma_\varepsilon^\dagger\|_{\mathcal{L}} = O_p(\varpi_{n,p}^{1-q} s_p)$ .  $\square$

**Lemma B.30.** *Under the assumptions of Theorem 6,*

$$\|\hat{\mathbf{B}}^\top (\hat{\Sigma}_\varepsilon^{\mathcal{A}})^\dagger \hat{\mathbf{B}} - \mathbf{B}^\top \Sigma_\varepsilon^\dagger \mathbf{B}\|_{\mathcal{L}} = O_p(p\varpi_{n,p}^{1-q} s_p) = o_p(p).$$

*Proof.* Consider

$$\begin{aligned} \|\hat{\mathbf{B}}^\top (\hat{\Sigma}_\varepsilon^{\mathcal{A}})^\dagger \hat{\mathbf{B}} - \mathbf{B}^\top \Sigma_\varepsilon^\dagger \mathbf{B}\|_{\mathcal{L}} &\leq 2\|(\hat{\mathbf{B}} - \mathbf{B})^\top (\hat{\Sigma}_\varepsilon^{\mathcal{A}})^\dagger \hat{\mathbf{B}}\|_{\mathcal{L}} + \|\mathbf{B}^\top \{(\hat{\Sigma}_\varepsilon^{\mathcal{A}})^\dagger - \Sigma_\varepsilon^\dagger\} \mathbf{B}\|_{\mathcal{L}} \\ &\leq 2\|\hat{\mathbf{B}} - \mathbf{B}\| \|\hat{\Sigma}_\varepsilon^{\mathcal{A}}\|_{\mathcal{L}} \|\hat{\mathbf{B}}\|_{\mathcal{L}} + \|\mathbf{B}\|^2 \|(\hat{\Sigma}_\varepsilon^{\mathcal{A}})^\dagger - \Sigma_\varepsilon^\dagger\|_{\mathcal{L}} \\ &= O_p(p\varpi_{n,p}) + O_p(p\varpi_{n,p}^{1-q} s_p) = O_p(p\varpi_{n,p}^{1-q} s_p) = o_p(p), \end{aligned}$$

where the last line follows from Lemmas B.23 and B.29.  $\square$

**Lemma B.31.** *Under the assumptions of Theorem 6,*

- (i)  $\|(\Sigma_f^\dagger + \mathbf{B}^\top \Sigma_\varepsilon^\dagger \mathbf{B})^\dagger\|_{\mathcal{L}} = O(p^{-1})$ ;
- (ii)  $\|\{\hat{\Sigma}_f^\dagger + \hat{\mathbf{B}}^\top (\hat{\Sigma}_\varepsilon^{\mathcal{A}})^\dagger \hat{\mathbf{B}}\}^\dagger\|_{\mathcal{L}} = O_p(p^{-1})$ .

*Proof.* (i) Note that  $\|(\Sigma_f^\dagger + \mathbf{B}^\top \Sigma_\varepsilon^\dagger \mathbf{B})^\dagger\|_{\mathcal{L}} \leq \|(\mathbf{B}^\top \Sigma_\varepsilon^\dagger \mathbf{B})^\dagger\|_{\mathcal{L}} \leq \{\lambda_{\min}(\mathbf{B}^\top \mathbf{B})\}^{-1} \|\Sigma_\varepsilon\|_{\mathcal{L}} = O(p^{-1})$ , where the first inequality follows from the fact  $\Sigma_f$  is a Mercer’s kernel.

(ii) Since  $\|\Sigma_f^\dagger\|_{\mathcal{L}} < c_4$  and  $\|\hat{\Sigma}_f - \Sigma_f\|_{\mathcal{L}} = O_p(\varpi_{n,p}) = o_p(1)$ , by Lemma B.15, we have  $\|\hat{\Sigma}_f^\dagger - \Sigma_f^\dagger\|_{\mathcal{L}} = O_p(\varpi_{n,p})$ . Thus, by Lemma B.30,  $\|\{\hat{\Sigma}_f^\dagger + \hat{\mathbf{B}}^\top (\hat{\Sigma}_\varepsilon^{\mathcal{A}})^\dagger \hat{\mathbf{B}}\}^\dagger - \{\Sigma_f^\dagger + \mathbf{B}^\top \Sigma_\varepsilon^\dagger \mathbf{B}\}^\dagger\|_{\mathcal{L}} = o_p(p)$ . Combining Lemma B.15 and part (i), we obtain that  $\|\{\hat{\Sigma}_f^\dagger + \hat{\mathbf{B}}^\top (\hat{\Sigma}_\varepsilon^{\mathcal{A}})^\dagger \hat{\mathbf{B}}\}^\dagger\|_{\mathcal{L}} = O_p(p^{-1})$ .  $\square$

We are now ready to prove Theorem 6. Using the functional version of the Sherman–Morrison–Woodbury formula, we have  $\|(\widehat{\Sigma}_y^{\mathcal{D}})^{\dagger} - \Sigma_y^{\dagger}\|_{\mathcal{L}} \leq \sum_{k=1}^4 L_k$ , where

$$\begin{aligned} L_1 &= \|(\widehat{\Sigma}_{\varepsilon}^{\mathcal{A}})^{\dagger} - \Sigma_{\varepsilon}^{\dagger}\|_{\mathcal{L}}, \\ L_2 &= \|\{(\widehat{\Sigma}_{\varepsilon}^{\mathcal{A}})^{\dagger} \widehat{\mathbf{B}} - \Sigma_{\varepsilon}^{\dagger} \mathbf{B}\} \{\widehat{\Sigma}_f^{\dagger} + \widehat{\mathbf{B}}^{\top} (\widehat{\Sigma}_{\varepsilon}^{\mathcal{A}})^{\dagger} \widehat{\mathbf{B}}\}^{\dagger} \widehat{\mathbf{B}}^{\top} (\widehat{\Sigma}_{\varepsilon}^{\mathcal{A}})^{\dagger}\|_{\mathcal{L}}, \\ L_3 &= \|\Sigma_{\varepsilon}^{\dagger} \mathbf{B} \{\widehat{\Sigma}_f^{\dagger} + \widehat{\mathbf{B}}^{\top} (\widehat{\Sigma}_{\varepsilon}^{\mathcal{A}})^{\dagger} \widehat{\mathbf{B}}\}^{\dagger} \{\widehat{\mathbf{B}}^{\top} (\widehat{\Sigma}_{\varepsilon}^{\mathcal{A}})^{\dagger} - \mathbf{B}^{\top} \Sigma_{\varepsilon}^{\dagger}\}\|_{\mathcal{L}}, \\ L_4 &= \|\Sigma_{\varepsilon}^{\dagger} \mathbf{B} [\{\widehat{\Sigma}_f^{\dagger} + \widehat{\mathbf{B}}^{\top} (\widehat{\Sigma}_{\varepsilon}^{\mathcal{A}})^{\dagger} \widehat{\mathbf{B}}\}^{\dagger} - \{\Sigma_f^{\dagger} + \mathbf{B}^{\top} \Sigma_{\varepsilon}^{\dagger} \mathbf{B}\}^{\dagger}] \mathbf{B}^{\top} \Sigma_{\varepsilon}^{\dagger}\|_{\mathcal{L}}. \end{aligned}$$

Clearly,  $L_1 = O_p(\varpi_{n,p}^{1-q} s_p)$  by Lemma B.29. Then, note that  $\|(\widehat{\Sigma}_{\varepsilon}^{\mathcal{A}})^{\dagger} \widehat{\mathbf{B}} - \Sigma_{\varepsilon}^{\dagger} \mathbf{B}\|_{\mathcal{L}} \leq \|(\widehat{\Sigma}_{\varepsilon}^{\mathcal{A}})^{\dagger} - \Sigma_{\varepsilon}^{\dagger}\|_{\mathcal{L}} \|\widehat{\mathbf{B}}\| + \|\Sigma_{\varepsilon}^{\dagger}\|_{\mathcal{L}} \|\widehat{\mathbf{B}} - \mathbf{B}\| = O_p(\sqrt{p} \varpi_{n,p}^{1-q} s_p)$ . From Lemma B.31, we obtain that  $L_2 \asymp L_3 = O_p(\varpi_{n,p}^{1-q} s_p)$ . Lastly, since  $\|(\Sigma_f^{\dagger} + \mathbf{B}^{\top} \Sigma_{\varepsilon}^{\dagger} \mathbf{B})^{\dagger}\|_{\mathcal{L}} = O(p^{-1})$  and  $\|\{\widehat{\Sigma}_f^{\dagger} + \widehat{\mathbf{B}}^{\top} (\widehat{\Sigma}_{\varepsilon}^{\mathcal{A}})^{\dagger} \widehat{\mathbf{B}}\} - \{\Sigma_f^{\dagger} + \mathbf{B}^{\top} \Sigma_{\varepsilon}^{\dagger} \mathbf{B}\}\|_{\mathcal{L}} = O_p(p \varpi_{n,p}^{1-q} s_p) = o_p(p)$ , we apply Lemma B.15 to obtain that

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

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

## References

- Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors, *Econometrica* **81**: 1203–1227.
- 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.
- Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models, *The Annals of Statistics* **43**: 1535–1567.
- 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.
- Chang, J., Chen, C., Qiao, X. and Yao, Q. (2024). An autocovariance-based learning framework for high-dimensional functional time series, *Journal of Econometrics* **239**: 105385.
- Chen, C., Guo, S. and Qiao, X. (2022). Functional linear regression: dependence and error contamination, *Journal of Business & Economic Statistics* **40**: 444–457.
- Chiou, J.-M., Chen, Y.-T. and Yang, Y.-F. (2014). Multivariate functional principal component analysis: A normalization approach, *Statistica Sinica* **24**: 1571–1596.
- 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., Li, K. and Liao, Y. (2021). Recent developments in factor models and applications in econometric learning, *Annual Review of Financial Economics* **13**: 401–430.
- 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., Wang, W. and Zhong, Y. (2018). An  $\ell_\infty$  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. (2024). Adaptive functional thresholding for sparse covariance function estimation in high dimensions, *Journal of the American Statistical Association* **119**: 1473–1485.
- 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., Wang, Q. and Wang, Z. (2025). Factor modelling for high-dimensional functional time series, *Journal of Business & Economic Statistics* pp. 1–24.
- Hallin, M., Nisol, G. and Tavakoli, S. (2023). Factor models for high-dimensional functional time series I: Representation results, *Journal of Time Series Analysis* **44**: 578–600.
- Han, Y., Chen, R. and Zhang, C.-H. (2022). Rank determination in tensor factor model, *Electronic Journal of Statistics* **16**: 1726–1803.
- 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.
- 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.
- Leng, C., Li, D., Shang, H. L. and Xia, Y. (2024). Covariance function estimation for high-dimensional functional time series with dual factor structures, *arXiv preprint arXiv:2401.05784*.
- Lou, D., Polk, C. and Skouras, S. (2019). A tug of war: Overnight versus intraday expected returns, *Journal of Financial Economics* **134**: 192–213.
- Solea, E. and Li, B. (2022). Copula Gaussian graphical models for functional data, *Journal of the American Statistical Association* **117**: 781–793.
- 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* **44**: 601–621.
- 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.

- Weyl, H. (1912). Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung), *Mathematische Annalen* **71**: 441–479.
- Xue, K., Yang, J. and Yao, F. (2024). Optimal linear discriminant analysis for high-dimensional functional data, *Journal of the American Statistical Association* **119**: 1055–1064.
- Yu, Y., Wang, T. and Samworth, R. J. (2015). A useful variant of the Davis–Kahan theorem for statisticians, *Biometrika* **102**: 315–323.
- Zhang, D. and Wu, W. B. (2021). Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes, *The Annals of Statistics* **49**: 233–254.
- Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data, *The Annals of Statistics* **35**: 1052–1079.
- 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 the proofs of the technical lemmas in the appendix in Section S.1, and the remaining technical proofs in Sections S.2–S.3, further derivations in Section S.5, additional simulation results in Section S.6 and additional real data result in Section S.7.

## S.1 Proofs of theoretical lemmas in the appendix

### S.1.1 Proof of Lemma A.5

(i) Note that  $\iint \mathbf{K}(u, v) \mathbf{K}(u, v)^\top du dv = \int \sum_{i=1}^{\infty} \lambda_i^2 \phi_i(u) \phi_i(u)^\top du$ , and thus

$$\text{tr} \left\{ \iint \mathbf{K}(u, v) \mathbf{K}(u, v)^\top du dv \right\} = \text{tr} \left\{ \sum_{i=1}^{\infty} \lambda_i^2 \int \phi_i(u)^\top \phi_i(u) du \right\} = \sum_{i=1}^{\infty} \lambda_i^2.$$

The equality  $\text{tr} \left\{ \iint \mathbf{K}(u, v) \mathbf{K}(u, v)^\top du dv \right\} = \|\mathbf{K}\|_{\mathcal{S}, \mathcal{F}}^2$  can be verified by simple calculation. The first equality can be obtained by  $\mathbf{K}(u, v)^\top = \mathbf{K}^\top(v, u)$  and the multiplication of kernel functions.

(ii) Similarly,

$$\begin{aligned} \left\| \iint \mathbf{K}(u, v) \mathbf{K}(u, v)^\top du dv \right\| &= \left\| \int \sum_{i=1}^{\infty} \lambda_i^2 \phi_i(u) \phi_i(u)^\top du \right\| = \lambda_{\max} \left\{ \int \sum_{i=1}^{\infty} \lambda_i^2 \phi_i(u) \phi_i(u)^\top du \right\} \\ &= \lambda_{\max} \left\{ \mathbf{\Lambda}^2 \int \mathbf{\Phi}(u) \mathbf{\Phi}(u)^\top du \right\} = \lambda_{\max} \left\{ \mathbf{\Lambda}^2 \int \mathbf{\Phi}(u)^\top \mathbf{\Phi}(u) du \right\} \\ &= \lambda_{\max}(\mathbf{\Lambda}^2) = \lambda_1^2 = \|\mathbf{K}\|_{\mathcal{L}}^2, \end{aligned}$$

where  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots)$ ,  $\mathbf{\Phi}(\cdot) = \{\phi_1(\cdot), \phi_2(\cdot), \dots\}$ , and the fact that the  $\infty \times \infty$  matrix  $\int \mathbf{\Phi}(u)^\top \mathbf{\Phi}(u) du$  shares the same nonzero eigenvalues with the  $p \times p$  matrix  $\int \mathbf{\Phi}(u) \mathbf{\Phi}(u)^\top du$ , which can be obtained following the proof of Proposition 2 in Bathia et al. (2010).

(iii) The equality holds by the definition of trace norm.

### S.1.2 Proof of Lemma A.6

(i) Note that

$$\begin{aligned} \left\| \iint \mathbf{K}(u, v) \mathbf{G}(u, v)^\top du dv \right\|_1 &= \max_{j \in [p]} \sum_{i=1}^p \left| \iint \sum_{k=1}^p K_{ik}(u, v) G_{jk}(u, v) du dv \right| \\ &\leq \max_{j \in [p]} \sum_{i=1}^p \sum_{k=1}^p \|K_{ik}\|_S \|G_{jk}\|_S \\ &\leq \left( \max_{k \in [p]} \sum_{i=1}^p \|K_{ik}\|_S \right) \left( \max_{j \in [p]} \sum_{k=1}^p \|G_{jk}\|_S \right) \\ &= \|\mathbf{K}\|_{S, 1} \|\mathbf{G}\|_{S, \infty}. \end{aligned} \tag{S.1}$$

(ii) By similar arguments, we obtain that

$$\left\| \int \int \mathbf{K}(u, v) \mathbf{G}(u, v)^T du dv \right\|_{\infty} \leq \|\mathbf{K}\|_{S, \infty} \|\mathbf{G}\|_{S, 1}. \quad (\text{S.2})$$

(iii) The inequality follows immediately from (S.1), (S.2), the matrix norm inequality  $\|\mathbf{A}\|^2 \leq \|\mathbf{A}\|_{\infty} \|\mathbf{A}\|_1$  for any  $p \times p$  matrix  $\mathbf{A}$  and the choice of  $\mathbf{A} = \int \int \mathbf{K}(u, v) \mathbf{G}(u, v)^T du dv$ .

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

### S.1.3 Proof of Lemma A.7

(i) By Lemma B.14 or Theorem 4.2.5 in Hsing and Eubank (2015), we have

$$\begin{aligned} \|\Sigma \tilde{\Sigma}\|_{\mathcal{L}} &= \max_{\mathbf{x} \in \mathbb{H}^p} \frac{\langle \mathbf{x}, \Sigma \tilde{\Sigma}(\mathbf{x}) \rangle}{\|\mathbf{x}\|^2} = \max_{\mathbf{x} \in \mathbb{H}^p} \frac{\langle \tilde{\Sigma}^{1/2}(\mathbf{x}), \Sigma \{\tilde{\Sigma}^{1/2}(\mathbf{x})\} \rangle}{\|\tilde{\Sigma}^{1/2}(\mathbf{x})\|^2} \cdot \frac{\langle \mathbf{x}, \tilde{\Sigma}(\mathbf{x}) \rangle}{\|\mathbf{x}\|^2} \\ &\leq \max_{\mathbf{y} \in \mathbb{H}^p} \frac{\langle \mathbf{y}, \Sigma(\mathbf{y}) \rangle}{\|\mathbf{y}\|^2} \max_{\mathbf{x} \in \mathbb{H}^p} \frac{\langle \mathbf{x}, \tilde{\Sigma}(\mathbf{x}) \rangle}{\|\mathbf{x}\|^2} = \|\Sigma\|_{\mathcal{L}} \cdot \|\tilde{\Sigma}\|_{\mathcal{L}}. \end{aligned}$$

(ii) By Lemma A.5,  $\|\Sigma\|_{\mathcal{L}} = \lambda_1$  and  $\|\Sigma\|_{\text{F}} = \sqrt{\sum_{i \geq 1} \lambda_i^2}$ , where  $\{\lambda_i\}_{i=1}^{\infty}$  are the eigenvalues of  $\Sigma$  in a descending order. Apparently,  $\|\Sigma\|_{\mathcal{L}} \leq \|\Sigma\|_{S, \text{F}}$  holds.

(iii) By Lemmas A.5(ii) and A.6(iii),  $\|\Sigma\|_{\mathcal{L}}^2 = \left\| \int \int \Sigma(u, v) \Sigma(u, v)^T du dv \right\| \leq \|\Sigma\|_{S, 1} \|\Sigma\|_{S, \infty}$ . Furthermore, if  $\|\Sigma_{ij}\|_{\mathcal{S}} = \|\Sigma_{ji}\|_{\mathcal{S}}$  for all  $i, j \in [p]$ , we have  $\|\Sigma\|_{S, 1} = \|\Sigma\|_{S, \infty}$ , and thus  $\|\Sigma\|_{\mathcal{L}} \leq \|\Sigma\|_{S, 1}$  holds.

### S.1.4 Proof of Lemma A.8

(i) Note that

$$\begin{aligned} \text{tr} \left\{ \int \mathbf{K} \mathbf{G}(u, u) du \right\} &= \text{tr} \left\{ \int \int \mathbf{K}(u, v) \mathbf{G}(v, u) du dv \right\} = \int \int \text{tr} \{ \mathbf{K}(u, v) \mathbf{G}(v, u) \} du dv \\ &= \text{tr} \left\{ \int \int \mathbf{G}(v, u) \mathbf{K}(v, u) dv du \right\} = \text{tr} \left\{ \int \mathbf{G} \mathbf{K}(v, v) dv \right\}. \end{aligned}$$

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

$$\begin{aligned} \text{tr} \left\{ \int \mathbf{K} \mathbf{G}(u, u) du \right\} &= \text{tr} \left\{ \int \int \mathbf{K}(u, v) \mathbf{G}(v, u) du dv \right\} \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \lambda_i \omega_j \int \phi_i(v)^T \psi_j(v) dv \int \psi_j(u)^T \phi_i(u) du \\ &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \lambda_i \omega_j |\langle \phi_i, \psi_j \rangle|^2 \leq \sum_{i=1}^{\infty} \lambda_i \omega_i \\ &\leq \left( \max_i \lambda_i \right) \sum_{j=1}^{\infty} \omega_j = \|\mathbf{K}\|_{\mathcal{L}} \|\mathbf{G}\|_{\mathcal{N}} = \|\mathbf{K}\|_{\mathcal{L}} \text{tr} \left\{ \int \mathbf{G}(u, u) du \right\}, \end{aligned}$$



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

(iii) From Lemma [A.5](#)(i) and the part (ii) above, we have

$$\begin{aligned}\|\mathbf{KG}\|_{\mathcal{S},\mathbf{F}}^2 &= \text{tr}\left\{\int \mathbf{KGG}^T\mathbf{K}^T(u,u)du\right\} = \text{tr}\left\{\int \mathbf{K}^T\mathbf{KGG}^T(u,u)du\right\} \\ &\leq \|\mathbf{KK}^T\|_{\mathcal{L}}\text{tr}\left\{\int \mathbf{GG}^T(u,u)du\right\} = \|\mathbf{K}\|_{\mathcal{L}}^2\|\mathbf{G}\|_{\mathcal{S},\mathbf{F}}^2,\end{aligned}$$

which implies the desired result.

### S.1.5 Proof of Lemma [A.9](#)

(i) It follows that

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

Further,

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

(ii) It follows that

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

where the inequality follows from the Cauchy-Schwarz inequality. Furthermore,

$$\|\mathbf{KA}^T\|_{\mathcal{S},\mathbf{F}} = \|(\mathbf{KA}^T)^T\|_{\mathcal{S},\mathbf{F}} = \|\mathbf{AK}^T\|_{\mathcal{S},\mathbf{F}} \leq \|\mathbf{A}\|_{\mathbf{F}}\|\mathbf{K}\|_{\mathcal{S},\mathbf{F}} = \|\mathbf{K}\|_{\mathcal{S},\mathbf{F}}\|\mathbf{A}^T\|_{\mathbf{F}}.$$

(iii)&(iv) It follows that

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

Furthermore,

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

The other two arguments can be proved similarly.

### S.1.6 Proof of Lemma A.10

(i) By the definition, it follows that

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

(ii) By the Cauchy–Schwarz inequality,

$$\begin{aligned} \|K\|_S &= \left[ \int \int \{\mathbf{f}(u)^T \mathbf{g}(v)\}^2 du dv \right]^{1/2} = \left[ \int \int \left\{ \sum_{j=1}^r f_j(u) g_j(v) \right\}^2 du dv \right]^{1/2} \\ &\leq \left\{ \int \int \sum_{j=1}^r f_j(u)^2 \sum_{j=1}^r g_j(v)^2 du dv \right\}^{1/2} = \left\{ \int \mathbf{f}(u)^T \mathbf{f}(u) du \int \mathbf{g}(v)^T \mathbf{g}(v) dv \right\}^{1/2} \\ &= \|\mathbf{f}\| \cdot \|\mathbf{g}\|. \end{aligned}$$

### S.1.7 Proof of Lemma A.11

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

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

which indicates that  $p\vartheta_j$  is the eigenvalue of  $\mathbf{Q}(\cdot) \mathbf{Q}(\cdot)^T$  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  $j \leq r$  and  $k \geq r+1$ , we obtain that  $\int \{\mathbf{Q}(u) \mathbf{Q}(v)^T\} \tilde{\mathbf{q}}_k(v) dv = 0$  for  $k \geq r+1$ . Thus, the rest eigenvalues of  $\mathbf{Q}(\cdot) \mathbf{Q}(\cdot)^T$  are zero.

### S.1.8 Proof of Lemma B.12

(i)&(ii) See Theorem 2 and equations (2.15) in Fang et al. (2022) for the corresponding proofs.

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

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

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

$$\begin{aligned}
\mathcal{M}_y &= 2\pi \cdot \operatorname{ess\,sup}_{\theta \in [-\pi, \pi], \phi \in \mathbb{H}_0^p} \frac{\langle \phi, \mathbf{f}_{y, \theta}(\phi) \rangle}{\langle \phi, \Sigma_y(\phi) \rangle} \\
&= 2\pi \cdot \operatorname{ess\,sup}_{\theta \in [-\pi, \pi], \phi \in \mathbb{H}_0^p} \frac{\iint \phi(u)^\top \mathbf{f}_{y, \theta}(u, v) \phi(v) du dv}{\iint \phi(u)^\top \Sigma_y(u, v) \phi(v) du dv} \\
&\leq 2\pi \cdot \operatorname{ess\,sup}_{\theta \in [-\pi, \pi], \phi \in \mathbb{H}_0^p} \frac{\iint \phi(u)^\top \mathbf{B} \mathbf{f}_{f, \theta}(u, v) \mathbf{B}^\top \phi(v) du dv}{\iint \phi(u)^\top \mathbf{B} \Sigma_f(u, v) \mathbf{B}^\top \phi(v) du dv} + 2\pi \cdot \operatorname{ess\,sup}_{\theta \in [-\pi, \pi], \phi \in \mathbb{H}_0^p} \frac{\iint \phi(u)^\top \mathbf{f}_{\varepsilon, \theta}(u, v) \phi(v) du dv}{\iint \phi(u)^\top \Sigma_\varepsilon(u, v) \phi(v) du dv} \\
&\leq 2\pi \cdot \operatorname{ess\,sup}_{\theta \in [-\pi, \pi], \xi \in \mathbb{H}_0^r} \frac{\langle \xi, \mathbf{f}_{f, \theta}(\xi) \rangle}{\langle \xi, \Sigma_f(\xi) \rangle} + 2\pi \cdot \operatorname{ess\,sup}_{\theta \in [-\pi, \pi], \phi \in \mathbb{H}_0^p} \frac{\langle \phi, \mathbf{f}_{\varepsilon, \theta}(\phi) \rangle}{\langle \phi, \Sigma_\varepsilon(\phi) \rangle} \\
&\asymp \mathcal{M}_f + \mathcal{M}_\varepsilon \asymp \mathcal{M}_\varepsilon.
\end{aligned}$$

The other conditions imposed by Fang et al. (2022) 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.

### S.1.9 Proof of Lemma B.15

(i) Notice that  $\langle \mathbf{x}, \mathbf{K}(\mathbf{x}) \rangle = \iint \mathbf{x}(u)^\top \mathbf{K}(u, v) \mathbf{x}(v) du dv$ . Then for any  $\mathbf{x} \in \mathbb{H}^p$  with  $\|\mathbf{x}\| = 1$ , we have

$$\left| \iint \mathbf{x}(u)^\top \{ \mathbf{K}(u, v) - \mathbf{G}(u, v) \} \mathbf{x}(v) du dv \right| \leq \|\mathbf{x}\|^2 \cdot \|\mathbf{K} - \mathbf{G}\|_{\mathcal{L}} = o_p(c_n^{-1})$$

Thus, for a sufficiently large  $n$ ,  $\langle \mathbf{x}, \mathbf{K}(\mathbf{x}) \rangle \geq \langle \mathbf{x}, \mathbf{G}(\mathbf{x}) \rangle - c_n^{-1}/2 > c_n^{-1}/2$  for any  $\mathbf{x} \in \operatorname{Ker}(\mathbf{G})^\perp$  and  $\|\mathbf{x}\| = 1$  with probability approaching one, since  $\langle \mathbf{x}, \mathbf{G}(\mathbf{x}) \rangle \geq (\|\mathbf{G}^\dagger\|_{\mathcal{L}})^{-1} > c_n^{-1}$  by using Lemma B.14. Hence, by the definition of  $\mathbf{K}^\dagger$ , it follows that  $\|\mathbf{K}^\dagger\|_{\mathcal{L}} < 2c_n$  with probability approaching one.

(ii) Notice that

$$\begin{aligned}
\|\mathbf{K}^\dagger - \mathbf{G}^\dagger\|_{\mathcal{L}} &= \max_{\mathbf{x} \in \operatorname{Im}(\mathbf{G}), \|\mathbf{x}\|=1} \langle \mathbf{x}, \mathbf{K}^\dagger(\mathbf{x}) - \mathbf{G}^\dagger(\mathbf{x}) \rangle = \max_{\mathbf{x} \in \operatorname{Im}(\mathbf{G}), \|\mathbf{x}\|=1} \langle \mathbf{x}, \mathbf{K}^\dagger(\mathbf{G} - \mathbf{K})\mathbf{G}^\dagger(\mathbf{x}) \rangle \\
&\leq \|\mathbf{G}^\dagger\|_{\mathcal{L}} \cdot \max_{\mathbf{y} \in \operatorname{Ker}(\mathbf{G})^\perp, \|\mathbf{y}\|=1} \langle \mathbf{y}, \mathbf{K}^\dagger(\mathbf{G} - \mathbf{K})(\mathbf{y}) \rangle \leq \|\mathbf{K} - \mathbf{G}\|_{\mathcal{L}} \|\mathbf{G}^\dagger\|_{\mathcal{L}} \cdot \max_{\mathbf{z} \in \operatorname{Im}(\mathbf{G}), \|\mathbf{z}\|=1} \langle \mathbf{z}, \mathbf{K}^\dagger(\mathbf{z}) \rangle \\
&\leq \|\mathbf{K}^\dagger\|_{\mathcal{L}} \|\mathbf{K} - \mathbf{G}\|_{\mathcal{L}} \|\mathbf{G}^\dagger\|_{\mathcal{L}} = O_p(c_n^2) \|\mathbf{K} - \mathbf{G}\|_{\mathcal{L}},
\end{aligned}$$

where the last equality follows from part (i).

### S.1.10 Proof of Lemma B.16

The proofs of this lemma utilize the tail probabilities of sub-Gaussian process as presented in Section S.5.3.

(i) Notice that  $\{\mathbf{f}_t(\cdot)\}$  and  $\{\varepsilon_t(\cdot)\}$  follow the sub-Gaussian functional linear process (see Section S.5.3) by Assumption 4, with  $\mathbb{E}\|\mathbf{f}_t\|^2 = O(1)$  and  $\mathbb{E}\|p^{-1/2}\mathbf{B}^\top \varepsilon_t\|^2 = O(1)$  by Assumption 3(ii). By Bonferroni's method, it yields that, for each  $j \in [r]$  and any given  $\eta > 0$ ,

$$\mathbb{P} \left\{ \max_{t \in [n]} (\|f_{tj}\|^2 - \mathbb{E}\|f_{tj}\|^2) \geq \eta \right\} \leq \sum_{t=1}^n \mathbb{P} (\|f_{tj}\|^2 - \mathbb{E}\|f_{tj}\|^2 \geq \eta) \leq 2n \exp\{-c \min(\eta^2, \eta)\}, \quad (\text{S.3})$$

where  $c > 0$  is some constant and the second inequality follows from Lemma 5 of Fang et al. (2022). Letting  $\eta = \log n$ , (S.3) shows that  $\max_{t \in [n]} \|\mathbf{f}_t\|^2 = O_p(\log n) + \mathbb{E}\|\mathbf{f}_t\|^2$ , which implies that  $\max_{t \in [n]} \|\mathbf{f}_t\| =$

$O_p(\sqrt{\log n})$ . For the second argument, to satisfy Condition 10 in Fang et al. (2022) and apply their Lemma 5, alternatively we consider that, for each  $j \in [r]$  and any given  $\eta > 0$ ,

$$\mathbb{P} \left\{ \max_{t \in [n]} (p^{-1} \|\mathbf{b}_j^\top \boldsymbol{\varepsilon}_t\|^2 - p^{-1} \mathbb{E} \|\mathbf{b}_j^\top \boldsymbol{\varepsilon}_t\|^2) \geq \eta \right\} \leq 2n \exp\{-c \min(\eta^2, \eta)\},$$

which implies that  $\max_{t \in [n]} \|p^{-1/2} \mathbf{B}^\top \boldsymbol{\varepsilon}_t\| = O_p(\sqrt{\log n})$ . Given that  $\mathbb{E} \|p^{-1/2} \boldsymbol{\varepsilon}_t\|^2 = O(1)$  by Assumption 3(iv), we can also show that  $\max_{t \in [n]} \|p^{-1/2} \boldsymbol{\varepsilon}_t\| = O_p(\sqrt{\log n})$  similarly.

(ii) The three arguments  $\max_{t \in [n]} \|\boldsymbol{\gamma}_t\| = O_p(\sqrt{\log n})$ ,  $\max_{t \in [n]} \|p^{-1/2} \int \mathbf{Q}(u)^\top \boldsymbol{\varepsilon}_t(u) du\| = O_p(\sqrt{\log n})$  and  $\max_{t' \in [n]} |p^{-1/2} \{\langle \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t'} \rangle - \mathbb{E} \langle \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t'} \rangle\}|^2 = O_p(\log n)$  for  $t \in [n]$  can be proved similarly to part (i), since  $\{\boldsymbol{\gamma}_t\}$  and  $\{\boldsymbol{\varepsilon}_t(\cdot)\}$  are sub-Gaussian (functional) linear processes by Assumption 4', with moment conditions  $\mathbb{E} \|\boldsymbol{\gamma}_t\|^2 = O(1)$ ,  $\mathbb{E} \|p^{-1/2} \int \mathbf{Q}(u)^\top \boldsymbol{\varepsilon}_t(u) du\|^2 = O(1)$  and  $\mathbb{E} |p^{-1/2} \{\langle \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t'} \rangle - \mathbb{E} \langle \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t'} \rangle\}|^4 = O(1)$  by Assumption 3'(ii).

### S.1.11 Proof of Lemma B.17

(i) In Assumption 1, we assume that

$$\iint \boldsymbol{\Sigma}_f(u, v) \boldsymbol{\Sigma}_f(u, v)^\top du dv = \text{diag}(\theta_1, \dots, \theta_r),$$

i.e.,

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

Then we have

$$\|\boldsymbol{\Sigma}_f\|_{\mathcal{S}, \max}^2 = \max_{i \in [r], j \in [r]} \|\Sigma_{f,ij}\|_{\mathcal{S}}^2 = \max_{i \in [r], j \in [r]} \iint \Sigma_{f,ij}(u, v)^2 du dv \leq \max_{i \in [r]} \iint \sum_{j=1}^r \Sigma_{f,ij}(u, v)^2 du dv = \max_{i \in [r]} \theta_i = \theta_1,$$

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

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

### S.1.12 Proof of Lemma B.18

(i) By Lemma A.9(i) and the fact  $\|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S}, \max} \leq \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{L}}$ , we have

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

(ii) By Lemma A.9(iii), we have

$$\begin{aligned} \|\boldsymbol{\Sigma}_y\|_{\mathcal{S}, \infty} &= \|\mathbf{B} \boldsymbol{\Sigma}_f \mathbf{B}^\top + \boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S}, \infty} \leq \|\mathbf{B} \boldsymbol{\Sigma}_f \mathbf{B}^\top\|_{\mathcal{S}, \infty} + \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S}, \infty} \\ &\leq \|\mathbf{B}\|_\infty \|\mathbf{B}^\top\|_\infty \|\boldsymbol{\Sigma}_f\|_{\mathcal{S}, \infty} + \|\boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S}, \infty} \leq r p C^2 C_\infty + s_p \lesssim p. \end{aligned}$$

Part (iii) can be proved similarly.

### S.1.13 Proof of Lemma B.19

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

$$\|\boldsymbol{\Omega}_{\mathcal{L}}\|_{\text{F}}^2 = \left\| p\mathbf{B} \int \int \boldsymbol{\Sigma}_f(u, v) \boldsymbol{\Sigma}_f(u, v)^{\text{T}} du dv \mathbf{B}^{\text{T}} \right\|_{\text{F}}^2 = p^4 \text{tr}(\text{diag}\{\theta_1, \dots, \theta_r\} \text{diag}\{\theta_1, \dots, \theta_r\}^{\text{T}}) = p^4 \sum_{j=1}^r \theta_j^2 \asymp p^4,$$

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

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

where the second inequality follows from Lemmas A.5(ii) and A.6(iv), and the last line follows from Assumption 3(iii) and Lemmas A.7(i)(ii) and B.17(ii).

### S.1.14 Proof of Lemma B.20

Notice that

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

where the second inequality follows from Lemma A.6(ii), the third inequality follows from Lemma A.9(iii), and the fourth inequality follows from Lemma B.17.

### S.1.15 Proof of Lemma B.23

By Proposition 1(ii) and Lemma B.19, if  $\boldsymbol{\xi}_j^{\text{T}} \tilde{\mathbf{b}}_j \geq 0$ , then

$$\|\boldsymbol{\xi}_j - \tilde{\mathbf{b}}_j\| = O_p(p^{-2}\|\boldsymbol{\Omega}_{\mathcal{R}}\|) = O_p(p^{-1}), \text{ for } j \in [r]. \quad (\text{S.4})$$

Applying Lemma A.2 yields that, if  $\hat{\boldsymbol{\xi}}_j^{\text{T}} \boldsymbol{\xi}_j \geq 0$ , we have

$$\|\hat{\boldsymbol{\xi}}_j - \boldsymbol{\xi}_j\| \leq \frac{\|\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}\|/\sqrt{2}}{\min(|\hat{\lambda}_{j-1} - \lambda_j|, |\lambda_j - \hat{\lambda}_{j+1}|)}, \quad (\text{S.5})$$

where  $\{\hat{\lambda}_j\}_{j=1}^p$  are the eigenvalues of  $\hat{\boldsymbol{\Omega}}$  in a descending order, and  $\{\hat{\boldsymbol{\xi}}_j\}_{j=1}^p$  are their corresponding eigenvectors. Then, for  $j \in [r]$ , we have  $|\hat{\lambda}_{j-1} - \lambda_j| \geq |\hat{\lambda}_{j-1} - \hat{\lambda}_j| - |\lambda_j - \hat{\lambda}_j|$ , where the first term  $|\hat{\lambda}_{j-1} - \hat{\lambda}_j| \gtrsim p^2$

with probability approaching one by Lemma B.22, and the second term  $|\lambda_j - \hat{\lambda}_j| = o_p(p^2)$  by Lemmas A.1 and B.21(i). Hence,  $|\hat{\lambda}_{j-1} - \lambda_j| \gtrsim p^2$  with probability approaching one 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 B.19 that  $\lambda_{r+1} = o(p^2)$ , and Lemmas A.1 and B.21(i) that  $|\hat{\lambda}_{r+1} - \lambda_{r+1}| = o_p(p^2)$ . Thus,

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

Applying (S.5), Lemma B.21(i) and the above argument, we have, if  $\hat{\boldsymbol{\xi}}_j^\top \boldsymbol{\xi}_j \geq 0$ , then

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

Combining with (S.4) we have, if  $\hat{\boldsymbol{\xi}}_j^\top \tilde{\mathbf{b}}_j \geq 0$ , then

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

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

$$\|\hat{\mathbf{B}} - \mathbf{B}\mathbf{U}^\top\|_F = O_p(\mathcal{M}_\varepsilon \sqrt{p/n} + 1/\sqrt{p}),$$

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

### S.1.16 Proof of Lemma B.27

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

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

which implies that

$$\begin{aligned} \|\hat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}^2 &= \int \int \hat{\Theta}_{ij}(u, v) du dv \leq \frac{2}{n} \int \int \sum_{t=1}^n \left\{ \hat{\varepsilon}_{ti}(u) \hat{\varepsilon}_{tj}(v) - \Sigma_{\varepsilon, ij}(u, v) \right\}^2 du dv + 2 \|\hat{\boldsymbol{\Sigma}}_\varepsilon - \boldsymbol{\Sigma}_\varepsilon\|_{\mathcal{S}, \max}^2 \\ &= \frac{2}{n} \int \int \sum_{t=1}^n \left\{ \hat{\varepsilon}_{ti}(u) \hat{\varepsilon}_{tj}(v) - \Sigma_{\varepsilon, ij}(u, v) \right\}^2 du dv + o_p(1), \end{aligned}$$

where the last line follows from Lemma B.26. Moreover

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

Here, we bound each term above as follows: (a) by Corollary 1, we have  $\max_{i \in [p], t \in [n]} \|\widehat{\varepsilon}_{ti} - \varepsilon_{ti}\|^2 = O_p(\varrho^2) = o_p(1)$  under Assumption 6; (b) by Lemma B.26(i), we have  $\max_{j \in [p]} n^{-1} \sum_{t=1}^n \|\widehat{\varepsilon}_{tj} - \varepsilon_{tj}\|^2 = O_p(\varpi_{n,p}^2) = o_p(1)$ ; (c) by Lemma B.12(ii) and Assumption 3(iv), we have  $\max_{j \in [p]} n^{-1} \sum_{t=1}^n \|\varepsilon_{tj}\|^2 \leq o_p(1) + \max_{j \in [p]} \int \Sigma_{\varepsilon,jj}(u, u) du = O_p(1)$ . Combining these results yields that

$$\|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}^2 \leq \frac{2}{n} \int \int \sum_{t=1}^n \left\{ \varepsilon_{ti}(u) \varepsilon_{tj}(v) - \Sigma_{\varepsilon,ij}(u, v) \right\}^2 du dv + o_p(1).$$

Similar arguments as those in the proof of Lemma 2 of Cai and Liu (2011) results in

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

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 one. Therefore,  $\max_{i, j \in [p]} \|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}$  is bounded away from infinity with probability approaching one.

We next prove the lower bound. Notice that

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

which implies that

$$\frac{1}{n} \int \int \sum_{t=1}^n \left\{ \varepsilon_{ti}(u) \varepsilon_{tj}(v) - \Sigma_{\varepsilon,ij}(u, v) \right\}^2 du dv \leq \frac{4}{n} \int \int \sum_{t=1}^n [\varepsilon_{ti}(u) \varepsilon_{tj}(v) - \widehat{\varepsilon}_{ti}(u) \widehat{\varepsilon}_{tj}(v)]^2 du dv + 4 \|\widehat{\Theta}_{ij}^{1/2}\|_{\mathcal{S}}^2 + o_p(1),$$

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

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



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

$$\frac{1}{n} \int \int \sum_{t=1}^n \{ \varepsilon_{ti}(u) \varepsilon_{tj}(v) - \hat{\varepsilon}_{ti}(u) \hat{\varepsilon}_{tj}(v) \}^2 du dv = o_p(1).$$

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

## S.2 Proofs of theoretical results in Section 3

### S.2.1 Technical lemmas and their proofs

**Lemma S.1.** *Under Assumptions 3'(iv) and 4', we have that,*

- (i) for any  $i, j \in [r]$ ,  $|n^{-1} \sum_{t=1}^n \gamma_{ti} \gamma_{tj} - \Sigma_{\gamma, ij}| = O_p(1/\sqrt{n})$ ,  $\|n^{-1} \sum_{t=1}^n \gamma_t \gamma_t^T - \Sigma_{\gamma}\|_{\max} = O_p(1/\sqrt{n})$ ;
- (ii) for any  $i, j \in [p]$ ,  $\|n^{-1} \sum_{t=1}^n \varepsilon_{ti} \varepsilon_{tj} - \Sigma_{\varepsilon, ij}\|_{\mathcal{S}} = O_p(\mathcal{M}_{\varepsilon}/\sqrt{n})$ ,  $\|n^{-1} \sum_{t=1}^n \varepsilon_t \varepsilon_t^T - \Sigma_{\varepsilon}\|_{\mathcal{S}, \max} = O_p(\mathcal{M}_{\varepsilon} \sqrt{\log p/n})$ ;
- (iii) for any  $i \in [p]$ ,  $j \in [r]$ ,  $\|n^{-1} \sum_{t=1}^n \varepsilon_{ti} \gamma_{tj}\| = O_p(\mathcal{M}_{\varepsilon}/\sqrt{n})$ ,  $\max_{i \in [p], j \in [r]} \|n^{-1} \sum_{t=1}^n \varepsilon_{ti} \gamma_{tj}\| = O_p(\mathcal{M}_{\varepsilon} \sqrt{\log p/n})$ .

*Proof.* For parts (i) and (ii), see Theorem 2 and equations (2.15) in Fang et al. (2022) for the corresponding proofs. For part (iii), see Remark 3 and equation (2.16) in Fang et al. (2022) for a proof.  $\square$

### S.2.2 Proof of Theorem 1'

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

**Lemma S.2.** *Under Assumption 4', it holds that*

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

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

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

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

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

□

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

$$\hat{\gamma}_t - \mathbf{H}\gamma_t = \left(\frac{\mathbf{V}}{p}\right)^{-1} \left\{ \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'} \frac{\mathbb{E}\langle \varepsilon_{t'}, \varepsilon_t \rangle}{p} + \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'} \zeta_{t't} + \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'} \eta_{t't} + \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'} \xi_{t't} \right\}, \quad (\text{S.6})$$

where

$$\zeta_{t't} = \frac{1}{p} \langle \varepsilon_{t'}, \varepsilon_t \rangle - \frac{1}{p} \mathbb{E}\langle \varepsilon_{t'}, \varepsilon_t \rangle, \quad \eta_{t't} = \frac{1}{p} \gamma_{t'}^T \sum_{i=1}^p \int \mathbf{q}_i(u) \varepsilon_{ti}(u) du, \quad \xi_{t't} = \frac{1}{p} \gamma_t^T \sum_{i=1}^p \int \mathbf{q}_i(u) \varepsilon_{t'i}(u) du.$$

**Lemma S.3.** *Under the assumptions of Theorem 1', it holds that*

- (i)  $\max_{t \in [n]} \|(np)^{-1} \sum_{t'=1}^n \hat{\gamma}_{t'} \mathbb{E}\langle \varepsilon_{t'}, \varepsilon_t \rangle\| = O_p(\mathcal{M}_\varepsilon/\sqrt{n})$ ;
- (ii)  $\max_{t \in [n]} \|n^{-1} \sum_{t'=1}^n \hat{\gamma}_{t'} \zeta_{t't}\| = O_p(\sqrt{\log n/p})$ ;
- (iii)  $\max_{t \in [n]} \|n^{-1} \sum_{t'=1}^n \hat{\gamma}_{t'} \eta_{t't}\| = O_p(\sqrt{\log n/p})$ ;
- (iv)  $\max_{t \in [n]} \|n^{-1} \sum_{t'=1}^n \hat{\gamma}_{t'} \xi_{t't}\| = O_p(\sqrt{\log n/p})$ .

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

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

where the last equality follows from Lemma S.2.

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

$$\begin{aligned} \max_{t \in [n]} \left\| \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'} \zeta_{t't} \right\| &\leq \max_{t \in [n]} \frac{1}{n} \left( \sum_{t'=1}^n \|\hat{\gamma}_{t'}\|^2 \sum_{t'=1}^n \zeta_{t't}^2 \right)^{1/2} = O_p(1) \left( \max_{t \in [n]} \frac{1}{n} \sum_{t'=1}^n \zeta_{t't}^2 \right)^{1/2} \\ &= O_p(1) \left\{ \max_{t \in [n]} \frac{1}{n} \sum_{t'=1}^n \left( \frac{1}{p} \langle \varepsilon_{t'}, \varepsilon_t \rangle - \frac{1}{p} \mathbb{E}\langle \varepsilon_{t'}, \varepsilon_t \rangle \right)^2 \right\}^{1/2} = O_p(\sqrt{\log n/p}), \end{aligned}$$

where the last equality follows from Lemma B.16(ii).

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

$$\max_{t \in [n]} \left\| \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'} \eta_{t't} \right\| \leq \left\| \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'} \gamma_{t'}^T \right\| \max_{t \in [n]} \left\| \frac{1}{p} \sum_{i=1}^p \int \mathbf{q}_i(u) \varepsilon_{ti}(u) du \right\| = O_p(\sqrt{\log n/p}),$$

where the last equality follows from Lemma B.16(ii).

(iv) By Assumption 3'(ii), we can show that  $\|(np)^{-1} \sum_{t'=1}^n \sum_{i=1}^p \int \mathbf{q}_i(u) \varepsilon_{t'i}(u) du \hat{\gamma}_{t'}\| = O_p(1/\sqrt{p})$ . Additionally,  $\max_{t \in [n]} \|\gamma_t\| = O_p(\sqrt{\log n})$  by Lemma B.16(ii). The desired result follows immediately that

$$\max_{t \in [n]} \left\| \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'} \xi_{t't} \right\| \leq \max_{t \in [n]} \|\gamma_t\| \left\| \frac{1}{np} \sum_{t'=1}^n \sum_{i=1}^p \int \mathbf{q}_i(u) \varepsilon_{t'i}(u) du \hat{\gamma}_{t'} \right\| = O_p(\sqrt{\log n/p}).$$

□

**Lemma S.4.** Denote  $\hat{\gamma}_t = (\hat{\gamma}_{t1}, \dots, \hat{\gamma}_{tr})^T$ . Under the assumptions of Theorem 1', it holds that, for  $i \in [r]$ ,

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

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

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

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

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

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

where the last equality follows from Lemma S.2.

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

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \left( \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'i} \zeta_{t't} \right)^2 &= \frac{1}{n^3} \sum_{t', l \in [n]} \left\{ \hat{\gamma}_{t'i} \hat{\gamma}_{li} \left( \sum_{t=1}^n \zeta_{t't} \zeta_{lt} \right) \right\} \\ &\leq \frac{1}{n^3} \left\{ \sum_{t', l \in [n]} \hat{\gamma}_{t'i}^2 \hat{\gamma}_{li}^2 \sum_{t \in [n]} \left( \sum_{t=1}^n \zeta_{t't} \zeta_{lt} \right)^2 \right\}^{1/2} \\ &\leq \frac{1}{n^3} \sum_{t'=1}^n \hat{\gamma}_{t'i}^2 \left\{ \sum_{t', l \in [n]} \left( \sum_{t=1}^n \zeta_{t't} \zeta_{lt} \right)^2 \right\}^{1/2} = \frac{1}{n^2} \left\{ \sum_{t', l \in [n]} \left( \sum_{t=1}^n \zeta_{t't} \zeta_{lt} \right)^2 \right\}^{1/2}. \end{aligned}$$

Notice that  $\mathbb{E}\left\{ \sum_{t', l \in [n]} \left( \sum_{t=1}^n \zeta_{t't} \zeta_{lt} \right)^2 \right\} = n^2 \mathbb{E}\left( \sum_{t=1}^n \zeta_{t't} \zeta_{lt} \right)^2 \leq n^4 \max_{t', t} \mathbb{E}|\zeta_{t't}|^4$ , 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–Schwarz inequality, and the facts that  $\sum_{t'=1}^n \hat{\gamma}_{t'i}^2 = n$  and  $n^{-1} \sum_{t'=1}^n \|\gamma_{t'}\|^2 = O_p(1)$ ,

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \left( \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'i} \eta_{t't} \right)^2 &\leq \left\| \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'i} \gamma_{t'}^T \right\|^2 \frac{1}{n} \sum_{t=1}^n \left\| \frac{1}{p} \sum_{j=1}^p \int \mathbf{q}_j(u) \varepsilon_{tj}(u) du \right\|^2 \\ &\leq \left( \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'i}^2 \frac{1}{n} \sum_{t'=1}^n \|\gamma_{t'}\|^2 \right) \frac{1}{np^2} \sum_{t=1}^n \left\| \sum_{j=1}^p \int \mathbf{q}_j(u) \varepsilon_{tj}(u) du \right\|^2 \\ &= O_p(1) \cdot \frac{1}{np^2} \sum_{t=1}^n \left\| \sum_{j=1}^p \int \mathbf{q}_j(u) \varepsilon_{tj}(u) du \right\|^2. \end{aligned}$$

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

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

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \left( \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'i} \xi_{t't} \right)^2 &\leq \left( \frac{1}{n} \sum_{t=1}^n \|\gamma_t\|^2 \right) \left\| \frac{1}{np} \sum_{t'=1}^n \hat{\gamma}_{t'i} \sum_{j=1}^p \int \mathbf{q}_j(u) \varepsilon_{t'j}(u) du \right\|^2 \\ &\leq O_p(1) \cdot \left( \frac{1}{n} \sum_{t'=1}^n \hat{\gamma}_{t'i}^2 \right) \frac{1}{np^2} \sum_{t'=1}^n \left\| \sum_{j=1}^p \int \mathbf{q}_j(u) \varepsilon_{t'j}(u) du \right\|^2 = O_p(1/p), \end{aligned}$$

where the result in the last equality has been obtained in part (iii).  $\square$

**Lemma S.5.** *Let  $\{\hat{\tau}_j\}_{j=1}^r$  be the first  $r$  largest eigenvalues of  $\hat{\Sigma}_y^S(\cdot, \cdot)$  in a descending order. Under the assumptions of Theorem 1', it holds that  $\hat{\tau}_r \gtrsim p$  with probability approaching one.*

*Proof.* By Proposition 2, we obtain that

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

To show  $\hat{\tau}_r \gtrsim p$  with probability approaching one, it suffices to show that  $|\hat{\tau}_r - \tau_r| = o_p(p)$ . By applying Lemma A.3 again, we only need to show  $\|\hat{\Sigma}_y^S - \Sigma_y\|_{\mathcal{S}, \mathcal{F}} = o_p(p)$ . Note that

$$\begin{aligned} \|\hat{\Sigma}_y^S - \Sigma_y\|_{\mathcal{S}, \mathcal{F}} &= \left\| \frac{1}{n} \sum_{t=1}^n (\mathbf{Q}\gamma_t + \varepsilon_t)(\mathbf{Q}\gamma_t + \varepsilon_t)^T - \mathbf{Q}\mathbf{Q}^T - \Sigma_\varepsilon \right\|_{\mathcal{S}, \mathcal{F}} \\ &\leq \left\| \mathbf{Q} \left( \frac{1}{n} \sum_{t=1}^n \gamma_t \gamma_t^T - \mathbf{I}_r \right) \mathbf{Q}^T \right\|_{\mathcal{S}, \mathcal{F}} + \left\| \frac{1}{n} \sum_{t=1}^n \varepsilon_t \varepsilon_t^T - \Sigma_\varepsilon \right\|_{\mathcal{S}, \mathcal{F}} \\ &\quad + \left\| \mathbf{Q} \left( n^{-1} \sum_{t=1}^n \gamma_t \varepsilon_t^T \right) \right\|_{\mathcal{S}, \mathcal{F}} + \left\| \left( \frac{1}{n} \sum_{t=1}^n \varepsilon_t \gamma_t^T \right) \mathbf{Q}^T \right\|_{\mathcal{S}, \mathcal{F}} \\ &\leq \left\| n^{-1} \sum_{t=1}^n \gamma_t \gamma_t^T - \mathbf{I}_r \right\|_{\mathcal{F}} \cdot \|\mathbf{Q}\mathbf{Q}^T\|_{\mathcal{S}, \mathcal{F}} + \left( \sum_{i=1}^p \sum_{j=1}^p \left\| n^{-1} \sum_{t=1}^n \varepsilon_{ti} \varepsilon_{tj} - \Sigma_{\varepsilon, ij} \right\|_{\mathcal{S}}^2 \right)^{1/2} \\ &\quad + 2 \left( \sum_{i=1}^p \sum_{j=1}^r \left\| n^{-1} \sum_{t=1}^n \varepsilon_{ti} \gamma_{tj} \right\|^2 \right)^{1/2} \cdot \sqrt{p} \max_{i \in [p]} \|\mathbf{q}_i\| \\ &= O_p(p/\sqrt{n}) + O_p(p\mathcal{M}_\varepsilon \sqrt{1/n}) + O_p(p\mathcal{M}_\varepsilon \sqrt{1/n}) = o_p(p), \end{aligned}$$

where the second inequality follows from Lemma A.9(ii), the fact  $\|\mathbf{K}\|_{\mathcal{S}, \mathcal{F}} \leq p\|\mathbf{K}\|_{\mathcal{S}, \max}$  any  $\mathbf{K}(\cdot, \cdot) \in$

$\mathbb{H}^p \otimes \mathbb{H}^p$ , and the Cauchy–Schwarz inequality. The last line of the above equation follows from Lemma S.1,  $\mathcal{M}_\varepsilon^2 = o(n)$ , and the fact

$$\|\mathbf{Q}\mathbf{Q}^\top\|_{\mathcal{S},\mathbf{F}} = \left[ \sum_{i=1}^p \sum_{j=1}^p \int \{\mathbf{q}_i(u)^\top \mathbf{q}_j(v)\}^2 du dv \right]^{1/2} \leq p \max_{i \in [p]} \|\mathbf{q}_i\|^2 = p.$$

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

**Lemma S.6.** *Under the assumptions of Theorem 1', it holds that*

- (i)  $\|\mathbf{H}\| = O_p(1)$ ;
- (ii)  $\mathbf{H}\mathbf{H}^\top = \mathbf{I}_r + O_p(\mathcal{M}_\varepsilon/\sqrt{n} + 1/\sqrt{p})$ ;
- (iii)  $\mathbf{H}^\top \mathbf{H} = \mathbf{I}_r + O_p(\mathcal{M}_\varepsilon/\sqrt{n} + 1/\sqrt{p})$ .

*Proof.* (i) By Lemma S.5,  $\|\mathbf{V}^{-1}\| = \hat{\tau}_r^{-1} = O_p(p^{-1})$ . Also,  $\|\hat{\mathbf{\Gamma}}\| = \lambda_{\max}^{1/2}(\hat{\mathbf{\Gamma}}^\top \hat{\mathbf{\Gamma}}) = \lambda_{\max}^{1/2}(n\mathbf{I}_r) = \sqrt{n}$  from the normalization (A.1), and  $\|\mathbf{\Gamma}\| = \lambda_{\max}^{1/2}(\mathbf{\Gamma}^\top \mathbf{\Gamma}) = \lambda_{\max}^{1/2}(\sum_{t=1}^n \gamma_t \gamma_t^\top) = O_p(\sqrt{n})$  by Lemma S.1(i). In addition,  $\|\int \mathbf{Q}(u)^\top \mathbf{Q}(u) du\| = O(p)$ . By the definition of  $\mathbf{H}$ , i.e.,  $\mathbf{H} = n^{-1} \mathbf{V}^{-1} \hat{\mathbf{\Gamma}}^\top \mathbf{\Gamma} \int \mathbf{Q}(u)^\top \mathbf{Q}(u) du$ , we have  $\|\mathbf{H}\| = O_p(1)$ , which is also satisfied for  $\|\mathbf{H}\|_{\mathbf{F}}$  since  $\mathbf{H} \in \mathbb{R}^{r \times r}$ .

(ii) Notice that

$$\|\mathbf{H}\mathbf{H}^\top - \mathbf{I}_r\|_{\mathbf{F}} \leq \left\| \mathbf{H}\mathbf{H}^\top - \frac{1}{n} \sum_{t=1}^n \mathbf{H} \gamma_t \gamma_t^\top \mathbf{H}^\top \right\|_{\mathbf{F}} + \left\| \frac{1}{n} \sum_{t=1}^n \mathbf{H} \gamma_t \gamma_t^\top \mathbf{H}^\top - \mathbf{I}_r \right\|_{\mathbf{F}}. \quad (\text{S.7})$$

In (S.7), the first term can be bound by  $\|\mathbf{H}\mathbf{H}^\top - n^{-1} \sum_{t=1}^n \mathbf{H} \gamma_t \gamma_t^\top \mathbf{H}^\top\|_{\mathbf{F}} \leq \|\mathbf{H}\|_{\mathbf{F}}^2 \|\mathbf{I}_r - n^{-1} \sum_{t=1}^n \gamma_t \gamma_t^\top\|_{\mathbf{F}} = O_p(1/\sqrt{n})$  by Lemma S.1(i). The second term can be bounded by

$$\begin{aligned} \left\| \frac{1}{n} \sum_{t=1}^n \mathbf{H} \gamma_t \gamma_t^\top \mathbf{H}^\top - \mathbf{I}_r \right\|_{\mathbf{F}} &= \left\| \frac{1}{n} \sum_{t=1}^n \mathbf{H} \gamma_t \gamma_t^\top \mathbf{H}^\top - \frac{1}{n} \sum_{t=1}^n \hat{\gamma}_t \hat{\gamma}_t^\top \right\|_{\mathbf{F}} \\ &\leq \left\| \frac{1}{n} \sum_{t=1}^n (\mathbf{H} \gamma_t - \hat{\gamma}_t) \gamma_t^\top \mathbf{H}^\top \right\|_{\mathbf{F}} + \left\| \frac{1}{n} \sum_{t=1}^n \hat{\gamma}_t (\hat{\gamma}_t^\top - \gamma_t^\top \mathbf{H}^\top) \right\|_{\mathbf{F}} \\ &\leq \left( \frac{1}{n} \sum_{t=1}^n \|\mathbf{H} \gamma_t - \hat{\gamma}_t\|^2 \frac{1}{n} \sum_{t=1}^n \|\mathbf{H} \gamma_t\|^2 \right)^{1/2} + \left( \frac{1}{n} \sum_{t=1}^n \|\mathbf{H} \gamma_t - \hat{\gamma}_t\|^2 \frac{1}{n} \sum_{t=1}^n \|\hat{\gamma}_t\|^2 \right)^{1/2} \\ &= O_p(\mathcal{M}_\varepsilon/\sqrt{n} + 1/\sqrt{p}), \end{aligned}$$

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

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

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

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

We are now ready to prove Theorem 1'.

(i) By Lemma S.5, the diagonal elements of  $\mathbf{V}/p = \text{diag}(\hat{\tau}_1/p, \dots, \hat{\tau}_r/p)$  are bounded away from 0. By the

inequality  $(a + b + c + d)^2 \leq 4(a^2 + b^2 + c^2 + d^2)$ , equation (S.6) and Lemma S.4, we have

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

The desired result immediately follows that

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

(ii) Note  $\|(\mathbf{V}/p)^{-1}\| = O(1)$ . Applying the inequality  $(a + b + c + d)^2 \leq 4(a^2 + b^2 + c^2 + d^2)$ , equation (S.6) and Lemma S.3, we have

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

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

$$\begin{aligned} \hat{\mathbf{q}}_i(\cdot) - \mathbf{H}\mathbf{q}_i(\cdot) &= \frac{1}{n} \sum_{t=1}^n y_{ti}(\cdot) \hat{\gamma}_t - \frac{1}{n} \sum_{t=1}^n \mathbf{H}\gamma_t \{y_{ti}(\cdot) - \mathbf{q}_i(\cdot)^\top \gamma_t + \varepsilon_{ti}(\cdot)\} - \mathbf{H}\mathbf{q}_i(\cdot) \\ &= \frac{1}{n} \sum_{t=1}^n \mathbf{H}\gamma_t \varepsilon_{ti}(u) + \frac{1}{n} \sum_{t=1}^n y_{ti}(\cdot) (\hat{\gamma}_t - \mathbf{H}\gamma_t) + \mathbf{H} \left( \frac{1}{n} \sum_{t=1}^n \gamma_t \gamma_t^\top - \mathbf{I}_r \right) \mathbf{q}_i(\cdot). \end{aligned} \tag{S.8}$$

The first term in (S.8) can be bounded by

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

where the inequality follows from Lemma A.10(i), and the last equality follows from Lemmas S.1(iii) and S.6(i). For the second term, since  $\Sigma_{y,ii} = \mathbf{q}_i^\top \Sigma_\gamma \mathbf{q}_i + \Sigma_{\varepsilon,ii}$  with  $\max_{i \in [p]} \|\mathbf{q}_i\| = O(1)$  and  $\|\Sigma_\varepsilon\|_{\mathcal{S}, \max} \leq \|\Sigma_\varepsilon\|_{\mathcal{L}} = O(1)$  by Assumption 3'(i)(iii), we have  $\max_{i \in [p]} \|\Sigma_{y,ii}\|_{\mathcal{S}} = \max_{i \in [p]} \mathbb{E}\|y_{ti}\|^2 = O(1)$ , and thus  $\max_{i \in [p]} n^{-1} \sum_{t=1}^n \|y_{ti}\|^2 = O_p(1)$  by Chebyshev's inequality. Using the Cauchy-Schwarz inequality in the second term of (S.8), we obtain that

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

In addition,  $\|\mathbf{H}\| = O_p(1)$  from Lemma S.6(i),  $\|n^{-1} \sum_{t=1}^n \gamma_t \gamma_t^\top - \mathbf{I}_r\| = O_p(1/\sqrt{n})$  from Lemma S.1(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

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

### S.2.3 Proof of Corollary 1'

By Theorem 1'(ii)(iii), Lemmas S.6(ii) and B.16(ii), we have

$$\begin{aligned} \max_{i \in [p], t \in [n]} \|\hat{\mathbf{q}}_i^\top \hat{\gamma}_t - \mathbf{q}_i^\top \gamma_t\| &\leq \max_{i \in [p]} \|\hat{\mathbf{q}}_i - \mathbf{H}\mathbf{q}_i\| \cdot \max_{t \in [n]} \|\hat{\gamma}_t - \mathbf{H}\gamma_t\| + \max_{i \in [p]} \|\mathbf{H}\mathbf{q}_i\| \cdot \max_{t \in [n]} \|\hat{\gamma}_t - \mathbf{H}\gamma_t\| \\ &\quad + \max_{i \in [p]} \|\hat{\mathbf{q}}_i - \mathbf{H}\mathbf{q}_i\| \cdot \max_{t \in [n]} \|\mathbf{H}\gamma_t\| + \max_{i \in [p]} \|\mathbf{q}_i\| \cdot \max_{t \in [n]} \|\gamma_t\| \cdot \|\mathbf{H}^\top \mathbf{H} - \mathbf{I}_r\| \\ &= O_p\left\{\varpi_{n,p} \cdot (\mathcal{M}_\varepsilon/\sqrt{n} + \sqrt{\log n/p})\right\} + O_p(\mathcal{M}_\varepsilon/\sqrt{n} + \sqrt{\log n/p}) \\ &\quad + O_p(\varpi_{n,p} \cdot \sqrt{\log n}) + O_p\left\{\sqrt{\log n} \cdot (\mathcal{M}_\varepsilon/\sqrt{n} + 1/\sqrt{p})\right\} \\ &= O_p\left\{\mathcal{M}_\varepsilon \sqrt{\log n \log p/n} + \sqrt{\log n/p}\right\}. \end{aligned}$$

### S.2.4 Proof of Theorem 4'

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

**Lemma S.7.** *Under the assumptions of Theorem 4', it holds that*

- (i)  $\max_{i \in [p]} n^{-1} \sum_{t=1}^n \|\hat{\varepsilon}_{ti} - \varepsilon_{ti}\|^2 = O_p(\varpi_{n,p}^2)$ ;
- (ii)  $\max_{i,j \in [p]} \|n^{-1} \sum_{t=1}^n \hat{\varepsilon}_{ti} \hat{\varepsilon}_{tj} - n^{-1} \sum_{t=1}^n \varepsilon_{ti} \varepsilon_{tj}\|_S = O_p(\varpi_{n,p})$ ;
- (iii)  $\|\tilde{\Sigma}_\varepsilon - \Sigma_\varepsilon\|_{S,\max} = O_p(\varpi_{n,p})$ .

*Proof.* (i) Note that  $\varepsilon_{ti}(\cdot) - \hat{\varepsilon}_{ti}(\cdot) = \{y_{ti}(\cdot) - \mathbf{q}_i(\cdot)^\top \gamma_t\} - \{y_{ti}(\cdot) - \hat{\mathbf{q}}_i(\cdot)^\top \hat{\gamma}_t\} = \hat{\mathbf{q}}_i(\cdot)^\top \hat{\gamma}_t - \mathbf{q}_i(\cdot)^\top \gamma_t$ , which can be decomposed as  $\hat{\mathbf{q}}_i(\cdot)^\top \hat{\gamma}_t - \mathbf{q}_i(\cdot)^\top \gamma_t = \{\hat{\mathbf{q}}_i(\cdot)^\top - \mathbf{q}_i(\cdot)^\top \mathbf{H}\} \hat{\gamma}_t + \mathbf{q}_i(\cdot)^\top \mathbf{H}^\top (\hat{\gamma}_t - \mathbf{H}\gamma_t) + \mathbf{q}_i(\cdot)^\top (\mathbf{H}^\top \mathbf{H} - \mathbf{I}_r) \gamma_t$ . Applying the inequality  $(a + b + c)^2 \leq 3a^2 + 3b^2 + 3c^2$  and the Cauchy–Schwarz inequality yields that

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

(ii) Notice that  $\max_{i \in [p]} \mathbb{E} \|\varepsilon_{ti}\|^2 = \max_{i \in [p]} \mathbb{E} \int \varepsilon_{ti}(u)^2 du = \max_{i \in [p]} \int \Sigma_{\varepsilon,ii}(u, u) du = O(1)$  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–Schwarz inequality, we have

$$\begin{aligned} \max_{i,j \in [p]} \left\| \frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_{ti} \hat{\varepsilon}_{tj} - \frac{1}{n} \sum_{t=1}^n \varepsilon_{ti} \varepsilon_{tj} \right\|_S &= \max_{i,j \in [p]} \left\| \frac{1}{n} \sum_{t=1}^n (\hat{\varepsilon}_{ti} - \varepsilon_{ti}) \hat{\varepsilon}_{tj} + \varepsilon_{ti} (\hat{\varepsilon}_{tj} - \varepsilon_{tj}) \right\|_S \\ &\leq \max_{i \in [p]} \frac{1}{n} \sum_{t=1}^n \|\hat{\varepsilon}_{ti} - \varepsilon_{ti}\|^2 + 2 \left( \max_{i \in [p]} \frac{1}{n} \sum_{t=1}^n \|\varepsilon_{ti}\|^2 \right)^{1/2} \left( \max_{j \in [p]} \frac{1}{n} \sum_{t=1}^n \|\hat{\varepsilon}_{tj} - \varepsilon_{tj}\|^2 \right)^{1/2} \\ &= O_p(\varpi_{n,p}^2) + O_p(\varpi_{n,p}) = O_p(\varpi_{n,p}). \end{aligned}$$



(iii) By part (ii) and Lemma S.1(ii), the result follows immediately.  $\square$

We are now ready to prove Theorem 4'. By Corollary 1' and Lemma S.7(i), we can follow nearly the same procedure as in the proof of Theorem 4 to show the similar argument that there exist some constants  $C_1, C_2 > 0$  such that with probability approaching one,

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

Together with Lemmas S.7(iii), we can show that for any  $\epsilon > 0$ , there exist some positive constant  $N$  such that each of events

$$\tilde{\Upsilon}_1 = \left\{ \max_{i \in [p], j \in [p]} \|\tilde{\Sigma}_{\epsilon, ij} - \Sigma_{\epsilon, ij}\|_{\mathcal{S}} < N\varpi_{n,p} \right\}, \quad \tilde{\Upsilon}_2 = \left\{ C_1 \leq \|\tilde{\Theta}_{ij}^{1/2}\|_{\mathcal{S}} \leq C_2, \text{ all } i, j \in [p] \right\}$$

hold with probability at least  $1 - \epsilon$ . Then for  $\dot{C} > 2NC_1^{-1}(\varpi_{n,p}/\omega_{n,p})$  and under the event  $\tilde{\Upsilon}_1 \cap \tilde{\Upsilon}_2$ , we obtain that  $\|\tilde{\Sigma}_{\epsilon}^{\mathcal{A}} - \Sigma_{\epsilon}\|_{\mathcal{S},1} \lesssim \varpi_{n,p}^{1-q} s_p$  by using the same way as the proof of Theorem 4. By Proposition 3, we know that  $\hat{\mathbf{R}}^{\mathcal{A}} = \tilde{\Sigma}_{\epsilon}^{\mathcal{A}}$ . Therefore, with probability at least  $1 - 2\epsilon$ ,  $\|\hat{\mathbf{R}}^{\mathcal{A}} - \Sigma_{\epsilon}\|_{\mathcal{S},1} \lesssim \varpi_{n,p}^{1-q} s_p$ . Considering that  $\epsilon > 0$  can be arbitrarily small, we have  $\|\hat{\mathbf{R}}^{\mathcal{A}} - \Sigma_{\epsilon}\|_{\mathcal{L}} \leq \|\hat{\mathbf{R}}^{\mathcal{A}} - \Sigma_{\epsilon}\|_{\mathcal{S},1} = O_p(\varpi_{n,p}^{1-q} s_p)$ .

### S.2.5 Proof of Theorem 5'

Under Assumption 1', we have  $\Sigma_y(u, v) = \mathbf{Q}(u)\mathbf{Q}(v)^{\top} + \Sigma_{\epsilon}(u, v)$ . By the Cauchy–Schwarz inequality and Lemma A.10, we have

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

where the last line follows from Theorem 1'(iii) and Lemma S.6. Then by Lemma S.7, we have  $\|\tilde{\Sigma}_{\epsilon} - \Sigma_{\epsilon}\|_{\mathcal{S},\max} = \max_{i,j \in [p]} \|\tilde{\Sigma}_{\epsilon, ij} - \Sigma_{\epsilon, ij}\|_{\mathcal{S}} = O_p(\varpi_{n,p})$ , and hence

$$\|\tilde{\Sigma}_{\epsilon}^{\mathcal{A}} - \Sigma_{\epsilon}\|_{\mathcal{S},\max} \leq \|\tilde{\Sigma}_{\epsilon}^{\mathcal{A}} - \tilde{\Sigma}_{\epsilon}\|_{\mathcal{S},\max} + \|\tilde{\Sigma}_{\epsilon} - \Sigma_{\epsilon}\|_{\mathcal{S},\max} \leq \max_{i,j \in [p]} (\|\tilde{\Theta}_{ij}^{1/2}\|_{\mathcal{S}} \lambda) + O_p(\varpi_{n,p}) = O_p(\varpi_{n,p}),$$

where  $\tilde{\Theta}_{ij}(u, v) \equiv n^{-1} \sum_{t=1}^n \{\hat{\varepsilon}_{ti}(u)\hat{\varepsilon}_{tj}(v) - \tilde{\Sigma}_{\epsilon, ij}(u, v)\}^2$ , the last line follows from Lemma S.7(iii), the choice of  $\lambda = \dot{C}(\sqrt{\log p/n} + \sqrt{1/p}) \lesssim \varpi_{n,p}$ , and the fact  $\max_{i,j \in [p]} \|\tilde{\Theta}_{ij}^{1/2}\|_{\mathcal{S}} = O_p(1)$  that can be proved following a similar argument compared to the proof of Lemma B.27. The desired result follows immediately.

### S.2.6 Proof of Theorem 6'

Given  $\text{Ker}(\Sigma_{\epsilon}) = \text{Ker}(\mathbf{Q}\Sigma_{\gamma}\mathbf{Q}^{\top})$ , with the covariance decomposition for FPOET estimator  $\Sigma_y(u, v) = \mathbf{Q}(u)\Sigma_{\gamma}\mathbf{Q}(v)^{\top} + \Sigma_{\epsilon}(u, v)$ , we apply Sherman–Morrison–Woodbury formula (Theorem 3.5.6 of Hsing and Eubank (2015)) to obtain its inverse  $\Sigma_y^{\dagger} = \Sigma_{\epsilon}^{\dagger} - \Sigma_{\epsilon}^{\dagger}\mathbf{Q}(\mathbf{I}_r + \mathbf{Q}^{\top}\Sigma_{\epsilon}^{\dagger}\mathbf{Q})^{\dagger}\mathbf{Q}^{\top}\Sigma_{\epsilon}^{\dagger}$  (note that under Assumption 1',

$\Sigma_\gamma = \mathbf{I}_r$  and  $\hat{\Sigma}_\gamma = \mathbf{I}_r$ ). Then, the plug-in inverse FPOET estimator is defined as  $(\hat{\Sigma}_y^\mathcal{F})^\dagger = (\hat{\mathbf{R}}^\mathcal{A})^\dagger - (\hat{\mathbf{R}}^\mathcal{A})^\dagger \hat{\mathbf{Q}} \{\mathbf{I}_r + \hat{\mathbf{Q}}^\mathbf{T} (\hat{\mathbf{R}}^\mathcal{A})^\dagger \hat{\mathbf{Q}}\}^\dagger \hat{\mathbf{Q}}^\mathbf{T} (\hat{\mathbf{R}}^\mathcal{A})^\dagger$ . In this section,  $(\hat{\mathbf{R}}^\mathcal{A})^\dagger$  and  $\{\mathbf{I}_r + \hat{\mathbf{Q}}^\mathbf{T} (\hat{\mathbf{R}}^\mathcal{A})^\dagger \hat{\mathbf{Q}}\}^\dagger$  are defined in a similar way to the inverse in Lemma B.15 associated with  $\Sigma_\varepsilon$  and  $\mathbf{I}_r + \mathbf{Q}^\mathbf{T} \Sigma_\varepsilon^\dagger \mathbf{Q}$ , respectively. To prove Theorem 6', we first present some technical lemmas with their proofs.

**Lemma S.8.** *Under the assumptions of Theorem 6', then,  $\hat{\mathbf{R}}^\mathcal{A}$  has a bounded Moore–Penrose inverse with probability approaching one, and  $\|(\hat{\mathbf{R}}^\mathcal{A})^\dagger - \Sigma_\varepsilon^\dagger\|_{\mathcal{L}} = O_p(\varpi_{n,p}^{1-q} s_p)$ .*

*Proof.* Provided that  $\varpi_{n,p}^{1-q} s_p = o(1)$  and  $\|\Sigma_\varepsilon^\dagger\|_{\mathcal{L}} > c_5$  for some constant  $c_5 > 0$ , we combine Lemma B.15 and Theorem 4' to yield that  $\|(\hat{\mathbf{R}}^\mathcal{A})^\dagger\|_{\mathcal{L}} < 2c_5$  with probability approaching one, and thus  $\hat{\mathbf{R}}^\mathcal{A}$  has a bounded Moore–Penrose inverse with probability approaching one together with the desired result  $\|(\hat{\mathbf{R}}^\mathcal{A})^\dagger - \Sigma_\varepsilon^\dagger\|_{\mathcal{L}} = O_p(\varpi_{n,p}^{1-q} s_p)$ .  $\square$

**Lemma S.9.** *Under the assumptions of Theorem 6',*

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

*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,  $\|\mathbf{Q}\|_{\mathcal{S}, \mathbf{F}}^2 = \sum_{i=1}^p \|\mathbf{q}_i\|^2 = \int \mathbf{Q}(u)^\mathbf{T} \mathbf{Q}(u) du = p(\vartheta_1 + \dots + \vartheta_r) \asymp p$  under Assumption 1'. By Theorem 1'(iii),  $\|\hat{\mathbf{Q}} - \mathbf{Q} \mathbf{H}^\mathbf{T}\|_{\mathcal{S}, \mathbf{F}} = \{\sum_{i=1}^p \|\hat{\mathbf{q}}_i - \mathbf{H} \mathbf{q}_i\|^2\}^{1/2} = O_p(\sqrt{p} \varpi_{n,p})$ . Hence,

$$\begin{aligned} \|\hat{\mathbf{Q}}^\mathbf{T} (\hat{\mathbf{R}}^\mathcal{A})^\dagger \hat{\mathbf{Q}} - \mathbf{H} \mathbf{Q}^\mathbf{T} \Sigma_\varepsilon^\dagger \mathbf{Q} \mathbf{H}^\mathbf{T}\|_{\mathcal{L}} &\leq 2\|(\hat{\mathbf{Q}} - \mathbf{Q} \mathbf{H}^\mathbf{T})^\mathbf{T} (\hat{\mathbf{R}}^\mathcal{A})^\dagger \hat{\mathbf{Q}}\|_{\mathcal{L}} + \|\mathbf{H} \mathbf{Q}^\mathbf{T} \{(\hat{\mathbf{R}}^\mathcal{A})^\dagger - \Sigma_\varepsilon^\dagger\} \mathbf{Q} \mathbf{H}^\mathbf{T}\|_{\mathcal{L}} \\ &\leq 2\|\hat{\mathbf{Q}} - \mathbf{Q} \mathbf{H}^\mathbf{T}\|_{\mathcal{S}, \mathbf{F}} \|(\hat{\mathbf{R}}^\mathcal{A})^\dagger\|_{\mathcal{L}} \|\mathbf{Q}\|_{\mathcal{S}, \mathbf{F}} + \|\mathbf{Q}\|_{\mathcal{S}, \mathbf{F}}^2 \|\mathbf{H}\|^2 \|(\hat{\mathbf{R}}^\mathcal{A})^\dagger - \Sigma_\varepsilon^\dagger\|_{\mathcal{L}} \\ &= O_p(p \varpi_{n,p}) + O_p(p \varpi_{n,p}^{1-q} s_p) = O_p(p \varpi_{n,p}^{1-q} s_p) = o_p(p), \end{aligned}$$

where the second inequality follows from Lemma A.7(i)(ii), and the last line follows from Lemmas S.6 and S.8.  $\square$

**Lemma S.10.** *Under the assumptions of Theorem 6',*

- (i)  $\|(\mathbf{I}_r + \mathbf{H} \mathbf{Q}^\mathbf{T} \Sigma_\varepsilon^\dagger \mathbf{Q} \mathbf{H}^\mathbf{T})^\dagger\|_{\mathcal{L}} = O(p^{-1})$ ;
- (ii)  $\|\{\mathbf{I}_r + \hat{\mathbf{Q}}^\mathbf{T} (\hat{\mathbf{R}}^\mathcal{A})^\dagger \hat{\mathbf{Q}}\}^\dagger\|_{\mathcal{L}} = O_p(p^{-1})$ ;
- (iii)  $\|(\mathbf{I}_r + \mathbf{Q}^\mathbf{T} \Sigma_\varepsilon^\dagger \mathbf{Q})^\dagger\|_{\mathcal{L}} = O(p^{-1})$ ;
- (iv)  $\|(\mathbf{H} \mathbf{H}^\mathbf{T})^\dagger + \mathbf{Q}^\mathbf{T} \Sigma_\varepsilon^\dagger \mathbf{Q}\|_{\mathcal{L}} = O(p^{-1})$ .

*Proof.* (i) By Lemma S.6, with probability approaching one,  $\lambda_{\min}(\mathbf{H} \mathbf{H}^\mathbf{T})$  is bounded away from 0. Hence,

$$\|(\mathbf{I}_r + \mathbf{H} \mathbf{Q}^\mathbf{T} \Sigma_\varepsilon^\dagger \mathbf{Q} \mathbf{H}^\mathbf{T})^\dagger\|_{\mathcal{L}} \leq \|(\mathbf{H} \mathbf{Q}^\mathbf{T} \Sigma_\varepsilon^\dagger \mathbf{Q} \mathbf{H}^\mathbf{T})^\dagger\|_{\mathcal{L}} \leq \{\lambda_{\min}(\mathbf{H}^\mathbf{T} \mathbf{H})\}^{-1} \|\Sigma_\varepsilon\|_{\mathcal{L}} \|(\mathbf{Q} \mathbf{Q}^\mathbf{T})^\dagger\|_{\mathcal{L}} = O(p^{-1}),$$

where  $\|(\mathbf{Q} \mathbf{Q}^\mathbf{T})^\dagger\|_{\mathcal{L}} = (p \vartheta_r)^{-1}$  by Assumption 1' and  $\|\Sigma_\varepsilon\|_{\mathcal{L}} = O(1)$  by Assumption 3'(iii).

(ii) The result follows from part (i) and Lemmas B.15 and S.9.

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

(iv) The result follows from part (iii) and Lemmas B.15 and S.6.  $\square$

We are now ready to prove Theorem 6'. Using the functional version of Sherman–Morrison–Woodbury formula, we have  $\|(\hat{\Sigma}_y^{\mathcal{F}})^\dagger - \Sigma_y^\dagger\|_{\mathcal{L}} \leq \sum_{k=1}^6 L_k$ , where

$$\begin{aligned} L_1 &= \|(\hat{\mathbf{R}}^{\mathcal{A}})^\dagger - \Sigma_\varepsilon^\dagger\|_{\mathcal{L}}, \\ L_2 &= \|\{(\hat{\mathbf{R}}^{\mathcal{A}})^\dagger - \Sigma_\varepsilon^\dagger\} \hat{\mathbf{Q}} \{\mathbf{I}_r + \hat{\mathbf{Q}}^\top (\hat{\mathbf{R}}^{\mathcal{A}})^\dagger \hat{\mathbf{Q}}\}^\dagger \hat{\mathbf{Q}}^\top (\hat{\mathbf{R}}^{\mathcal{A}})^\dagger\|_{\mathcal{L}}, \\ L_3 &= \|\{(\hat{\mathbf{R}}^{\mathcal{A}})^\dagger - \Sigma_\varepsilon^\dagger\} \hat{\mathbf{Q}} \{\mathbf{I}_r + \hat{\mathbf{Q}}^\top (\hat{\mathbf{R}}^{\mathcal{A}})^\dagger \hat{\mathbf{Q}}\}^\dagger \hat{\mathbf{Q}}^\top \Sigma_\varepsilon^\dagger\|_{\mathcal{L}}, \\ L_4 &= \|\Sigma_\varepsilon^\dagger (\hat{\mathbf{Q}} - \mathbf{Q} \mathbf{H}^\top) \{\mathbf{I}_r + \hat{\mathbf{Q}}^\top (\hat{\mathbf{R}}^{\mathcal{A}})^\dagger \hat{\mathbf{Q}}\}^\dagger \hat{\mathbf{Q}}^\top \Sigma_\varepsilon^\dagger\|_{\mathcal{L}}, \\ L_5 &= \|\Sigma_\varepsilon^\dagger (\hat{\mathbf{Q}} - \mathbf{Q} \mathbf{H}^\top) \{\mathbf{I}_r + \hat{\mathbf{Q}}^\top (\hat{\mathbf{R}}^{\mathcal{A}})^\dagger \hat{\mathbf{Q}}\}^\dagger \mathbf{H} \mathbf{Q}^\top \Sigma_\varepsilon^\dagger\|_{\mathcal{L}}, \\ L_6 &= \|\Sigma_\varepsilon^\dagger \mathbf{Q} \mathbf{H}^\top [\{\mathbf{I}_r + \hat{\mathbf{Q}}^\top (\hat{\mathbf{R}}^{\mathcal{A}})^\dagger \hat{\mathbf{Q}}\}^\dagger - (\mathbf{I}_r + \mathbf{H} \mathbf{Q}^\top \Sigma_\varepsilon^\dagger \mathbf{Q} \mathbf{H}^\top)^\dagger] \mathbf{H} \mathbf{Q}^\top \Sigma_\varepsilon^\dagger\|_{\mathcal{L}}. \end{aligned}$$

Combining with Lemmas S.8 and S.10, the desired result follows from a similar argument to the proof of Theorem 6.

### S.3 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 B.23–B.28 is an identity matrix, which means that, when we perform eigen-decomposition on  $\hat{\mathbf{\Omega}}$ , we can always select the correct direction of  $\hat{\xi}_j$  to ensure  $\hat{\xi}_j^\top \tilde{\mathbf{b}}_j \geq 0$ . The proofs in Section 3 verify that the choice of  $\mathbf{U}$  does not affect the theoretical results.

#### S.3.1 Proposition 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  $\Sigma = \{\Sigma_{ij}(\cdot, \cdot)\}_{p \times p}$ , and  $\hat{\Sigma} = \{\hat{\Sigma}_{ij}(\cdot, \cdot)\}_{p \times p}$  with each  $\Sigma_{ij}, \hat{\Sigma}_{ij} \in \mathbb{S}$ . For any fixed  $\mathbf{w}(\cdot) \in \mathbb{H}^p$ , we have*

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

*Proof.* Consider that

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

Thus,  $\left| \langle \mathbf{w}, \hat{\Sigma}(\mathbf{w}) \rangle - \langle \mathbf{w}, \Sigma(\mathbf{w}) \rangle \right| = \left| \langle \mathbf{w}, (\hat{\Sigma} - \Sigma)\mathbf{w} \rangle \right| \leq \|\hat{\Sigma} - \Sigma\|_{\mathcal{S}, \max} \left( \sum_{i=1}^p \|w_i\| \right)^2.$   $\square$

**Proposition S.2.** Suppose  $\Sigma$  and  $\hat{\Sigma}$  are Mercer's kernels and  $\Sigma$  has a bounded inverse. For any fixed  $\mathbf{w}(\cdot) \in \text{Ker}(\Sigma)^\perp$ , we have

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

*Proof.* For any given  $\mathbf{w} \in \text{Ker}(\Sigma)^\perp$ , we denote  $\mathbf{x} = \Sigma^{1/2} \mathbf{w} \in \text{Im}(\Sigma)$  and  $\mathbf{w} = (\Sigma^\dagger)^{1/2} \mathbf{x}$ , provided that  $\Sigma$  has a bounded inverse. Consider that

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

The relative error can be bounded by

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

where the last line follows from Lemma B.14. □

### S.3.2 Proof of Theorem 7

Since there exist constants  $c_3$  and  $c_4 > 0$  such that  $\|\Sigma_\epsilon^\dagger\|_{\mathcal{L}} < c_3$ ,  $\|\Sigma_f^\dagger\|_{\mathcal{L}} < c_4$ , we can obtain that  $\|\Sigma_y^\dagger\|_{\mathcal{L}} = O(1)$ , and thus for any  $\mathbf{K} \in \mathbb{H}^p \otimes \mathbb{H}^p$ ,  $\|\mathbf{K}\|_{\mathcal{S}, \Sigma_y}^2 = O(p^{-1}) \|\mathbf{K}\|_{\mathcal{S}, \mathbf{F}}^2$  by Lemma A.8(iii).

To prove Theorem 7, we first present some technical lemmas with their proofs.

**Lemma S.11.** Under the assumptions of Theorem 7, we have  $\|\mathbf{B}^\top \Sigma_y^\dagger \mathbf{B}\|_{\mathcal{L}} = O(1)$ .

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

$$\Sigma_y^\dagger = \Sigma_\epsilon^\dagger - \Sigma_\epsilon^\dagger \mathbf{B} (\Sigma_f^\dagger + \mathbf{B}^\top \Sigma_\epsilon^\dagger \mathbf{B})^\dagger \mathbf{B}^\top \Sigma_\epsilon^\dagger.$$

Then it follows that

$$\begin{aligned} \mathbf{B}^\top \Sigma_y^\dagger \mathbf{B} &= \mathbf{B}^\top \Sigma_\epsilon^\dagger \mathbf{B} - \mathbf{B}^\top \Sigma_\epsilon^\dagger \mathbf{B} (\Sigma_f^\dagger + \mathbf{B}^\top \Sigma_\epsilon^\dagger \mathbf{B})^\dagger \mathbf{B}^\top \Sigma_\epsilon^\dagger \mathbf{B} \\ &= \mathbf{B}^\top \Sigma_\epsilon^\dagger \mathbf{B} (\Sigma_f^\dagger + \mathbf{B}^\top \Sigma_\epsilon^\dagger \mathbf{B})^\dagger \Sigma_f^\dagger \\ &= \Sigma_f^\dagger - \Sigma_f^\dagger (\Sigma_f^\dagger + \mathbf{B}^\top \Sigma_\epsilon^\dagger \mathbf{B})^\dagger \Sigma_f^\dagger, \end{aligned} \tag{S.9}$$

where the last two equalities follow from the assumption  $\text{Ker}(\Sigma_\epsilon) = \text{Ker}(\mathbf{B} \Sigma_f \mathbf{B}^\top)$ . (S.9) also implies that  $\Sigma_f^\dagger \geq \Sigma_f^\dagger (\Sigma_f^\dagger + \mathbf{B}^\top \Sigma_\epsilon^\dagger \mathbf{B})^\dagger \Sigma_f^\dagger$  since  $\mathbf{B}^\top \Sigma_y^\dagger \mathbf{B} \geq 0$ . Here, for two Mercer's kernels  $\mathbf{K}, \mathbf{G} \in \mathbb{H}^r \otimes \mathbb{H}^r$ , we denote  $\mathbf{K} \geq \mathbf{G}$  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.,  $\mathbf{K} \geq \mathbf{G}$  implies  $\|\mathbf{K}\|_{\mathcal{L}} \geq \|\mathbf{G}\|_{\mathcal{L}}$ . Thus, from (S.9) we have

$$\|\mathbf{B}^\top \Sigma_y^\dagger \mathbf{B}\|_{\mathcal{L}} \leq \|\Sigma_f^\dagger\|_{\mathcal{L}} + \|\Sigma_f^\dagger (\Sigma_f^\dagger + \mathbf{B}^\top \Sigma_\epsilon^\dagger \mathbf{B})^\dagger \Sigma_f^\dagger\|_{\mathcal{L}} \leq 2\|\Sigma_f^\dagger\|_{\mathcal{L}} \leq 2c_4 = O(1),$$

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

**Lemma S.12.** *Under the assumptions of Theorem 7, it follows that*

- (i)  $\|\hat{\Sigma}_\varepsilon^{\mathcal{A}} - \Sigma_\varepsilon\|_{\mathcal{S}, \Sigma_y}^2 = O_p(\varpi_{n,p}^{2-2q} s_p^2)$ ;
- (ii)  $\|(\hat{\mathbf{B}} - \mathbf{B})\hat{\Sigma}_f(\hat{\mathbf{B}} - \mathbf{B})^\top\|_{\mathcal{S}, \Sigma_y}^2 = O_p(\mathcal{M}_\varepsilon^4 p/n^2 + 1/p^3)$ ;
- (iii)  $\|\mathbf{B}\hat{\Sigma}_f(\hat{\mathbf{B}} - \mathbf{B})^\top\|_{\mathcal{S}, \Sigma_y}^2 = O_p(\mathcal{M}_\varepsilon^2/n + 1/p^2)$ ;
- (iv)  $\|\mathbf{B}(\hat{\Sigma}_f - \Sigma_f)\mathbf{B}^\top\|_{\mathcal{S}, \Sigma_y}^2 = O_p\{\mathcal{M}_\varepsilon/(\sqrt{np}) + 1/p^{3/2}\}$ .

*Proof.* (i) Since all the eigenvalues of  $\Sigma_y$  are bounded away from zero, and by Theorem 4,

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

(ii) By applying Lemmas A.8(iii) and B.23, we have

$$\|(\hat{\mathbf{B}} - \mathbf{B})\hat{\Sigma}_f(\hat{\mathbf{B}} - \mathbf{B})^\top\|_{\mathcal{S}, \Sigma_y}^2 \leq p^{-1} \|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathbf{F}}^4 \|\hat{\Sigma}_f\|_{\mathcal{L}}^2 \|\Sigma_y^\dagger\|_{\mathcal{L}}^2 = O_p(\mathcal{M}_\varepsilon^4 p/n^2 + 1/p^3).$$

(iii) Consider

$$\begin{aligned} \|\mathbf{B}\hat{\Sigma}_f(\hat{\mathbf{B}} - \mathbf{B})^\top\|_{\mathcal{S}, \Sigma_y}^2 &= p^{-1} \text{tr} \left\{ \int \left[ \hat{\Sigma}_f(\hat{\mathbf{B}} - \mathbf{B})^\top \Sigma_y^\dagger (\hat{\mathbf{B}} - \mathbf{B}) \hat{\Sigma}_f \mathbf{B}^\top \Sigma_y^\dagger \mathbf{B} \right] (u, u) du \right\} \\ &\leq p^{-1} \|\mathbf{B}^\top \Sigma_y^\dagger \mathbf{B}\|_{\mathcal{L}} \|\Sigma_y^\dagger\|_{\mathcal{L}} \|\hat{\mathbf{B}} - \mathbf{B}\|_{\mathbf{F}}^2 \|\hat{\Sigma}_f\|_{\mathcal{N}}^2 \\ &= O_p(\mathcal{M}_\varepsilon^2/n + 1/p^2), \end{aligned}$$

where the first inequality follows from Lemma A.8(ii) and the last line follows from Lemmas S.11 and B.23.

(iv) A similar argument shows that

$$\begin{aligned} \|\mathbf{B}(\hat{\Sigma}_f - \Sigma_f)\mathbf{B}^\top\|_{\mathcal{S}, \Sigma_y}^2 &= p^{-1} \text{tr} \left\{ \int \left[ (\hat{\Sigma}_f - \Sigma_f) \mathbf{B}^\top \Sigma_y^\dagger \mathbf{B} (\hat{\Sigma}_f - \Sigma_f) \mathbf{B}^\top \Sigma_y^\dagger \mathbf{B} \right] (u, u) du \right\} \\ &\leq p^{-1} \|\mathbf{B}^\top \Sigma_y^\dagger \mathbf{B}\|_{\mathcal{L}}^2 \|\hat{\Sigma}_f - \Sigma_f\|_{\mathcal{L}} \|\hat{\Sigma}_f - \Sigma_f\|_{\mathcal{N}} \\ &= O_p\{\mathcal{M}_\varepsilon/(\sqrt{np}) + 1/p^{3/2}\}, \end{aligned}$$

where the first inequality follows from Lemma A.8(ii) and the last line follows from Lemmas S.11 and B.23.  $\square$

We are now ready to prove Theorem 7. By Lemma S.12,

$$\begin{aligned} \|\hat{\Sigma}_y^{\mathcal{D}} - \Sigma_y\|_{\mathcal{S}, \Sigma_y}^2 &\leq 2\|\hat{\Sigma}_\varepsilon^{\mathcal{A}} - \Sigma_\varepsilon\|_{\mathcal{S}, \Sigma_y}^2 + 2\|(\hat{\mathbf{B}} - \mathbf{B})\hat{\Sigma}_f(\hat{\mathbf{B}} - \mathbf{B})^\top\|_{\mathcal{S}, \Sigma_y}^2 \\ &\quad + 4\|\mathbf{B}\hat{\Sigma}_f(\hat{\mathbf{B}} - \mathbf{B})^\top\|_{\mathcal{S}, \Sigma_y}^2 + 2\|\mathbf{B}(\hat{\Sigma}_f - \Sigma_f)\mathbf{B}^\top\|_{\mathcal{S}, \Sigma_y}^2 \\ &= O_p\left(\frac{\mathcal{M}_\varepsilon^4 p}{n^2} + \varpi_{n,p}^{2-2q} s_p^2\right), \end{aligned}$$

which then implies that

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

## S.4 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),

$$\Sigma_y(u, v) = \mathbf{B} \left\{ \sum_{i=1}^{50} i^{-2} \phi_i(u) \phi_i(v) (\mathbf{I}_r - \mathbf{A}^2)^{-1} \right\} \mathbf{B}^\top + \sum_{l=1}^{25} 3^{-1} 2^{2-l} \phi_l(u) \phi_l(v) \mathbf{C}_\zeta.$$

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

$$\Sigma_y(u, v) = \mathbf{Q}(u) (\mathbf{I}_r - \mathbf{A}^2)^{-1} \mathbf{Q}(v)^\top + \sum_{l=1}^{25} 3^{-1} 2^{2-l} \phi_l(u) \phi_l(v) \mathbf{C}_\zeta.$$

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

*Proof.* (i) Let  $\Xi_t = (\xi_{t1}, \dots, \xi_{t50}) \in \mathbb{R}^{r \times 50}$  and  $\phi(\cdot) = \{1^{-1} \phi_1(\cdot), 2^{-1} \phi_2(\cdot), \dots, 50^{-1} \phi_{50}(\cdot)\}^\top$ . Note that  $\mathbb{E}\{\mathbf{f}_t(\cdot)\} = \mathbf{0}$  and

$$\Sigma_f(u, v) = \text{Cov}\{\Xi_t \phi(u), \Xi_t \phi(v)\} = \sum_{i=1}^{50} i^{-2} \phi_i(u) \phi_i(v) \text{Var}(\xi_{ti}),$$

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

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

Similarly,  $\Sigma_\varepsilon(u, v) = \sum_{l=1}^{25} 2^{-l} \phi_l(u) \phi_l(v) \text{Var}(\psi_{tl}) = \sum_{l=1}^{25} 3^{-1} 2^{2-l} \phi_l(u) \phi_l(v) \mathbf{C}_\zeta$ , provided that  $\text{Var}(\psi_{tl}) = \sum_{s=0}^{\infty} 0.5^{2s} \mathbf{C}_\zeta = 4\mathbf{C}_\zeta/3$ . Hence we have

$$\Sigma_y(u, v) = \mathbf{B} \Sigma_f(u, v) \mathbf{B}^\top + \Sigma_\varepsilon(u, v) = \mathbf{B} \left\{ \sum_{i=1}^{50} i^{-2} \phi_i(u) \phi_i(v) (\mathbf{I}_r - \mathbf{A}^2)^{-1} \right\} \mathbf{B}^\top + \sum_{l=1}^{25} 3^{-1} 2^{2-l} \phi_l(u) \phi_l(v) \mathbf{C}_\zeta.$$

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

(iii) To see the functional sparsity condition on  $\Sigma_\varepsilon$ , notice that

$$\sigma_i(u) = \Sigma_{\varepsilon, ii}(u, u) = \sum_{l=1}^{25} 3^{-1} 2^{2-l} \phi_l(u)^2 D_i^2 (1 + \tilde{\delta}) \quad \text{and} \quad \|\sigma_i\|_{\mathcal{N}} = \int \sigma_i(u) du = \tilde{c} D_i^2,$$

where  $\tilde{c} = 3^{-1}(4 - 2^{-23})(1 + \tilde{\delta})$  is a constant. Then, for  $q = 0$  in (7), we have

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

□

## S.5 Further derivations and definitions

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

### S.5.1 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  $\mathbf{Y}(\cdot) = \{\mathbf{y}_1(\cdot), \dots, \mathbf{y}_n(\cdot)\} \in \mathbb{R}^{p \times n}$ , and  $\mathbf{F}(u)^{\top} = \{\mathbf{f}_1(\cdot), \dots, \mathbf{f}_n(\cdot)\} \in \mathbb{R}^{r \times n}$ . Consider solving the least-squares minimization problem

$$\arg \min_{\mathbf{B}, \mathbf{F}(\cdot)} \int \|\mathbf{Y}(u) - \mathbf{B}\mathbf{F}(u)^{\top}\|_{\mathbb{F}}^2 du = \arg \min_{\mathbf{B}, \mathbf{F}(\cdot)} \sum_{t=1}^n \|\mathbf{y}_t - \mathbf{B}\mathbf{f}_t\|^2, \quad (\text{S.10})$$

subject to the normalization  $p^{-1}\mathbf{B}^{\top}\mathbf{B} = \mathbf{I}_r$ . Following the similar procedure in Section 2.2, we obtain that, for each given  $\mathbf{B}$ , the constrained least squares estimator  $\tilde{\mathbf{F}}(\cdot) = p^{-1}\mathbf{B}^{\top}\mathbf{Y}(\cdot)$ . Plugging this into (S.10), objective function becomes  $\int \text{tr}[(\mathbf{I}_p - p^{-1}\mathbf{B}\mathbf{B}^{\top})\mathbf{Y}(u)\mathbf{Y}(u)^{\top}]du$ , whose minimizer is equivalent to the maximizer of  $\text{tr}\{\mathbf{B}^{\top}[\int \mathbf{Y}(u)\mathbf{Y}(u)^{\top}du]\mathbf{B}\}$ . Apparently,  $\hat{\mathbf{B}}/\sqrt{p}$  are the eigenvectors corresponding to the  $r$  largest eigenvalues of the  $p \times p$  matrix  $\int \mathbf{Y}(u)\mathbf{Y}(u)^{\top}du = n \int \hat{\Sigma}_y^{\mathcal{S}}(u, u)du$ .

For the DIGIT method, the loading matrix  $\mathbf{B}$  is estimated by the eigenanalysis of  $\int \int \hat{\Sigma}_y^{\mathcal{S}}(u, v)\hat{\Sigma}_y^{\mathcal{S}}(u, v)^{\top}dudv$ , while the above shows that minimizing the least squares criterion (S.10) is equivalent to performing eigenanalysis of  $\int \hat{\Sigma}_y^{\mathcal{S}}(u, u)du$ . By comparison, the DIGIT method contains more covariance information by taking into account not only the diagonal entries  $\hat{\Sigma}_y^{\mathcal{S}}(u, u)$  but also the off-diagonal entries  $\hat{\Sigma}_y^{\mathcal{S}}(u, v)$  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.

### S.5.2 Relationship between two FFMs

Notice that  $\chi_t(\cdot) = \mathbf{B}\mathbf{f}_t(\cdot)$  and  $\kappa_t(\cdot) = \mathbf{Q}(\cdot)\gamma_t$  are the common components of the two FFMs (1) and (2), respectively. The covariance matrix function of  $\chi_t(\cdot)$  is

$$\Sigma_{\chi}(u, v) = \mathbf{B}\Sigma_f(u, v)\mathbf{B}^{\top} = \mathbf{B}\left\{\sum_{i=1}^{\infty} \omega_i \phi_i(u)\phi_i(v)^{\top}\right\}\mathbf{B}^{\top} = \sum_{i=1}^{\infty} p\omega_i \psi_i(u)\psi_i(v)^{\top}, \quad (\text{S.11})$$

where, by Mercer's theorem,  $\Sigma_f(u, v) = \sum_{i=1}^{\infty} \omega_i \phi_i(u)\phi_i(v)^{\top}$  and  $\psi_i(\cdot) = \mathbf{B}\phi_i(\cdot)/\sqrt{p}$ . Suppose that Assumption A.1 is satisfied with  $\Sigma_{\gamma} = \text{diag}(\check{\vartheta}_1, \dots, \check{\vartheta}_r)$ . The covariance matrix function of  $\kappa_t(\cdot)$  is

$$\Sigma_{\kappa}(u, v) = \mathbf{Q}(u)\Sigma_{\gamma}\mathbf{Q}(v)^{\top} = \sum_{j=1}^r \check{\vartheta}_j \check{\mathbf{q}}_j(u)\check{\mathbf{q}}_j(v)^{\top} = \sum_{j=1}^r p\check{\vartheta}_j \boldsymbol{\nu}_j(u)\boldsymbol{\nu}_j(v)^{\top}, \quad (\text{S.12})$$



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

$$\begin{aligned} \int \boldsymbol{\psi}_i(u)^\top \boldsymbol{\psi}_j(u) du &= \int \boldsymbol{\phi}_i(u)^\top p^{-1} \mathbf{B}^\top \mathbf{B} \boldsymbol{\phi}_j(u) du = \int \boldsymbol{\phi}_i(u)^\top \boldsymbol{\phi}_j(u) du = I(i=j), \text{ and} \\ \int \boldsymbol{\nu}_i(u)^\top \boldsymbol{\nu}_j(u) du &= p^{-1} \int \check{\mathbf{q}}_i(u)^\top \check{\mathbf{q}}_j(u) du = I(i=j) \text{ from Assumption A.1.} \end{aligned}$$

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  $\{p\check{\nu}_j\}_{j=1}^\infty$  satisfying  $\check{\nu}_j = 0$  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  $r$ .

### S.5.3 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  $\Sigma_0 : \mathbb{H} \rightarrow \mathbb{H}$  be a covariance operator. Then  $x_t(\cdot)$  is a sub-Gaussian process if there exists a constant  $\alpha \geq 0$  such that for all  $x \in \mathbb{H}$ ,  $\mathbb{E}(\exp\langle x, x_t \rangle) \leq \exp\{\alpha^2 \langle x, \Sigma_0(x) \rangle / 2\}$ .

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  $\mathbf{z}_t(\cdot) = \{z_{t1}(\cdot), \dots, z_{tp}(\cdot)\}^\top \in \mathbb{H}^p$  admits the representation

$$\mathbf{z}_t(\cdot) = \sum_{l=0}^{\infty} \mathbf{A}_l(\mathbf{x}_{t-l}), \quad t \in \mathbb{Z}, \quad (\text{S.13})$$

where  $\mathbf{A}_l = (A_{l,ij})_{p \times p}$  with each  $A_{l,ij} \in \mathbb{S}$  and  $\mathbf{x}_t(\cdot) = \{x_{t1}(\cdot), \dots, x_{tp}(\cdot)\}^\top \in \mathbb{H}^p$ , 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.13) and Definition S.1.

### S.5.4 Optimal functional portfolio allocation

In this section, we derive the optimal functional portfolio allocation  $\hat{\mathbf{w}}(\cdot)$  that is required in Section 6. Specifically, we aim to solve the following constrained minimization problem:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{H}^p} \langle \mathbf{w}, \hat{\boldsymbol{\Sigma}}_y(\mathbf{w}) \rangle \quad \text{subject to} \quad \int_{\mathcal{U}} \mathbf{w}(u)^\top \mathbf{1}_p du = 1.$$

To solve this, we apply the method of Lagrange multipliers by defining the Lagrangian function as

$$L(\mathbf{w}, \lambda) = \int_{\mathcal{U}} \int_{\mathcal{U}} \mathbf{w}(u)^T \hat{\Sigma}_y(u, v) \mathbf{w}(v) du dv - 2\lambda \int_{\mathcal{U}} \{\mathbf{w}(u)^T \mathbf{1}_p - 1\} du,$$

where  $\lambda \in \mathbb{R}$  is the Lagrange multiplier. Setting the functional derivative of  $L(\mathbf{w}, \lambda)$  with respect to  $\mathbf{w}(\cdot)$  to zero, i.e.,  $2 \int_{\mathcal{U}} \hat{\Sigma}_y(u, v) \hat{\mathbf{w}}(v) dv - 2\lambda \cdot \mathbf{1}_p = 0$  for  $u \in \mathcal{U}$ , we obtain that

$$\hat{\mathbf{w}}(u) = \lambda \cdot \hat{\Sigma}_y^{-1}(\mathbf{1}_p)(u) = \lambda \int_{\mathcal{U}} \hat{\Sigma}_y^{-1}(u, v) \mathbf{1}_p dv, \quad u \in \mathcal{U}.$$

With the constraint  $\int_{\mathcal{U}} \hat{\mathbf{w}}(u)^T \mathbf{1}_p du = 1$  and the use of Moore–Penrose inverse, it yields the desired solution

$$\hat{\mathbf{w}}(u) = \frac{\int_{\mathcal{U}} \hat{\Sigma}_y^{\dagger}(u, v) \mathbf{1}_p dv}{\int_{\mathcal{U}} \int_{\mathcal{U}} \mathbf{1}_p^T \hat{\Sigma}_y^{\dagger}(z, v) \mathbf{1}_p dz dv}, \quad u \in \mathcal{U}.$$

### S.5.5 Optimal portfolio allocation based on POET

Instead of modeling the CIDR data as  $p$ -vector of functional time series, we can treat the data at each intraday time point  $u_k$  as  $p$ -vector time series, i.e.,  $\{\mathbf{y}_t(u_k)\}_{t \in [n]}$  for  $k \in [K]$ , where  $K$  denotes the number of intraday time points. For each vector time series, we assume the following factor model:

$$\mathbf{y}_t(u_k) = \mathbf{B}_k \mathbf{f}_{k,t} + \boldsymbol{\varepsilon}_{k,t}, \quad t \in [n], k \in [K].$$

Then, the standard POET estimator (Fan et al., 2013) can be applied to obtain the estimated covariance matrix of  $\mathbf{y}_t(u_k)$  as  $\hat{\Sigma}_y(u_k, u_k)$ .

To incorporate the non-functional method into our functional risk management formulation, we also need to estimate the cross-covariance matrix  $\Sigma_y(u_k, u_l) = \text{Cov}\{\mathbf{y}_t(u_k), \mathbf{y}_t(u_l)\}$  for  $k \neq l \in [K]$ . Assuming that  $\text{Cov}(\mathbf{f}_{k,t}, \boldsymbol{\varepsilon}_{l,t}) = \mathbf{0}$  for  $k, l \in [K]$ , it follows that  $\Sigma_y(u_k, u_l) = \mathbf{B}_k \text{Cov}(\mathbf{f}_{k,t}, \mathbf{f}_{l,t}) \mathbf{B}_l + \text{Cov}(\boldsymbol{\varepsilon}_{k,t}, \boldsymbol{\varepsilon}_{l,t})$ . Thus,  $\Sigma_y(u_k, u_l)$  can be estimated as

$$\hat{\Sigma}_y(u_k, u_l) = \hat{\mathbf{B}}_k \left( n^{-1} \sum_{t=1}^n \hat{\mathbf{f}}_{k,t} \hat{\mathbf{f}}_{l,t}^T \right) \hat{\mathbf{B}}_l^T + n^{-1} \sum_{t=1}^n \hat{\boldsymbol{\varepsilon}}_{k,t} \hat{\boldsymbol{\varepsilon}}_{l,t}^T,$$

where  $\hat{\mathbf{B}}_k$ ,  $\{\hat{\mathbf{f}}_{k,t}\}_{t \in [n]}$  and  $\{\hat{\boldsymbol{\varepsilon}}_{k,t}\}_{t \in [n]}$  are obtained by the POET method. With the (cross-)covariance matrix estimates  $\hat{\Sigma}_y(u_k, u_l)$  for  $k, l \in [K]$ , the optimal portfolio allocation can be obtained by minimizing the perceived portfolio risk:

$$\{\hat{\mathbf{w}}(u_k)\}_{k \in [K]} = \arg \min_{\{\mathbf{w}(u_k)\}_{k \in [K]}} \frac{1}{K^2} \sum_{k=1}^K \sum_{l=1}^K \mathbf{w}(u_k)^T \hat{\Sigma}_y(u_k, u_l) \mathbf{w}(u_l) \quad \text{s.t.} \quad \frac{1}{K} \sum_{k=1}^K \mathbf{w}(u_k)^T \mathbf{1}_p = 1. \quad (\text{S.14})$$

Define the matrix  $\tilde{\Sigma}_y = (\hat{\Sigma}_y(u_k, u_l)) \in \mathbb{R}^{Kp \times Kp}$ , whose  $(k, l)$ -th block is  $\hat{\Sigma}_y(u_k, u_l)$ . Following similar

derivations as in Section S.5.4, the solution of (S.14) is given by

$$\{\widehat{\mathbf{w}}(u_1)^T, \dots, \widehat{\mathbf{w}}(u_K)^T\}^T = \frac{\widetilde{\boldsymbol{\Sigma}}_y^{-1} \mathbf{1}_{Kp}}{\mathbf{1}_{Kp}^T \widetilde{\boldsymbol{\Sigma}}_y^{-1} \mathbf{1}_{Kp}} \in \mathbb{R}^{Kp}.$$

## S.6 Additional simulation results

This section provides additional results supporting Section 5. Figure S.1 presents boxplots of  $\Delta\text{IC}_i$  ( $i \in [3]$ ) for two DGPs under the setting  $p = 100, n = 50, \alpha = 0.05, 0.1$ , and  $r = 3, 5, 7$ . Table S.1 reports the model selection accuracies for different values of  $\alpha$ . The results show that the proposed criteria can select the correct model with high probability even when  $\alpha$  and  $n$  are relatively small.

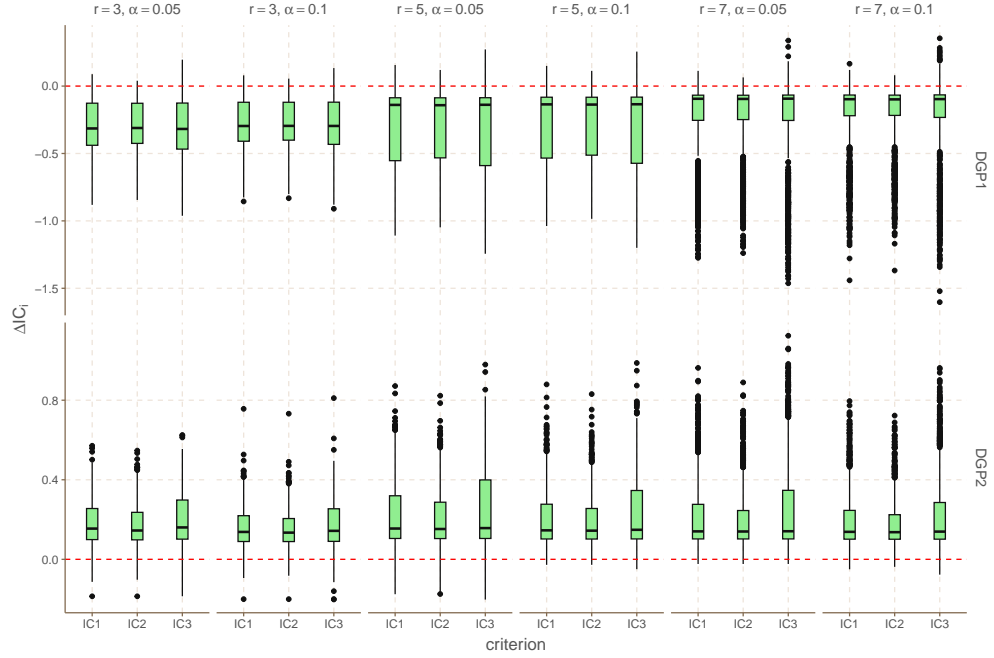


Figure S.1: The boxplots of  $\Delta\text{IC}_i$  ( $i \in [3]$ ) for DGP1 and DGP2 with  $p = 100, n = 50, \alpha = 0.05, 0.1$  and  $r = 3, 5, 7$  over 1000 simulation runs.

Table S.1: The average relative frequency estimates of  $\mathbb{P}\{\text{IC}^{\mathcal{D}}(\hat{r}^{\mathcal{D}}) < \text{IC}^{\mathcal{F}}(\hat{r}^{\mathcal{F}})\}$  for DGP1, and  $\mathbb{P}\{\text{IC}^{\mathcal{D}}(\hat{r}^{\mathcal{D}}) < \text{IC}^{\mathcal{F}}(\hat{r}^{\mathcal{F}})\}$  for DGP2 with  $p = 100, n = 50, 100, \alpha = 0.05, 0.1, 0.25, 0.5$  and  $r = 3, 5, 7$  over 1000 simulation runs.

$\alpha$	$n$	criterion	$r = 3$		$r = 5$		$r = 7$	
			DGP1	DGP2	DGP1	DGP2	DGP1	DGP2
0.05	50	$\Delta\text{IC}_1$	0.990	0.977	0.984	0.988	0.977	0.998
		$\Delta\text{IC}_2$	0.993	0.980	0.992	0.989	0.988	0.998
		$\Delta\text{IC}_3$	0.979	0.972	0.976	0.986	0.962	0.997
	100	$\Delta\text{IC}_1$	0.994	0.974	0.994	0.992	0.989	0.997
		$\Delta\text{IC}_2$	0.998	0.975	0.996	0.993	0.996	0.996
		$\Delta\text{IC}_3$	0.993	0.971	0.991	0.992	0.985	0.995
0.10	50	$\Delta\text{IC}_1$	0.992	0.974	0.975	0.993	0.980	0.998
		$\Delta\text{IC}_2$	0.998	0.976	0.984	0.996	0.984	0.997
		$\Delta\text{IC}_3$	0.988	0.962	0.965	0.991	0.962	0.994
	100	$\Delta\text{IC}_1$	0.996	0.974	0.995	0.990	0.995	0.991
		$\Delta\text{IC}_2$	1.000	0.974	0.999	0.990	1.000	0.992
		$\Delta\text{IC}_3$	0.993	0.971	0.990	0.989	0.988	0.991
0.25	50	$\Delta\text{IC}_1$	0.996	0.983	0.980	0.995	0.973	0.996
		$\Delta\text{IC}_2$	0.996	0.985	0.985	0.997	0.984	0.998
		$\Delta\text{IC}_3$	0.991	0.982	0.973	0.992	0.959	0.996
	100	$\Delta\text{IC}_1$	0.998	0.980	0.993	0.989	0.993	0.994
		$\Delta\text{IC}_2$	0.999	0.981	0.999	0.989	0.998	0.994
		$\Delta\text{IC}_3$	0.997	0.978	0.991	0.989	0.987	0.992
0.50	50	$\Delta\text{IC}_1$	0.998	0.997	0.995	1.000	0.994	1.000
		$\Delta\text{IC}_2$	0.999	0.998	0.996	1.000	0.995	1.000
		$\Delta\text{IC}_3$	0.998	0.997	0.993	1.000	0.990	1.000
	100	$\Delta\text{IC}_1$	1.000	0.998	0.999	1.000	0.998	1.000
		$\Delta\text{IC}_2$	1.000	0.998	1.000	1.000	0.999	1.000
		$\Delta\text{IC}_3$	1.000	0.998	0.998	1.000	0.997	1.000

We conduct additional simulations with data generated using  $p = 100, n = 50, \alpha = 0.5$  and  $r = 3$  for DGP1 and DGP2. The identified number of factors is set to  $\hat{r} = 1, 3, 5, 7$  when calculating  $\text{IC}_i^{\mathcal{D}}$  and  $\text{IC}_i^{\mathcal{F}}$ , with the results reported in Figure S.2. It is observed that the proposed criteria exhibit strong robustness against the misidentification of the number of factors.

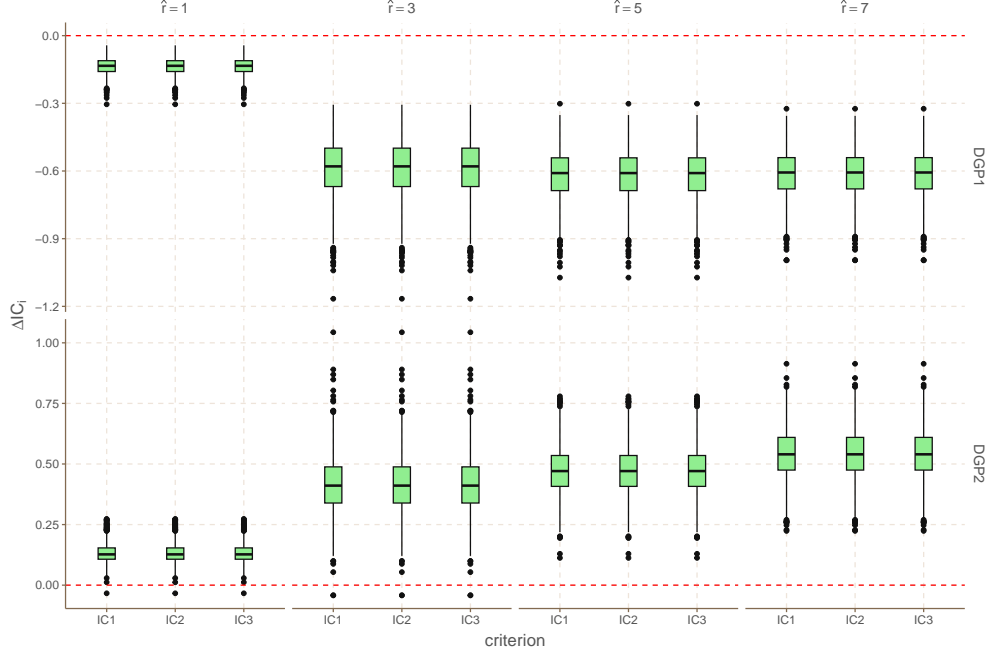


Figure S.2: The boxplots of  $\Delta IC_i$  ( $i \in [3]$ ) for DGP1 and DGP2 with  $p = 100, n = 50, \alpha = 0.5, r = 3$ , and the number of factors to calculate  $IC_i^D$  and  $IC_i^F$  being fixed as  $\hat{r} = 1, 3, 5, 7$  over 1000 simulation runs.

The model selection criteria assessed above use the same penalty function when calculating both  $IC^D$  and  $IC^F$ . According to the discussion in Section 2.4, different penalty functions from the perspective of model complexity can be applied with  $g^D(p, n) = (p + ns_n - k)/pn$  for  $IC^D(k)$  and  $g^F(p, n) = (ps_n + n - k)/pn$  for  $IC^F(k)$ . When  $\mathcal{M}_\varepsilon = o\{\sqrt{\log(p \wedge n)}\}$  and  $p \asymp n$ , it can be shown that  $s_n = \log n$  guarantees that both  $g^D(p, n)$  and  $g^F(p, n)$  satisfy the conditions in Theorem 3. Table S.2 presents the numerical summaries for the model selection accuracies when different penalty functions are employed. The results show that the proposed criteria remain effective.

Table S.2: The average relative frequency estimates of  $\mathbb{P}\{IC^D(\hat{r}^D) < IC^F(\hat{r}^F)\}$  for DGP1, and  $\mathbb{P}\{IC^D(\hat{r}^D) < IC^F(\hat{r}^F)\}$  for DGP2 with  $p = 100, n = 50, 100, \alpha = 0.05, 0.1, 0.25, 0.5, r = 3, 5, 7$  and different penalty functions for  $IC^D$  and  $IC^F$  over 1000 simulation runs.

$\alpha$	$n$	$r = 3$		$r = 5$		$r = 7$	
		DGP1	DGP2	DGP1	DGP2	DGP1	DGP2
0.05	50	0.994	0.945	0.971	0.982	0.961	0.988
	100	0.998	0.980	0.991	0.994	0.985	0.998
0.10	50	0.991	0.941	0.979	0.966	0.973	0.990
	100	0.991	0.971	0.993	0.986	0.987	0.996
0.25	50	0.995	0.956	0.980	0.987	0.972	0.992
	100	0.998	0.979	0.998	0.992	0.989	0.997
0.50	50	0.999	0.996	0.991	0.999	0.993	1.000
	100	0.999	1.000	1.000	1.000	0.997	1.000

Then we examine the performance of model selection criteria when there exist moderate to even strong correlations in idiosyncratic components. Specifically, we generate new idiosyncratic components  $\tilde{\varepsilon}_t(\cdot) = cp^{\delta/2}\varepsilon_t(\cdot)$  for  $\delta \in (0, 1)$  and  $c = 2$  (or 0.6) for DGP1 (or DGP2) which results in  $\|\Sigma_{\tilde{\varepsilon}}\|_{\mathcal{L}} = O(p^\delta)$ . The model selection accuracies are presented in Table S.3. The results indicate that the model selection accuracies decrease (particularly noticeable for DGP2) as  $\delta$  increases.

Table S.3: The average relative frequency estimates of  $\mathbb{P}\{\text{IC}^{\mathcal{D}}(\hat{r}^{\mathcal{D}}) < \text{IC}^{\mathcal{F}}(\hat{r}^{\mathcal{F}})\}$  for DGP1, and  $\mathbb{P}\{\text{IC}^{\mathcal{D}}(\hat{r}^{\mathcal{D}}) < \text{IC}^{\mathcal{F}}(\hat{r}^{\mathcal{F}})\}$  for DGP2 with  $p = 100, n = 50, 100, \delta = 0.25, 0.5, \alpha = 0.5$  and  $r = 3, 5, 7$  over 1000 simulation runs.

$\delta$	$n$	criterion	$r = 3$		$r = 5$		$r = 7$	
			DGP1	DGP2	DGP1	DGP2	DGP1	DGP2
0.25	50	$\Delta\text{IC}_1$	0.964	0.997	0.961	1.000	0.941	1.000
		$\Delta\text{IC}_2$	0.958	0.997	0.948	1.000	0.938	1.000
		$\Delta\text{IC}_3$	0.809	0.997	0.832	1.000	0.827	1.000
	100	$\Delta\text{IC}_1$	1.000	1.000	1.000	0.999	1.000	1.000
		$\Delta\text{IC}_2$	0.998	1.000	1.000	0.999	1.000	1.000
		$\Delta\text{IC}_3$	0.939	1.000	0.962	0.999	0.954	1.000
0.50	50	$\Delta\text{IC}_1$	0.956	0.767	0.925	0.878	0.892	0.920
		$\Delta\text{IC}_2$	0.947	0.734	0.910	0.869	0.882	0.921
		$\Delta\text{IC}_3$	0.818	0.779	0.776	0.873	0.735	0.916
	100	$\Delta\text{IC}_1$	0.999	0.832	0.998	0.935	0.997	0.964
		$\Delta\text{IC}_2$	0.992	0.819	0.986	0.925	0.984	0.958
		$\Delta\text{IC}_3$	0.947	0.829	0.946	0.917	0.941	0.946

We compare our AFT estimator in (10) with two related methods for estimating the idiosyncratic covariance  $\Sigma_\varepsilon$ , specifically, the sample covariance estimator defined as  $\hat{\Sigma}_\varepsilon^S(u, v) = n^{-1} \sum_{t=1}^n \hat{\varepsilon}_t(u) \hat{\varepsilon}_t(v)^\top$ , and Fang et al. (2024)'s AFT estimator in (11). Figures S.3 and S.4 plot average losses of  $\hat{\Sigma}_\varepsilon$  measured by functional matrix  $\ell_1$  norm and operator norm for DGP1 and DGP2, respectively, under the settings  $n = p = 60, 80, \dots, 200$  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  $\hat{\Sigma}_\varepsilon^S$ . Second, despite our AFT proposal requiring weaker assumptions compared to Fang et al. (2024)'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 do not show significant upward trends. This phenomenon can be attributed to the fact that  $\{(\log p/n)^{1/2} + p^{-1/2}\}p^{1-\alpha} = o(1)$  as  $p, n \rightarrow \infty$  if  $\alpha > 0.5$ , which is implied by Theorems 4 and 4' under the setting  $n = p, q = 0, \mathcal{M}_\varepsilon = O(1)$ .

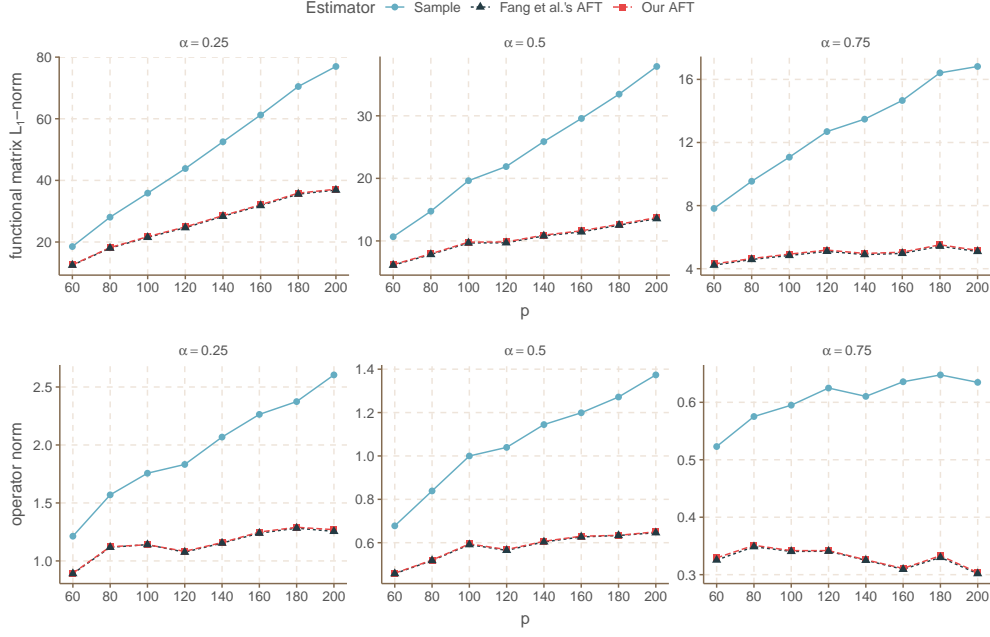


Figure S.3: The average losses of  $\hat{\Sigma}_\varepsilon$  in functional matrix  $\ell_1$  norm (top row) and operator norm (bottom row) for DGP1 over 1000 simulation runs with  $n = p = 60, 80, \dots, 200$  and  $\alpha = 0.25, 0.5, 0.75$ .

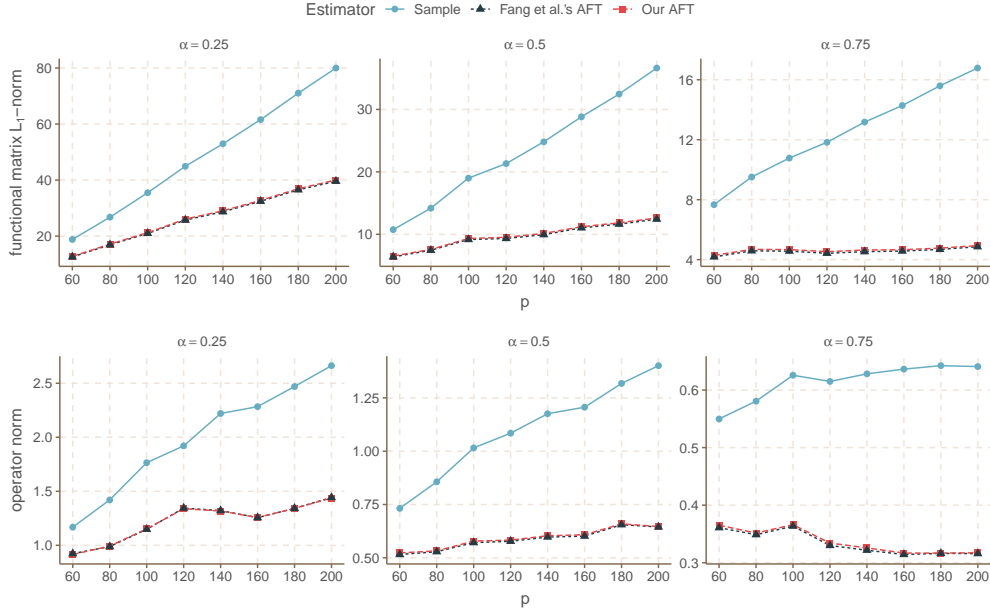


Figure S.4: The average losses of  $\hat{\Sigma}_\varepsilon$  in functional matrix  $\ell_1$  norm (top row) and operator norm (bottom row) for DGP2 over 1000 simulation runs with  $n = p = 60, 80, \dots, 200$  and  $\alpha = 0.25, 0.5, 0.75$ .

Figures S.5 and S.6 plot average losses of  $\hat{\Sigma}_y$  measured by functional versions of elementwise  $\ell_\infty$  norm, Frobenius norm and matrix  $\ell_1$  norm for DGP1 and DGP2, respectively, when  $\dot{C} = 1$ .



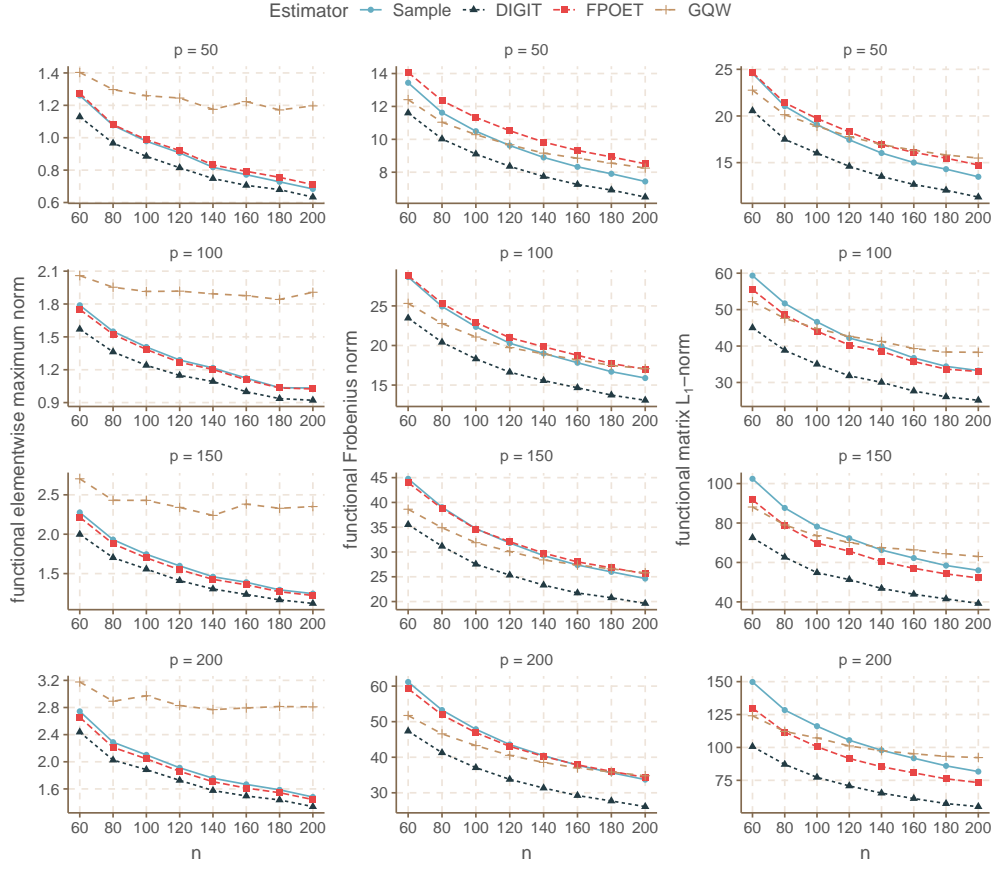


Figure S.5: The average losses of  $\hat{\Sigma}_y$  in functional versions of elementwise  $\ell_\infty$  norm (left column), Frobenius norm (middle column) and matrix  $\ell_1$  norm (right column) for DGP1 with  $\dot{C} = 1$  over 1000 simulation runs.

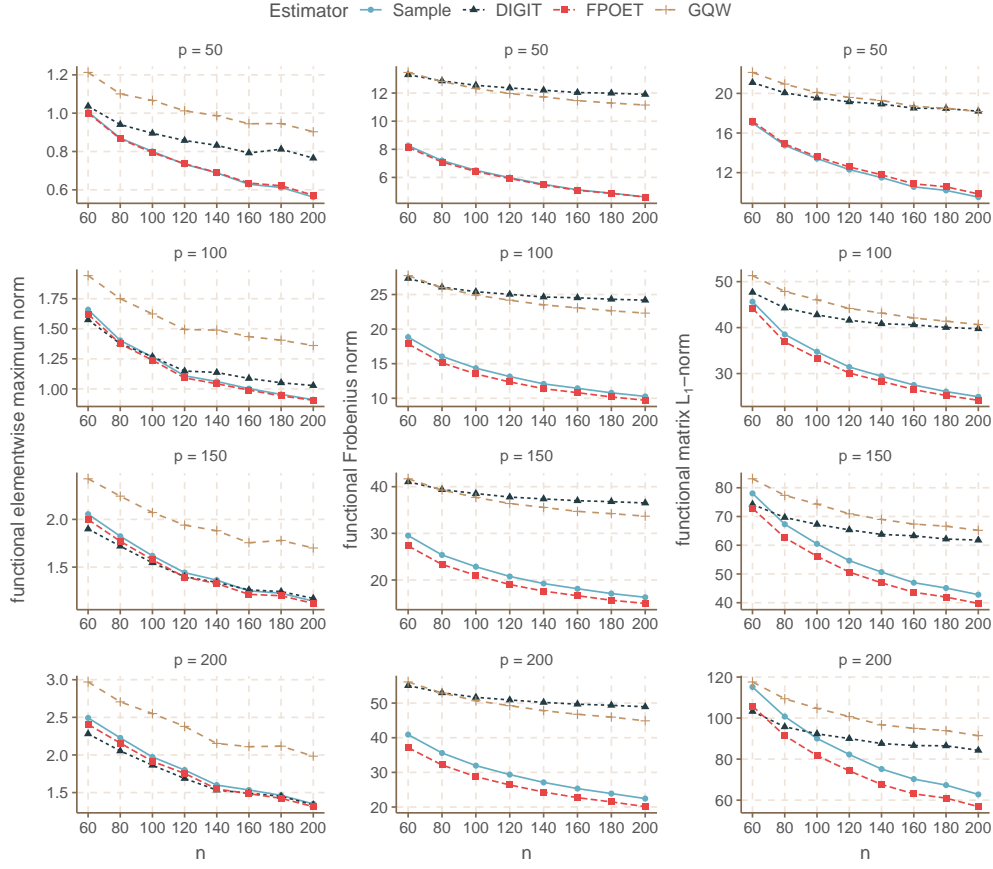


Figure S.6: The average losses of  $\hat{\Sigma}_y$  in functional versions of elementwise  $\ell_\infty$  norm (left column), Frobenius norm (middle column) and matrix  $\ell_1$  norm (right column) for DGP2 with  $\dot{C} = 1$  over 1000 simulation runs.

## S.7 Additional real data result

Table S.4 presents the list of S&P 100 component stocks used in Section 6.

Table S.4: List of S&amp;P 100 stocks.

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

## 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.
- 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.
- 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.
- 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. (2024). Adaptive functional thresholding for sparse covariance function estimation in high dimensions, *Journal of the American Statistical Association* **119**: 1473–1485.
- Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, John Wiley & Sons, Chichester.
- Zhang, D. and Wu, W. B. (2021). Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes, *The Annals of Statistics* **49**: 233–254.