Journal of Econometrics 239 (2024) 105385

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

An autocovariance-based learning framework for high-dimensional functional time series☆

Jinyuan Chang [a], Cheng Chen [a], Xinghao Qiao [b,*], Qiwei Yao [b]

[a] Joint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics, Chengdu, Sichuan Province 611130, China
[b] Department of Statistics, London School of Economics, London, WC2A 2AE, UK

ARTICLE INFO

Article history:
Received 30 August 2021
Received in revised form 20 August 2022
Accepted 13 January 2023
Available online 1 February 2023

JEL classification:
C13
C32
C55

Keywords:
Block regularized minimum distance estimation
Dimension reduction
Functional time series
High-dimensional data
Non-asymptotics
Sparsity

ABSTRACT

Many scientific and economic applications involve the statistical learning of high-dimensional functional time series, where the number of functional variables is comparable to, or even greater than, the number of serially dependent functional observations. In this paper, we model observed functional time series, which are subject to errors in the sense that each functional datum arises as the sum of two uncorrelated components, one dynamic and one white noise. Motivated from the fact that the autocovariance function of observed functional time series automatically filters out the noise term, we propose a three-step framework by first performing autocovariance-based dimension reduction, then formulating a novel autocovariance-based block regularized minimum distance estimation to produce block sparse estimates, and based on which obtaining the final functional sparse estimates. We investigate theoretical properties of the proposed estimators, and illustrate the proposed estimation procedure with the corresponding convergence analysis via three sparse high-dimensional functional time series models. We demonstrate via both simulated and real datasets that our proposed estimators significantly outperform their competitors.

1. Introduction

Functional time series refers to functional data objects that are observed consecutively over time. Existing research on functional time series has mainly focused on extending the univariate or low-dimensional multivariate time series methods to the functional domain. An incomplete list of the relevant references includes Bosq (2000), Bathia et al. (2010), Hörmann and Kokoszka (2010), Panaretos and Tavakoli (2013), Aue et al. (2015), Hörmann et al. (2015), Li et al. (2020) and Chen et al. (2022). The rapid development of data collection technology has made high-dimensional functional time series datasets increasingly common. Examples include hourly measured concentrations of various pollutants such as PM10 trajectories (Hörmann et al., 2015) collected at different measuring stations, daily electricity load curves (Cho et al., 2013) for a large number of households, cumulative intraday return trajectories (Horváth et al., 2014), daily return density curves (Bathia et al., 2010) and functional volatility processes (Müller et al., 2011) for a collection of stocks.

☆ The authors equally contributed to the paper. We thank the editor, the associate editor and two anonymous referees for their constructive comments and suggestions. Chang and Chen were supported in part by the National Natural Science Foundation of China (grant nos. 71991472, 72125008 and 11871401). Chang was also supported by the Center of Statistical Research at Southwestern University of Finance and Economics. Yao was supported in part by the U.K. Engineering and Physical Sciences Research Council (grant no. EP/V007556/1).

* Corresponding author.
E-mail addresses: changjinyuan@swufe.edu.cn (J. Chang), chenc@swufe.edu.cn (C. Chen), x.qiao@lse.ac.uk (X. Qiao), q.yao@lse.ac.uk (Q. Yao).

https://doi.org/10.1016/j.jeconom.2023.01.007

We consider in this paper a setting for modelling high-dimensional functional time series as follows. Let $\mathbf{W}_t(\cdot) = \{W_{t1}(\cdot), \ldots, W_{tp}(\cdot)\}^\top$, $t = 1, \ldots, n$, be the observed p-vector of functional time series defined on a compact interval $\mathcal{U}$, where the dimension p is large in relation to n, and p may be greater than n. Suppose that $\mathbf{W}_t(\cdot)$ is subject to an error:

$$\mathbf{W}_t(\cdot) = \mathbf{X}_t(\cdot) + \mathbf{e}_t(\cdot), \tag{1}$$

where $\mathbf{X}_t(\cdot) = \{X_{t1}(\cdot), \ldots, X_{tp}(\cdot)\}^\top$ is a functional time series of interest, $\mathbf{e}_t(\cdot) = \{e_{t1}(\cdot), \ldots, e_{tp}(\cdot)\}^\top$ is white noise in the sense (3) below, and $\{\mathbf{X}_t(\cdot)\}_{t=1}^n$ and $\{\mathbf{e}_t(\cdot)\}_{t=1}^n$ are uncorrelated. Note that both $\mathbf{X}_t(\cdot)$ and $\mathbf{e}_t(\cdot)$ are latent. We assume that both $\mathbf{W}_t(\cdot)$ and $\mathbf{X}_t(\cdot)$ are weakly stationary, and $\mathbb{E}\{\mathbf{W}_t(u)\} = \mathbf{0}$ for any $u \in \mathcal{U}$. For any integer h and $u, v \in \mathcal{U}$, put

$$\boldsymbol{\Sigma}_h^W(u, v) = \mathrm{Cov}\{\mathbf{W}_{t-h}(u), \mathbf{W}_t(v)\}, \quad \boldsymbol{\Sigma}_h^X(u, v) = \mathrm{Cov}\{\mathbf{X}_{t-h}(u), \mathbf{X}_t(v)\}, \quad \boldsymbol{\Sigma}_h^e(u, v) = \mathrm{Cov}\{\mathbf{e}_{t-h}(u), \mathbf{e}_t(v)\}. \tag{2}$$

We call $\mathbf{e}_t(\cdot)$ a white noise if

$$\mathbb{E}\{\mathbf{e}_t(u)\} = \mathbf{0} \text{ and } \boldsymbol{\Sigma}_h^e(u, v) = \mathbf{0} \text{ for any } u, v \in \mathcal{U} \text{ and } h \neq 0. \tag{3}$$

Furthermore, we assume that $\mathbf{a}^\top \mathbf{X}_t(\cdot)$ is not white noise for any non-zero constant vector $\mathbf{a} \in \mathbb{R}^p$. Under this setting, the linear dynamic structure of $\mathbf{W}_t(\cdot)$ is entirely determined by that of $\mathbf{X}_t(\cdot)$, and all white noise elements in $\mathbf{W}_t(\cdot)$ are absorbed into $\mathbf{e}_t(\cdot)$. The presence of $\mathbf{e}_t(\cdot)$ reflects that signal curves $\mathbf{X}_t(\cdot)$ are seldom completely observed. Instead, they are often only measured, with errors, on a grid. These noisy discrete data are smoothed to yield 'observed' curves $\mathbf{W}_t(\cdot)$. See Bathia et al. (2010) for the univariate version of model (1). When $\mathbf{W}_1(\cdot), \ldots, \mathbf{W}_n(\cdot)$ are univariate and independent, Hall and Vial (2006) considered the same model under a 'low noise' setting assuming that $\mathbf{e}_t(\cdot)$ goes to 0 as n grows to ∞. To separate $\mathbf{X}_t(\cdot)$ from $\mathbf{e}_t(\cdot)$, e.g., via the covariance function, even in the univariate case, some special structures were imposed; see, e.g., diagonal $\boldsymbol{\Sigma}_0^e$ of Yao et al. (2005) and banded $\boldsymbol{\Sigma}_0^e$ of Descary and Panaretos (2019). In contrast, we do not impose any structures on $\boldsymbol{\Sigma}_0^e$ in this paper, and our estimation filters out the impact of $\mathbf{e}_t(\cdot)$ automatically.

The standard estimation procedures for univariate functional time series models usually consist of three steps (Aue et al., 2015). Dimension-reduction is performed first via, e.g., functional principal components analysis (FPCA). Each observed curve is then approximated by a finite truncation. This effectively transforms functional time series into a vector time series of FPC scores. In the second step the estimation of the function-valued parameters in the model is transformed to that of some appropriate parameter vectors/matrices based on estimated FPC scores. Finally the estimated principal component functions are utilized to obtain function-valued estimates based on the estimated parameter vectors/matrices. To overcome the difficulties caused by high-dimensionality (i.e. large p in relation to n), some functional sparsity assumptions are imposed, which results in the estimation under block sparsity constraints in the second step in the sense that variables belonging to the same block (or group) are simultaneously included or excluded. In regression setups, the group-lasso penalized least squares estimation (Yuan and Lin, 2006) is often adopted in the second step to obtain block sparse estimates. Similar three-step procedures have been developed to estimate sparse high-dimensional functional models, see, e.g., vector functional autoregression (VFAR) (Guo and Qiao, 2023), scalar-on-function linear additive regression (SFLR) (Fan et al., 2015; Kong et al., 2016; Xue and Yao, 2021; Fang et al., 2022) and function-on-function linear additive regression (FFLR) (Fan et al., 2014; Luo and Qi, 2017; Fang et al., 2022). However, those estimation procedures are developed under an assumption that signal curves are observed directly.

In our setting the observed curves $\mathbf{W}_t(\cdot)$ are subject to the error contamination as in model (1). Both FPCA and penalized least squares estimation based on the estimated covariance function $\widehat{\boldsymbol{\Sigma}}_0^W$ of $\mathbf{W}_t(\cdot)$ are inappropriate since $\boldsymbol{\Sigma}_0^W = \boldsymbol{\Sigma}_0^X + \boldsymbol{\Sigma}_0^e$ and, hence, $\widehat{\boldsymbol{\Sigma}}_0^W$ is no longer a consistent estimator for $\boldsymbol{\Sigma}_0^X$. Motivated from the fact that $\boldsymbol{\Sigma}_h^W = \boldsymbol{\Sigma}_h^X$ for any $h \neq 0$, which automatically removes the impact from the noise $\mathbf{e}_t(\cdot)$ and ensures that the estimator for $\boldsymbol{\Sigma}_h^W$ is also legitimate for $\boldsymbol{\Sigma}_h^X$, we propose an autocovariance-based three-step learning procedure. Differing from FPCA based on the Karhunen–Loève expansion, our first dimension reduction step is formulated under an alternative data-driven basis expansion of each $X_{tj}(\cdot)$ based on the eigenanalysis of a positive-definite operator defined in terms of the autocovariance functions of $W_{tj}(\cdot)$. Different from the penalized least squares estimation, our second step makes use of the autocovariance of basis coefficients to construct high-dimensional moment equations and then applies the proposed block regularized method to estimate the associated block sparse parameter vectors/matrices. Our third step re-transforms block sparse estimates to functional sparse estimates via estimated basis functions obtained in the first step.

Our theoretical development stands at the intersection between high-dimensional statistics and functional time series, facing several challenges due to non-asymptotics and infinite-dimensionality with serial dependence. Firstly, in the proposed second step we deal with the estimated basis coefficients to produce block sparse estimates whereas the conventional sparse estimation is applied directly to observed data. Accounting for such approximation is a major undertaking. Secondly, under a high-dimensional and dependent setting, it is essential to develop non-asymptotic error bounds on the relevant estimated terms as a function of n, p and the truncated dimension, and to assess how the serial dependence affects non-asymptotic results. Thirdly, compared to non-functional data, the infinite-dimensional nature of functional data leads to the additional theoretical complexity that arises from specifying the block structure and controlling bias terms formed by truncation errors from the dimension reduction step.

The main contribution of our paper is three-fold.

1. Our autocovariance-based learning framework can address the error contamination model (1) in the presence of infinite-dimensional signal curve dynamics with the addition of 'genuinely functional' noise. It makes the good use of the serial correlation information, which is the most relevant in the context of time series modelling.

2. To provide theoretical guarantees for the first and the third steps and to verify imposed high-level regularity conditions in the second step, we establish useful non-asymptotic error bounds on the relevant estimated terms under the autocovariance-based dimension reduction framework.
3. We utilize the autocovariance among basis coefficients to construct high-dimensional moment equations with partitioned group structure, based on which we formulate the second step in a novel block regularized minimum distance (RMD) estimation framework to produce block sparse estimates. The group information can be explicitly encoded in a convex optimization targeting at minimizing the block ℓ_1 norm objective function subject to the block ℓ_∞ norm constraint. To theoretically support the second step, we investigate convergence properties of the block RMD estimator. Besides being useful in the second step, the block RMD estimation framework itself is of independent interest and can be applied more broadly.

Our paper is set out as follows. In Section 2, we present Step 1, i.e. the autocovariance-based dimension reduction technique. We also establish some essential deviation bounds on the relevant estimated terms. In Section 3, we first use an example to illustrate the construction of high-dimensional moment equations. We then formulate a general block RMD estimation method (i.e. Step 2) and investigate its theoretical properties. In Section 4, we illustrate the proposed three-step framework using three examples of sparse high-dimensional functional time series models, i.e. SFLR, FFLR and VFAR. Theoretically, we study convergence rates of the associated estimators in these models. In Section 5, we examine the finite-sample performance of the proposed estimators through both simulations and an analysis of a real financial dataset. All technical proofs are relegated to the Appendix.

Notation. For a positive integer q, we write $[q] = \{1, \ldots, q\}$. Let $L_2(\mathcal{U})$ be a Hilbert space of square-integrable functions on a compact interval $\mathcal{U}$. The inner product of $f, g \in L_2(\mathcal{U})$ is defined as $\langle f, g\rangle = \int_{\mathcal{U}} f(u)g(u)\,\mathrm{d}u$. For a Hilbert space $\mathbb{H} \subset L_2(\mathcal{U})$, we denote the p-fold Cartesian product by $\mathbb{H}^p = \mathbb{H} \times \cdots \times \mathbb{H}$ and the tensor product by $\mathbb{S} = \mathbb{H} \otimes \mathbb{H}$. For any $\mathbf{f} = (f_1, \ldots, f_p)^\top$ and $\mathbf{g} = (g_1, \ldots, g_p)^\top$ in $\mathbb{H}^p$, we define $\langle \mathbf{f}, \mathbf{g}\rangle = \sum_{i=1}^p \langle f_i, g_i\rangle$. We use $\|\mathbf{f}\| = \langle \mathbf{f}, \mathbf{f}\rangle^{1/2}$ and $\|\mathbf{f}\|_0 = \sum_{i=1}^p I(\|f_i\| \neq 0)$ with $I(\cdot)$ being the indicator function to denote functional versions of induced norm and ℓ_0-norm, respectively. For an integral operator $\mathbf{K} : \mathbb{H}^p \to \mathbb{H}^q$ induced from the kernel function $\mathbf{K} = (K_{ij})_{q\times p}$ with each $K_{ij} \in \mathbb{S}$, $\mathbf{K}(\mathbf{f})(u) = \{\sum_{j=1}^p \langle K_{1j}(u, \cdot), f_j(\cdot)\rangle, \ldots, \sum_{j=1}^p \langle K_{qj}(u, \cdot), f_j(\cdot)\rangle\}^\top \in \mathbb{H}^q$ for any $\mathbf{f} = (f_1, \ldots, f_p)^\top \in \mathbb{H}^p$. For notational economy, we will also use $\mathbf{K}$ to denote both the kernel and the operator. We define functional versions of Frobenius and matrix ℓ_∞-norms by $\|\mathbf{K}\|_{\mathrm{F}} = (\sum_{i=1}^q \sum_{j=1}^p \|K_{ij}\|_{\mathcal{S}}^2)^{1/2}$ and $\|\mathbf{K}\|_\infty = \max_{i\in[q]} \sum_{j=1}^p \|K_{ij}\|_{\mathcal{S}}$, respectively, where $\|K_{ij}\|_{\mathcal{S}} = \{\int_{\mathcal{U}}\int_{\mathcal{U}} K_{ij}^2(u, v)\,\mathrm{d}u\mathrm{d}v\}^{1/2}$ denotes the Hilbert–Schmidt norm of K_{ij}. For any real matrix $\mathbf{B} = (b_{ij})_{q\times p}$, we write $\|\mathbf{B}\|_{\max} = \max_{i\in[q],j\in[p]} |b_{ij}|$ and use $\|\mathbf{B}\|_{\mathrm{F}} = (\sum_{i=1}^q \sum_{j=1}^p |b_{ij}|^2)^{1/2}$ and $\|\mathbf{B}\|_2 = \lambda_{\max}^{1/2}(\mathbf{B}^\top\mathbf{B})$ to denote its Frobenius norm and ℓ_2-norm, respectively. For two sequences of positive numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ if there exists a positive constant c such that $\limsup_{n\to\infty} a_n/b_n \leq c$. We write $a_n \asymp b_n$ if and only if $a_n \lesssim b_n$ and $b_n \lesssim a_n$ hold simultaneously.

2. Autocovariance-based dimension reduction

2.1. Methodology

Our Step 1 is to approximate each curve $X_{tj}(\cdot)$ by a finite linear combination: we expand curve $X_{tj}(\cdot)$ using the data-driven orthonormal basis functions $\{\psi_{ji}(\cdot)\}_{i=1}^\infty$, and truncate the expansion to the first d_j terms:

$$X_{tj}(\cdot) = \sum_{l=1}^{\infty} \eta_{tjl}\psi_{jl}(\cdot) \approx \boldsymbol{\eta}_{tj}^\top \boldsymbol{\psi}_j(\cdot), \quad j \in [p], \tag{4}$$

where $\eta_{tjl} = \langle X_{tj}, \psi_{jl}\rangle$, $\boldsymbol{\eta}_{tj} = (\eta_{tj1}, \ldots, \eta_{tjd_j})^\top \in \mathbb{R}^{d_j}$ and $\boldsymbol{\psi}_j(\cdot) = \{\psi_{j1}(\cdot), \ldots, \psi_{jd_j}(\cdot)\}^\top$. Different from the conventional Karhunen–Loève expansion, the eigenvalues $\lambda_{j1} \geq \lambda_{j2} \geq \cdots > 0$ and the corresponding eigenfunctions $\psi_{j1}(\cdot), \psi_{j2}(\cdot), \ldots$ are taken from the spectral decomposition of an operator defined as

$$K_{jj}(u, v) = \sum_{h=1}^{L} \int_{\mathcal{U}} \Sigma_{h,jj}^X(u, z)\Sigma_{h,jj}^X(v, z)\,\mathrm{d}z, \tag{5}$$

where $L > 0$ is some prescribed fixed integer, and $\Sigma_{h,ij}^X(u, v)$ denotes the (i, j)th element of $\boldsymbol{\Sigma}_h^X(u, v)$ in (2). Also denote by $\Sigma_{h,ij}^W$ and $\Sigma_{h,ij}^e$ the (i, j)th element of, respectively, $\boldsymbol{\Sigma}_h^W$ and $\boldsymbol{\Sigma}_h^e$. The idea of using non-zero lagged autocovariances was initiated by Bathia et al. (2010). A direct consequence is the identity

$$K_{jj}(u, v) = \sum_{h=1}^{L} \int_{\mathcal{U}} \Sigma_{h,jj}^W(u, z)\Sigma_{h,jj}^W(v, z)\,\mathrm{d}z,$$

since $\Sigma_{h,jj}^X(u, z) = \Sigma_{h,jj}^W(u, z)$ for all $(u, z) \in \mathcal{U}^2$ and $h \neq 0$. This paves the way to estimate K_{jj}, and therefore also $\boldsymbol{\psi}_j(\cdot)$, directly based on observations $W_{1j}(\cdot), \ldots, W_{nj}(\cdot)$. The impact of the noise terms $e_{tj}(\cdot)$ is filtered out automatically. It is

worth noting that we choose not to use autocovariance functions $\Sigma^W_{h,jj}$ directly in defining K_{jj} as they are not nonnegative definite. The definition of K_{jj} in (5) ensures that it is nonnegative definite, and there is no cancellation of the information accumulated from lags 1 to L. Hence the estimation is not sensitive to the choice of L. In practice, we choose small L such as $1 \le L \le 5$, as the most significant autocorrelations typically occur at small lags.

In the standard Karhunen–Loève expansion, $\{\psi_{jl}(\cdot)\}_{l=1}^{\infty}$ is deduced from the spectral decomposition of $\Sigma^X_{0,jj}$. Since

$$\Sigma^X_{0,jj}(u,v) = \Sigma^W_{0,jj}(u,v) - \Sigma^e_{0,jj}(u,v)\,,$$

some strong assumptions have to be imposed to eliminate the impact of $\Sigma^e_{0,jj}(u,v)$ in order to obtain consistent estimates for $\psi_{jl}(\cdot)$. For example, Hall and Vial (2006) assumes that $W_{1j}(\cdot), \ldots, W_{nj}(\cdot)$ are independent and the noise $e_{tj}(\cdot)$ goes to 0 as n grows to ∞. Note that the dimension reduction via FPCA can also be performed based on the spectral decomposition of $\Sigma^W_{0,jj}$ instead of $\Sigma^X_{0,jj}$, as any basis could be used for expanding the data. However, because of $\Sigma^W_{0,jj} = \Sigma^X_{0,jj} + \Sigma^e_{0,jj}$, using $\Sigma^W_{0,jj}$ may require a larger truncated dimension to capture the sufficient signal information, leading to reduced statistical efficiency. It is also worth mentioning that the penalized least squares approach adopted in the covariance-based second step is based on $\Sigma^X_{0,jk}(u,v) = \Sigma^W_{0,jk}(u,v) - \Sigma^e_{0,jk}(u,v)$ and hence is inappropriate under model (1).

With the available observations $\{\mathbf{W}_t(\cdot)\}_{t\in[n]}$, a natural estimator for K_{jj} in (5) is defined as

$$\begin{aligned}\hat{K}_{jj}(u,v) &= \sum_{h=1}^{L}\int_{\mathcal{U}} \widehat{\Sigma}^W_{h,jj}(u,z)\widehat{\Sigma}^W_{h,jj}(v,z)\,\mathrm{d}z \\ &= \frac{1}{(n-L)^2}\sum_{h=1}^{L}\sum_{t,s=h+1}^{n} W_{(t-h)j}(u)W_{(s-h)j}(v)\langle W_{tj}, W_{sj}\rangle\,,\end{aligned} \tag{6}$$

where

$$\widehat{\boldsymbol{\Sigma}}^W_h(u,v) = \frac{1}{n-h}\sum_{t=h+1}^{n}\mathbf{W}_{t-h}(u)\mathbf{W}_t(v)^{\top} = \{\widehat{\Sigma}^W_{h,jk}(u,v)\}_{j,k\in[p]}\,, \quad (u,v)\in\mathcal{U}^2\,,\ h\ge 0\,. \tag{7}$$

Performing the spectral decomposition

$$\hat{K}_{jj}(u,v) = \sum_{l=1}^{\infty}\hat{\lambda}_{jl}\hat{\psi}_{jl}(u)\hat{\psi}_{jl}(v)\,, \tag{8}$$

where $\hat{\lambda}_{j1} \ge \hat{\lambda}_{j2} \ge \cdots > 0$ are the eigenvalues, and $\hat{\psi}_{j1}(\cdot)$, $\hat{\psi}_{j2}(\cdot)$, ... are the corresponding eigenfunctions.

Let $\mathbb{E}\{\boldsymbol{\eta}_{(t-h)j}\boldsymbol{\eta}_{tk}^{\top}\} = \{\sigma^{(h)}_{jklm}\}_{l\in[d_j],m\in[d_k]}$ with its estimator $(n-h)^{-1}\sum_{t=h+1}^{n}\hat{\boldsymbol{\eta}}_{(t-h)j}\hat{\boldsymbol{\eta}}_{tk}^{\top} = \{\hat{\sigma}^{(h)}_{jklm}\}_{l\in[d_j],m\in[d_k]}$ for $j,k\in[p]$ and $h\ge 0$, where $\hat{\boldsymbol{\eta}}_{tj} = (\hat{\eta}_{tj1},\ldots,\hat{\eta}_{tjd_j})^{\top}$. Our proposed autocovariance-based Step 2 and Step 3 explicitly rely on the sample autocovariance among estimated basis coefficients, $\{\hat{\sigma}^{(h)}_{jklm} : j,k\in[p], l\in[d_j], m\in[d_k], h\in[L]\}$, and the estimated basis functions $\{\hat{\psi}_{jl}(\cdot) : j\in[p], l\in[d_j]\}$, respectively. See details in Sections 3.1 and 4. Their convergence properties in elementwise ℓ_∞-norm under high-dimensional scaling are investigated in Section 2.2 below.

2.2. Rates in elementwise ℓ_∞-norm

To characterize the effect of serial dependence on the relevant estimated terms, we will use the functional stability measure of $\{\mathbf{W}_t(\cdot)\}_{t\in\mathbb{Z}}$ (Guo and Qiao, 2023).

Condition 1. For $\{\mathbf{W}_t(\cdot)\}_{t\in\mathbb{Z}}$, the spectral density operator $\mathbf{f}^W_\theta = (2\pi)^{-1}\sum_{h\in\mathbb{Z}}\boldsymbol{\Sigma}^W_h e^{-ih\theta}$ for $\theta\in[-\pi,\pi]$ exists and the functional stability measure defined in (9) is finite, i.e.

$$\mathcal{M}^W = 2\pi\cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\,\boldsymbol{\Phi}\in\mathbb{H}^p_0}\frac{\langle\boldsymbol{\Phi},\mathbf{f}^W_\theta(\boldsymbol{\Phi})\rangle}{\langle\boldsymbol{\Phi},\boldsymbol{\Sigma}^W_0(\boldsymbol{\Phi})\rangle} < \infty\,, \tag{9}$$

where $\mathbb{H}^p_0 = \{\boldsymbol{\Phi}\in\mathbb{H}^p : \langle\boldsymbol{\Phi},\boldsymbol{\Sigma}^W_0(\boldsymbol{\Phi})\rangle\in(0,\infty)\}$.

The quantity $\mathcal{M}^W$ in (9) is expressed proportional to functional Rayleigh quotients of $\mathbf{f}^W_\theta$ relative to $\boldsymbol{\Sigma}^W_0$. Hence it can more precisely capture the effect of small decaying eigenvalues of $\boldsymbol{\Sigma}^W_0$ on the numerator in (9), which is essential to handle truly infinite-dimensional functional objects $\{W_{tj}(\cdot)\}$. We next define the functional stability measure of all k-dimensional subsets of $\{\mathbf{W}_t(\cdot)\}_{t\in\mathbb{Z}}$, i.e. $\{(W_{tj}(\cdot) : j\in J)^{\top}\}_{t\in\mathbb{Z}}$ for $J\subset[p]$ with cardinality $|J|\le k$, by

$$\mathcal{M}^W_k = 2\pi\cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\,\|\boldsymbol{\Phi}\|_0\le k,\,\boldsymbol{\Phi}\in\mathbb{H}^p_0}\frac{\langle\boldsymbol{\Phi},\mathbf{f}^W_\theta(\boldsymbol{\Phi})\rangle}{\langle\boldsymbol{\Phi},\boldsymbol{\Sigma}^W_0(\boldsymbol{\Phi})\rangle}\,,\quad k\in[p]\,. \tag{10}$$

Under Condition 1, it is easy to verify that $\mathcal{M}^W_k \le \mathcal{M}^W < \infty$.

Our non-asymptotic results are developed using the infinite-dimensional analog of Hanson–Wright inequality (Rudelson and Vershynin, 2013) in a general Hilbert space $\mathbb{H}$, for which we need to impose the sub-Gaussian condition.

Definition 1. Let $Z_t(\cdot)$ be a mean zero random variable in $\mathbb{H}$ for any fixed t and $\Sigma_0 : \mathbb{H} \to \mathbb{H}$ be a covariance operator. Then $Z_t(\cdot)$ is a sub-Gaussian process if there exists a constant $c > 0$ such that $\mathbb{E}(e^{\langle x, Z\rangle}) \le e^{c^2\langle x, \Sigma_0(x)\rangle/2}$ for all $x \in \mathbb{H}$.

Condition 2. (i) $\{\mathbf{W}_t(\cdot)\}_{t\in\mathbb{Z}}$ is a sequence of multivariate functional linear processes with sub-Gaussian errors, namely sub-Gaussian functional linear processes, $\mathbf{W}_t(\cdot) = \sum_{l=0}^{\infty} \mathbf{B}_l(\boldsymbol{\varepsilon}_{t-l})$ for any $t \in \mathbb{Z}$, where $\mathbf{B}_l = (B_{l,jk})_{p\times p}$ with each $B_{l,jk} \in \mathbb{S}$, $\boldsymbol{\varepsilon}_t(\cdot) = \{\varepsilon_{t1}(\cdot), \ldots, \varepsilon_{tp}(\cdot)\}^\top \in \mathbb{H}^p$ and the components in $\{\boldsymbol{\varepsilon}_t(\cdot)\}_{t\in\mathbb{Z}}$ are independent sub-Gaussian processes satisfying Definition 1; (ii) The coefficient functions satisfy $\sum_{l=0}^{\infty} \|\mathbf{B}_l\|_\infty = O(1)$; (iii) $\omega_0^{\varepsilon} = \max_{j\in[p]} \int_{\mathcal{U}} \Sigma^{\varepsilon}_{0,jj}(u,u)\,du = O(1)$, where $\Sigma^{\varepsilon}_{0,jj}(u,u) = \mathrm{Cov}\{\varepsilon_{tj}(u), \varepsilon_{tj}(u)\}$.

The multivariate functional linear process can be seen as the generalization of functional linear process (Bosq, 2000) to the multivariate setting and also the extension of multivariate linear process (Hamilton, 1994) to the functional domain. Condition 2(ii) ensures the stationarity of $\{\mathbf{W}_t(\cdot)\}_{t\in\mathbb{Z}}$ and, together with Condition 2(iii), implies that $\omega_0^W = \max_{j\in[p]} \int_{\mathcal{U}} \Sigma^W_{0,jj}(u,u)\,du = O(1)$ (see Lemma 5 in Appendix B), which is essential in deriving non-asymptotic results. The sub-Gaussian condition is imposed on the functional process to facilitate the use of Hanson–Wright-type inequality in our non-asymptotic analysis. We believe that a Nagaev-type concentration bound can be established to accommodate functional linear process with functional errors under a weaker finite polynomial moments condition. It is also interesting to develop non-asymptotic results for more general non-Gaussian functional time series under other commonly adopted dependence framework.

Condition 3. (i) For each $j \in [p]$, $\lambda_{j1} > \lambda_{j2} > \cdots > 0$, and there exist some constants $c_0 > 0$ and $\alpha > 1$ such that $\lambda_{jl} - \lambda_{j(l+1)} \ge c_0 l^{-\alpha-1}$ for any $l \ge 1$; (ii) For each $j \in [p]$, the linear space spanned by $\{\nu_{jl}(\cdot)\}_{l=1}^{\infty}$ (i.e. eigenfunctions of $\Sigma^X_{0,jj}$) is the same as that spanned by $\{\psi_{jl}(\cdot)\}_{l=1}^{\infty}$.

Condition 3(i) controls the lower bound of eigengaps with larger values of α yielding tighter gaps between adjacent eigenvalues. See similar conditions in Hall and Horowitz (2007) and Kong et al. (2016). To simplify notation, we assume the same α across j, but this condition can be relaxed by allowing α to depend on j and our theoretical results can be generalized accordingly.

We next establish the deviation bounds on estimated eigenpairs, $\{\hat\lambda_{jl}, \hat\psi_{jl}(\cdot)\}$, and the sample autocovariance among estimated basis coefficients, $\{\hat\sigma^{(h)}_{jklm}\}$, in elementwise ℓ_∞-norm.

Theorem 1. *Let Conditions 1–3 hold, and d be a positive integer possibly depending on (n,p). For $n \gtrsim \log p$, there exist some positive constants c_1 and c_2 independent of (n,p,d) such that*

$$\max_{j\in[p],l\in[d]} \left\{ |\hat\lambda_{jl} - \lambda_{jl}| + \left\| \frac{\hat\psi_{jl} - \psi_{jl}}{l^{\alpha+1}} \right\| \right\} \lesssim \mathcal{M}_1^W \sqrt{\frac{\log p}{n}} \tag{11}$$

holds with probability greater than $1 - c_1 p^{-c_2}$, where $\mathcal{M}_1^W$ is defined in (10).

Theorem 2. *Let conditions in Theorem 1 hold and $h \ge 1$ be fixed. For $n \gtrsim d^{2\alpha+2}(\mathcal{M}_1^W)^2 \log p$, there exist some positive constants c_3 and c_4 independent of (n,p,d) such that*

$$\max_{j,k\in[p],l,m\in[d]} \frac{|\hat\sigma^{(h)}_{jklm} - \sigma^{(h)}_{jklm}|}{(l\vee m)^{\alpha+1}} \lesssim \mathcal{M}_1^W \sqrt{\frac{\log p}{n}} \tag{12}$$

holds with probability greater than $1 - c_3 p^{-c_4}$, where $\mathcal{M}_1^W$ is defined in (10).

Remark 1. (i) The parameter d in Theorems 1 and 2 can be understood as the truncated dimension of infinite-dimensional functional objects under the expansion in (4). In general, d can depend on j, say d_j, then the maximums in (11) and (12) are taken over $j, k \in [p]$, $l \in [d_j]$, $m \in [d_k]$ and the corresponding right-sides remain the same.

(ii) Compared with the normalized deviation bounds under FPCA framework established in Guo and Qiao (2023), we obtain slower rates in (11) and (12) for decaying eigenvalues. Note that $\{\nu_{jl}(\cdot)\}_{l=1}^{\infty}$ provides the unique basis with respect to which $X_{tj}(\cdot)$ can be expressed as Karhunen–Loève expansion with uncorrelated coefficients. It gives the most rapidly convergent representation of $X_{tj}(\cdot)$ in the L_2 sense. By comparison, the expansion of $X_{tj}(\cdot)$ through $\{\psi_{jl}(\cdot)\}_{l=1}^{\infty}$ in (4) results in a suboptimal convergent representation with correlated coefficients. From a theoretical viewpoint, whether the rates in (11) and (12) are minimax optimal is of interest and requires further investigation.

3. Block RMD estimation framework

Resulting from Step 1, the estimation of sparse function-valued parameters is transformed to the block sparse estimation of parameter vectors/matrices in Step 2. To identify these parameters, we choose $\{\hat{\boldsymbol{\eta}}_{(t-h)k} : h \in [L], k \in [p]\}$ as the vector-valued instrumental variables and construct autocovariance-based moment equations, which is illustrated using an example of SFLR in Section 3.1. We then formulate a general block RMD estimation method in Section 3.2 and study its theoretical properties in Section 3.3.

3.1. An illustrative example

We illustrate via the high-dimensional SFLR:

$$Y_t = \sum_{j=1}^{p} \int_{\mathcal{U}} X_{tj}(u)\beta_{0j}(u)\,\mathrm{d}u + \varepsilon_t\,, \quad t \in [n]\,, \tag{13}$$

where $\{X_{tj}(\cdot)\}_{t\in[n],j\in[p]}$ satisfy model (1), $\{\varepsilon_t\}_{t\in[n]}$ are i.i.d. and mean-zero random errors, and $\{X_{tj}(\cdot)\}$ and $\{\varepsilon_t\}$ are independent. Given observations $\{(\mathbf{W}_t(\cdot), Y_t)\}_{t\in[n]}$, our goal is to estimate p functional coefficients $\boldsymbol{\beta}_0(\cdot) = \{\beta_{01}(\cdot), \ldots, \beta_{0p}(\cdot)\}^\top$. To guarantee a feasible solution under high-dimensional scaling, we assume that $\boldsymbol{\beta}_0(\cdot)$ is functional s-sparse, i.e. s components in $\boldsymbol{\beta}_0(\cdot)$ are nonzero with $s \ll p$.

Resulting from the truncated expansion of $X_{tj}(\cdot)$ via (4) in Step 1, (13) can be rewritten as

$$Y_t = \sum_{j=1}^{p} \boldsymbol{\eta}_{tj}^\top \mathbf{b}_{0j} + r_t + \varepsilon_t\,,$$

where $\mathbf{b}_{0j} = \int_{\mathcal{U}} \boldsymbol{\psi}_j(u)\beta_{0j}(u)\,\mathrm{d}u \in \mathbb{R}^{d_j}$ and $r_t = \sum_{j=1}^{p}\sum_{l=d_j+1}^{\infty} \eta_{tjl}\langle\psi_{jl}, \beta_{0j}\rangle$ is the truncation error. Given some prescribed positive integer L, in Step 2, we choose $\{\boldsymbol{\eta}_{(t-h)k} : h \in [L], k \in [p]\}$ as the vector-valued instrumental variables. Then $\mathbf{b}_0 = (\mathbf{b}_{01}^\top, \ldots, \mathbf{b}_{0p}^\top)^\top \in \mathbb{R}^{\sum_{j=1}^{p} d_j}$ can be identified by the following moment equations:

$$\mathbb{E}\{\boldsymbol{\eta}_{(t-h)k}\varepsilon_t\} = \mathbf{g}_{hk}(\mathbf{b}_0) + \mathbf{R}_{hk} = \mathbf{0}\,, \quad k \in [p]\,,\ h \in [L]\,, \tag{14}$$

where $\mathbf{g}_{hk}(\mathbf{b}_0) = \mathbb{E}\{\boldsymbol{\eta}_{(t-h)k}Y_t\} - \sum_{j=1}^{p}\mathbb{E}\{\boldsymbol{\eta}_{(t-h)k}\boldsymbol{\eta}_{tj}^\top\mathbf{b}_{0j}\}$ and the bias term $\mathbf{R}_{hk} = -\mathbb{E}\{\boldsymbol{\eta}_{(t-h)k}r_t\}$.

With $\{\hat{\boldsymbol{\eta}}_{tj}\}_{t\in[n],j\in[p]}$ and $\{\hat{\boldsymbol{\psi}}_j(\cdot)\}_{j\in[p]}$ obtained in Step 1, for any $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_p^\top)^\top \in \mathbb{R}^{\sum_{j=1}^{p} d_j}$, we define

$$\hat{\mathbf{g}}_{hk}(\mathbf{b}) = \frac{1}{n-h}\sum_{t=h+1}^{n} \hat{\boldsymbol{\eta}}_{(t-h)k}Y_t - \frac{1}{n-h}\sum_{t=h+1}^{n}\sum_{j=1}^{p} \hat{\boldsymbol{\eta}}_{(t-h)k}\hat{\boldsymbol{\eta}}_{tj}^\top\mathbf{b}_j\,, \quad k \in [p]\,,\ h \in [L]\,, \tag{15}$$

which provides the empirical version of $\mathbf{g}_{hk}(\mathbf{b}) = \mathbb{E}\{\boldsymbol{\eta}_{(t-h)k}Y_t\} - \sum_{j=1}^{p}\mathbb{E}\{\boldsymbol{\eta}_{(t-h)k}\boldsymbol{\eta}_{tj}^\top\mathbf{b}_j\}$. Applying the block RMD estimation introduced in Section 3.2 below results in a block sparse estimator $\hat{\mathbf{b}} = (\hat{\mathbf{b}}_1^\top, \ldots, \hat{\mathbf{b}}_p^\top)^\top$.

3.2. A general estimation procedure

In this section, we present the proposed Step 2 in a general block RMD estimation framework. Note that Step 2 considers the block sparse estimation of some matrix-valued parameters, $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{01}^\top, \ldots, \boldsymbol{\theta}_{0p}^\top)^\top \in \mathbb{R}^{\sum_{j=1}^{p} d_j\times\tilde{d}}$ with each $\boldsymbol{\theta}_{0j} \in \mathbb{R}^{d_j\times\tilde{d}}$. For SFLR with a scalar response, $\tilde{d} = 1$. Given some prescribed positive integer L and $q = pL$ target moment functions $\boldsymbol{\theta} \mapsto \mathbf{g}_i(\boldsymbol{\theta})$ mapping $\boldsymbol{\theta} \in \mathbb{R}^{\sum_{j=1}^{p} d_j\times\tilde{d}}$ to $\mathbf{g}_i(\boldsymbol{\theta}) \in \mathbb{R}^{d_k\times\tilde{d}}$ with $i = (h-1)p + k$ and $k \in [p]$ for $h \in [L]$, where both p and q are large, we assume that $\boldsymbol{\theta}_0$ can be identified by the following moment equations:

$$\mathbf{g}_i(\boldsymbol{\theta}_0) + \mathbf{R}_i = \mathbf{0}\,, \quad i \in [q]\,, \tag{16}$$

where $\mathbf{R}_i$'s are formed by autocovariance-based truncation errors due to finite approximations in Step 1. We are interested in estimating the block sparse $\boldsymbol{\theta}_0$ based on empirical mappings $\boldsymbol{\theta} \mapsto \hat{\mathbf{g}}_i(\boldsymbol{\theta})$ of $\boldsymbol{\theta} \mapsto \mathbf{g}_i(\boldsymbol{\theta})$ for $i \in [q]$. See Sections 4 and 3.1 for detailed expressions of $\mathbf{g}_i(\cdot)$ and $\hat{\mathbf{g}}_i(\cdot)$ in some exemplified models.

It follows from (16) that

$$\hat{\mathbf{g}}_i(\boldsymbol{\theta}_0) \approx \mathbf{0}\,, \quad i \in [q]\,. \tag{17}$$

Based on (17), we define the block RMD estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1^\top, \ldots, \hat{\boldsymbol{\theta}}_p^\top)^\top \in \mathbb{R}^{\sum_{j=1}^{p} d_j\times\tilde{d}}$ as a solution to the following convex optimization problem:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{j=1}^{p} \|\boldsymbol{\theta}_j\|_{\mathrm{F}} \ \text{ subject to } \ \max_{i\in[q]} \|\hat{\mathbf{g}}_i(\boldsymbol{\theta})\|_{\mathrm{F}} \le \gamma_n\,, \tag{18}$$

where $\gamma_n \ge 0$ is a regularization parameter. The group information is encoded in the objective function, which forces the elements of $\hat{\boldsymbol{\theta}}_j$ to either all be zero or nonzero, thus producing the block sparsity in $\hat{\boldsymbol{\theta}}$. It is worth noting that, without the bias terms $\mathbf{R}_i$'s in (16), our proposed block RMD estimation framework can be seen as a blockwise generalization of the RMD estimation (Belloni et al., 2018) by replacing $|\cdot|$ by $\|\cdot\|_{\mathrm{F}}$. To solve the large-scale convex optimization problem in (18), we use the R package `CVXR` (Fu et al., 2020), which is easy to implement and converges fast. In Sections 4.1–4.3, we will illustrate our proposed autocovariance-based block RMD estimation framework using examples of SFLR, FFLR and VFAR, respectively.

3.3. Theoretical properties

For a block matrix $\mathbf{B} = (\mathbf{B}_{ij})_{i\in[N_1],j\in[N_2]} \in \mathbb{R}^{N_1 m_1 \times N_2 m_2}$ with the (i,j)th block $\mathbf{B}_{ij} \in \mathbb{R}^{m_1\times m_2}$, we write $\|\mathbf{B}\|_{\max}^{(m_1,m_2)} = \max_{i\in[N_1],j\in[N_2]} \|\mathbf{B}_{ij}\|_{\mathrm{F}}$, and $\|\mathbf{B}\|_1^{(m_1,m_2)} = \sum_{i=1}^{N_1} \|\mathbf{B}_i\|_{\mathrm{F}}$ when $N_2 = 1$. To simplify notation in this section and theoretical analysis in Section 4, we assume the same truncated dimension $d_j = d$ across $j \in [p]$, but our theoretical results can be extended naturally to the more general setting where d_j's are different.

Let $\mathbf{g}(\boldsymbol{\theta}) = \{\mathbf{g}_1(\boldsymbol{\theta})^\top, \ldots, \mathbf{g}_q(\boldsymbol{\theta})^\top\}^\top$ and $\mathbf{R} = (\mathbf{R}_1^\top, \ldots, \mathbf{R}_q^\top)^\top \in \mathbb{R}^{qd\times\tilde{d}}$. We focus on the case of which the moment function $\boldsymbol{\theta} \mapsto \mathbf{g}(\boldsymbol{\theta})$ mapping from $\mathbb{R}^{pd\times\tilde{d}}$ to $\mathbb{R}^{qd\times\tilde{d}}$ is linear with respect to $\boldsymbol{\theta}$ in the form of $\mathbf{g}(\boldsymbol{\theta}) = \mathbf{G}\boldsymbol{\theta} + \mathbf{g}(\mathbf{0})$ for some $\mathbf{G} \in \mathbb{R}^{qd\times pd}$. This together with (16) implies that

$$\mathbf{G}\boldsymbol{\theta}_0 + \mathbf{g}(\mathbf{0}) + \mathbf{R} = \mathbf{0}\,, \tag{19}$$

the form of which can be easily verified for, e.g., SFLR, FFLR and VFAR models considered in Section 4. Now we reformulate the optimization task in (18) as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1^{(d,\tilde{d})} \ \text{ subject to } \ \|\hat{\mathbf{g}}(\boldsymbol{\theta})\|_{\max}^{(d,\tilde{d})} \le \gamma_n\,, \tag{20}$$

where $\hat{\mathbf{g}}(\boldsymbol{\theta}) = \widehat{\mathbf{G}}\boldsymbol{\theta} + \hat{\mathbf{g}}(\mathbf{0})$ is the empirical version of $\mathbf{g}(\boldsymbol{\theta})$. It is worth noting that $\boldsymbol{\theta}_0$ is block s-sparse with support $S = \{j \in [p] : \|\boldsymbol{\theta}_{0j}\|_{\mathrm{F}} \neq 0\}$ and its cardinality $s = |S|$.

Before presenting properties of the block RMD estimator $\hat{\boldsymbol{\theta}}$, we impose some high-level regularity conditions.

Condition 4. (i) There exist $\epsilon_{n1}, \delta_{n1} > 0$ such that $\|\widehat{\mathbf{G}} - \mathbf{G}\|_{\max}^{(d,d)} \vee \|\hat{\mathbf{g}}(\mathbf{0}) - \mathbf{g}(\mathbf{0})\|_{\max}^{(d,\tilde{d})} \le \epsilon_{n1}$ with probability at least $1 - \delta_{n1}$; (ii) There exists $\epsilon_2 > 0$ such that $\|\mathbf{R}\|_{\max}^{(d,\tilde{d})} \le \epsilon_2$; (iii) There exists $\delta_{n2} > 0$ such that $\|\hat{\mathbf{g}}(\boldsymbol{\theta}_0)\|_{\max}^{(d,\tilde{d})} \le \gamma_n$ with probability at least $1 - \delta_{n2}$.

Conditions 4(i) and 4(ii) together ensure that the empirical moment functions are nicely concentrated around the target moment functions. Using our derived non-asymptotic results in Section 2.2, we can easily specify the concentration bounds in Condition 4(i) for SFLR, FFLR and VFAR. With further imposed smoothness conditions on coefficient functions, Condition 4(ii) can also be verified. Condition 4(iii) indicates that $\boldsymbol{\theta}_0$ is feasible in the optimization problem (20) with high probability, in which case a solution $\hat{\boldsymbol{\theta}}$ of (20) exists and satisfies $\|\hat{\boldsymbol{\theta}}\|_1^{(d,\tilde{d})} \le \|\boldsymbol{\theta}_0\|_1^{(d,\tilde{d})}$. The non-block version of such property typically plays a crucial role to tackle high-dimensional models in the literature.

Let $\boldsymbol{\delta} = \boldsymbol{\theta} - \boldsymbol{\theta}_0$. We define a block ℓ_1-sensitivity coefficient

$$\kappa(\boldsymbol{\theta}_0) = \inf_{T:|T|\le s} \inf_{\boldsymbol{\delta}\in C_T:\, \|\boldsymbol{\delta}\|_1^{(d,\tilde{d})}>0} \frac{\|\mathbf{G}\boldsymbol{\delta}\|_{\max}^{(d,\tilde{d})}}{\|\boldsymbol{\delta}\|_1^{(d,\tilde{d})}}\,, \tag{21}$$

where $C_T = \{\boldsymbol{\delta} \in \mathbb{R}^{pd\times\tilde{d}} : \|\boldsymbol{\delta}_{T^{\mathrm{c}}}\|_1^{(d,\tilde{d})} \le \|\boldsymbol{\delta}_T\|_1^{(d,\tilde{d})}\}$ for $T \subset [p]$. Provided that $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \in C_S$ under Condition 4(iii) as justified in Lemma 1 in Appendix B, the lower bound of $\kappa(\boldsymbol{\theta}_0)$ is useful to establish the error bound for $\|\hat{\boldsymbol{\delta}}\|_1^{(d,\tilde{d})}$. See also Gautier and Rose (2019) for non-block ℓ_q-sensitivity quantities to handle high-dimensional instruments. We then need Condition 5 below to determine such lower bound. Note that $\mathbf{G}$ can be divided into $q \times p$ blocks of the size $d \times d$. Let $\mathbf{G}_{J,M}$ be the submatrix of $\mathbf{G}$ consisting of all the (j,k)-blocks with $j \in J \subset [q]$ and $k \in M \subset [p]$. For an integer $m \ge s$, let

$$\sigma_{\min}(m,\mathbf{G}) = \min_{|M|\le m} \max_{|J|\le m} \sigma_{\min}(\mathbf{G}_{J,M}) \ \text{ and } \ \sigma_{\max}(m,\mathbf{G}) = \max_{|M|\le m} \max_{|J|\le m} \sigma_{\max}(\mathbf{G}_{J,M})\,,$$

where $\sigma_{\min}(\mathbf{G}_{J,M})$ and $\sigma_{\max}(\mathbf{G}_{J,M})$ are the smallest and largest singular values of $\mathbf{G}_{J,M}$.

Condition 5. There exist universal constants $c_5 > 0$ and $\mu > 0$ such that $\sigma_{\max}(m,\mathbf{G}) \ge c_5$ and $\sigma_{\min}(m,\mathbf{G})/\sigma_{\max}(m,\mathbf{G}) \ge \mu$ for $m = 16s/\mu^2$.

In Condition 5, the quantity μ serves as a key factor to determine the lower bound of $\kappa(\boldsymbol{\theta}_0)$, which is justified in Lemma 4 in Appendix B. When μ is bounded away from zero, we have a strongly-identified model. When $\mu \to 0$, it corresponds to the scenario with weak instruments. See also Belloni et al. (2018) for similar conditions.

Theorem 3. *Let Conditions 4–5 hold. If $\|\boldsymbol{\theta}_0\|_1^{(d,\tilde{d})} \le K$ for some $K > 0$ and the regularization parameter $\gamma_n \lesssim (K+1)\epsilon_{n1} + \epsilon_2$, then with probability at least $1 - (\delta_{n1} + \delta_{n2})$, the block RMD estimator $\hat{\boldsymbol{\theta}}$ satisfies*

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1^{(d,\tilde{d})} \lesssim s\mu^{-2}\{(K+1)\epsilon_{n1} + \epsilon_2\}\,. \tag{22}$$

Remark 2. (i) The error bound in (22) has the familiar variance-bias tradeoff as commonly considered in nonparametric statistics, suggesting us to carefully select the truncated dimension d so as to balance the variance and bias terms for the optimal estimation.

(ii) With commonly imposed smoothness conditions on functional coefficients, it is easy to verify that $K \vee \epsilon_2 = o(s)$ for SFLR, FFLR and VFAR in Section 4.

(iii) For three examples we consider, $\mathbf{G}$ is formed by $\{\sigma^{(h)}_{jklm} : j,k \in [p], l,m \in [d], h \in [L]\}$ with the components $\sigma^{(h)}_{jklm}$ satisfying $|\sigma^{(h)}_{jklm}| \le [\mathbb{E}\{\eta^2_{(t-h)jl}\}]^{1/2}\{\mathbb{E}(\eta^2_{tkm})\}^{1/2} = \lambda^{1/2}_{jl}\lambda^{1/2}_{km} \to 0$ as $l,m \to \infty$. Consider a general cross-covariance matrix $\mathbf{G} = \mathbb{E}(\mathbf{x}\mathbf{y}^\top) \in \mathbb{R}^{qd\times pd}$ with entries decaying to zero as $d \to \infty$, where $\mathbf{x} = (x_1,\ldots,x_{qd})^\top$ with $\mathbb{E}(\mathbf{x}) = \mathbf{0}$ and $\mathbf{y} = (y_1,\ldots,y_{pd})^\top$ with $\mathbb{E}(\mathbf{y}) = \mathbf{0}$, it is more sensible to impose Condition 5 on its normalized version $\widetilde{\mathbf{G}} = \mathbf{D}_x\mathbf{G}\mathbf{D}_y$ instead of $\mathbf{G}$ itself, where $\mathbf{D}_x = \mathrm{diag}[\{\mathrm{Var}(x_1)\}^{-1/2},\ldots,\{\mathrm{Var}(x_{qd})\}^{-1/2}]$ and $\mathbf{D}_y = \mathrm{diag}[\{\mathrm{Var}(y_1)\}^{-1/2},\ldots,\{\mathrm{Var}(y_{pd})\}^{-1/2}]$. For three exemplified models, $\mathbf{D}_x$ and $\mathbf{D}_y$ are formed by $\{\lambda^{-1/2}_{jl} : j \in [p], l \in [d]\}$.

Remark 2(iii) motivates us to present the following proposition that will be used in the theoretical analysis of associated estimators for SFLR, FFLR and VFAR in Section 4.

Proposition 1. *Suppose that all conditions in Theorem 3 hold except that Condition 5 holds for $\widetilde{\mathbf{G}}$, then with probability at least $1-(\delta_{n1}+\delta_{n2})$, the block RMD estimator $\hat{\boldsymbol{\theta}}$ satisfies*

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_1^{(d,\tilde{d})} \lesssim s\mu^{-2}\|\mathbf{D}_x\|_{\max}\|\mathbf{D}_y\|_{\max}\{(K+1)\epsilon_{n1}+\epsilon_2\}\,. \tag{23}$$

4. Applications

In this section, we illustrate the proposed estimation procedures with the three concrete models, namely SFLR, FFLR and VFAR.

4.1. High-dimensional SFLR

Consider the high-dimensional SFLR in (13), we first perform autocovariance-based dimension reduction on $\{W_{tj}(\cdot)\}_{t\in[n]}$ for each $j \in [p]$. According to Section 3.1 and following the optimization framework in (18), we then develop the block RMD estimator $\hat{\mathbf{b}}$ as a solution to the constrained optimization problem:

$$\hat{\mathbf{b}} = \arg\min_{\mathbf{b}} \sum_{j=1}^{p} \|\mathbf{b}_j\|_2 \;\text{ subject to }\; \max_{k\in[p],h\in[L]} \|\hat{\mathbf{g}}_{hk}(\mathbf{b})\|_2 \le \gamma_n\,,$$

where $\gamma_n \ge 0$ is a regularization parameter and $\hat{\mathbf{g}}_{hk}(\mathbf{b})$ is defined in (15). Given that the recovery of functional sparsity in $\boldsymbol{\beta}_0(\cdot)$ is equivalent to estimating the block sparsity in $\mathbf{b}_0$, in Step 3, we estimate functional sparse coefficients by

$$\hat{\beta}_j(\cdot) = \hat{\boldsymbol{\psi}}_j(\cdot)^\top\hat{\mathbf{b}}_j\,,\quad j \in [p]\,. \tag{24}$$

We next present the convergence analysis of $\{\hat{\beta}_j(\cdot)\}_{j\in[p]}$. To simplify the notation, we assume the same truncated dimension $d_j = d$ across $j \in [p]$. We rewrite (14) in the form of (19), where $\mathbf{g} = (\mathbf{g}_{11}^\top,\ldots,\mathbf{g}_{1p}^\top,\ldots,\mathbf{g}_{L1}^\top,\ldots,\mathbf{g}_{Lp}^\top)^\top$, $\mathbf{R} = (\mathbf{R}_{11}^\top,\ldots,\mathbf{R}_{1p}^\top,\ldots,\mathbf{R}_{L1}^\top,\ldots,\mathbf{R}_{Lp}^\top)^\top$ and $\mathbf{G} = (\mathbf{G}_{ij}) \in \mathbb{R}^{pLd\times pd}$ whose (i,j)th block is $\mathbf{G}_{ij} = \mathbb{E}\{\boldsymbol{\eta}_{(t-h)k}\boldsymbol{\eta}_{tj}^\top\} \in \mathbb{R}^{d\times d}$ with $i = (h-1)p+k$ and $k \in [p]$ for $h \in [L]$. Applying Theorem 2 and Proposition 3 in Appendix A on $\widehat{\mathbf{G}}$ and $\hat{\mathbf{g}}(\mathbf{0})$, respectively, we can verify Condition 4(i) with the choice of $\epsilon_{n1} \asymp \mathcal{M}_{W,Y}d^{\alpha+2}(n^{-1}\log p)^{1/2}$, where $\mathcal{M}_{W,Y}$ is defined in the same manner as $\mathcal{M}_{W,Z}$ specified in Proposition 3 with selecting $\mathbf{Z}_t = Y_t$. Before presenting the main theorem, we list the regularity conditions below.

Condition 6. (i) For each $j \in S = \{j \in [p] : \|\beta_{0j}\| \not\equiv 0\}$, $\beta_{0j}(\cdot) = \sum_{l=1}^{\infty} a_{jl}\psi_{jl}(\cdot)$ and there exists some positive constant $\tau > \alpha + 1/2$ such that $|a_{jl}| \lesssim l^{-\tau}$ for $l \ge 1$; (ii) Let $\widetilde{\mathbf{G}} = (\widetilde{\mathbf{G}}_{ij})$ be the normalized version of $\mathbf{G} = (\mathbf{G}_{ij})$ by replacing each $\mathbf{G}_{ij}$ by $\widetilde{\mathbf{G}}_{ij} = \mathbb{E}\{\mathbf{D}_k\boldsymbol{\eta}_{(t-h)k}\boldsymbol{\eta}_{tj}^\top\mathbf{D}_j\}$, $i = (h-1)p+k$, $k \in [p]$ for $h \in [L]$ and $j \in [p]$, where $\mathbf{D}_j = \mathrm{diag}(\lambda_{j1}^{-1/2},\ldots,\lambda_{jd}^{-1/2})$. There exist universal constants $c_6 > 0$ and $\mu > 0$ such that $\sigma_{\max}(m,\widetilde{\mathbf{G}}) \ge c_6$ and $\sigma_{\min}(m,\widetilde{\mathbf{G}})/\sigma_{\max}(m,\widetilde{\mathbf{G}}) \ge \mu$ for $m = 16s/\mu^2$.

Condition 6(i) restricts each component in $\{\beta_{0j}(\cdot) : j \in S\}$ based on its expansion through basis $\{\psi_{jl}(\cdot)\}_{l=1}^{\infty}$. The parameter τ determines the decay rate of basis coefficients and hence controls the level of smoothness with large values yielding smoother functions in $\{\beta_{0j}(\cdot) : j \in S\}$. See similar conditions in Hall and Horowitz (2007) and Kong et al. (2016). Noting that components of $\mathbf{G}$ decay to zero as d grows to infinity, we impose Condition 6(ii) on $\widetilde{\mathbf{G}}$, which can be viewed as the normalized counterpart of Condition 5 for SFLR.

Applying Proposition 1 and Theorem 1 yields the convergence rate of the SFLR estimate $\hat{\boldsymbol{\beta}}(\cdot) = \{\hat{\beta}_1(\cdot),\ldots,\hat{\beta}_p(\cdot)\}^\top$ under functional ℓ_1 norm in the following theorem.

Theorem 4. *Suppose that Conditions 1–3, 6 and 9(ii) in Appendix A hold, and $\{Y_t\}_{t\in[n]}$ is sub-Gaussian linear process. Let the regularization parameter $\gamma_n \asymp s\{d^{\alpha+2}\mathcal{M}_{W,Y}(n^{-1}\log p)^{1/2} + d^{-\tau+1/2}\}$, where $\mathcal{M}_{W,Y}$ is defined in the same manner as $\mathcal{M}_{W,Z}$ specified in Proposition 3 in Appendix A with selecting $\mathbf{Z}_t = Y_t$. Then the estimate $\hat{\boldsymbol{\beta}}(\cdot)$ satisfies*

$$\sum_{j=1}^{p}\|\hat{\beta}_j - \beta_{0j}\| = O_{\mathrm{p}}\left\{\mu^{-2}s^2\left(d^{2\alpha+2}\mathcal{M}_{W,Y}\sqrt{\frac{\log p}{n}} + d^{\alpha-\tau+1/2}\right)\right\}. \tag{25}$$

Remark 3. (i) The rate of convergence in (25) is governed by both dimensionality parameters (n, p, s) and internal parameters $(\mathcal{M}_{W,Y}, d, \alpha, \tau, \mu)$. Typically, the rate is better when τ, μ are large and $\mathcal{M}_{W,Y}, \alpha$ are small. To balance the variance and bias terms in (25) for the optimal estimation, we can choose the optimal truncated dimension $d \asymp (\mathcal{M}_{W,Y}^2 n^{-1} \log p)^{-1/(2\tau+2\alpha+3)}$.

(ii) Note that our convergence analysis relies on (12) rather than the normalized deviation bounds in Guo and Qiao (2023), the rate in (25) is slightly slower than that in Fang et al. (2022) by a multiplicative factor $d^{\alpha/2}$. For univariate functional linear regression, we similarly observe a slower rate for the autocovariance-based generalized methods-of-moments estimator (Chen et al., 2022) compared to the covariance-based least squares estimator (Hall and Horowitz, 2007).

4.2. High-dimensional FFLR

Consider high-dimensional FFLR in the form of

$$Y_t(v) = \sum_{j=1}^{p} \int_{\mathcal{U}} X_{tj}(u)\beta_{0j}(u, v)\,\mathrm{d}u + \varepsilon_t(v)\,, \quad t \in [n]\,,\ v \in \mathcal{V}\,, \tag{26}$$

where $\{\mathbf{X}_t(\cdot)\}_{t\in[n]}$ satisfy model (1) and are independent of i.i.d. mean-zero functional errors $\{\varepsilon_t(\cdot)\}_{t\in[n]}$, and $\{\beta_{0j}(\cdot,\cdot)\}_{j\in[p]}$ are functional coefficients to be estimated. With observed data $\{(\mathbf{W}_t(u), Y_t(v)) : (u, v) \in \mathcal{U} \times \mathcal{V}, t \in [n]\}$, we target to estimate $\boldsymbol{\beta}_0 = \{\beta_{01}(\cdot,\cdot), \ldots, \beta_{0p}(\cdot,\cdot)\}^\top$ under a functional sparsity constraint when p is large. Specifically, we assume $\boldsymbol{\beta}_0$ is functional s-sparse with support $S = \{j \in [p] : \|\beta_{0j}\|_{\mathcal{S}} \neq 0\}$ and cardinality $s = |S| \ll p$.

Provided that each observed $Y_t(\cdot)$ is decomposed into the sum of dynamic and white noise components in (26), we approximate $Y_t(\cdot)$ under the Karhunen–Loève expansion truncated at $\tilde{d}$, i.e. $Y_t(\cdot) \approx \boldsymbol{\zeta}_t^\top \boldsymbol{\phi}(\cdot)$, where $\boldsymbol{\zeta}_t = (\zeta_{t1}, \ldots, \zeta_{t\tilde{d}})^\top$ and $\boldsymbol{\phi}(\cdot) = \{\phi_1(\cdot), \ldots, \phi_{\tilde{d}}(\cdot)\}^\top$. Note that we can relax the independence assumption for $\{\varepsilon_t(\cdot)\}_{t\in[n]}$ and model observed responses via $\widetilde{Y}_t(\cdot) = Y_t(\cdot) + e_t^Y(\cdot)$, where $Y_t(\cdot)$ and $e_t^Y(\cdot)$ correspond to the dynamic signal and white noise elements, respectively. Then $Y_t(\cdot)$ can be approximated under the autocovariance-based expansion in the sense of (4) and our subsequent analysis still follow.

For each $j \in [p]$, we expand $X_{tj}(\cdot)$ according to (4) truncated at d_j. Some specific calculations lead to the representation of (26) as

$$\boldsymbol{\zeta}_t^\top = \sum_{j=1}^{p} \boldsymbol{\eta}_{tj}^\top \mathbf{B}_{0j} + \mathbf{r}_t^\top + \boldsymbol{\varepsilon}_t^\top\,, \tag{27}$$

where $\mathbf{B}_{0j} = \int_{\mathcal{U}\times\mathcal{V}} \boldsymbol{\psi}_j(u)\beta_{0j}(u, v)\boldsymbol{\phi}(v)^\top \mathrm{d}u\mathrm{d}v \in \mathbb{R}^{d_j\times\tilde{d}}$, and $\mathbf{r}_t = (r_{t1}, \ldots, r_{t\tilde{d}})^\top \in \mathbb{R}^{\tilde{d}}$ is the truncation error with each $r_{tm} = \sum_{j=1}^{p}\sum_{l=d_j+1}^{\infty} \eta_{tjl}\langle\langle\psi_{jl}, \beta_{0j}\rangle, \phi_m\rangle$ for $m \in [\tilde{d}]$. Let $\mathbf{B}_0 = (\mathbf{B}_{01}^\top, \ldots, \mathbf{B}_{0p}^\top)^\top \in \mathbb{R}^{\sum_{j=1}^{p} d_j\times\tilde{d}}$. We choose $\{\boldsymbol{\eta}_{(t-h)k} : h \in [L], k \in [p]\}$ as the vector-valued instrumental variables, which are assumed to be uncorrelated with the random error $\boldsymbol{\varepsilon}_t$ in (27). Within the framework of (16), we assume that $\mathbf{B}_0$ is the unique solution to the following moment equations:

$$\mathbf{0} = \mathbb{E}\{\boldsymbol{\eta}_{(t-h)k}\boldsymbol{e}_t^\top\} = \mathbf{g}_{hk}(\mathbf{B}_0) + \mathbf{R}_{hk}\,, \quad h \in [L]\,,\ k \in [p]\,, \tag{28}$$

where $\mathbf{g}_{hk}(\mathbf{B}_0) = \mathbb{E}\{\boldsymbol{\eta}_{(t-h)k}\boldsymbol{\zeta}_t^\top\} - \sum_{j=1}^{p}\mathbb{E}\{\boldsymbol{\eta}_{(t-h)k}\boldsymbol{\eta}_{tj}^\top\mathbf{B}_{0j}\}$ and $\mathbf{R}_{hk} = -\mathbb{E}\{\boldsymbol{\eta}_{(t-h)k}\mathbf{r}_t^\top\}$.

Given the recovery equivalence between functional sparsity in $\boldsymbol{\beta}_0$ and the block sparsity in $\mathbf{B}_0$, we aim to estimate the block sparse matrix $\mathbf{B}_0$ using the empirical versions $\mathbf{B} \mapsto \hat{\mathbf{g}}_{hk}(\mathbf{B})$ for $h \in [L]$ and $k \in [p]$,

$$\hat{\mathbf{g}}_{hk}(\mathbf{B}) = \frac{1}{n-h}\sum_{t=h+1}^{n} \hat{\boldsymbol{\eta}}_{(t-h)k}\hat{\boldsymbol{\xi}}_t^\top - \frac{1}{n-h}\sum_{t=h+1}^{n}\sum_{j=1}^{p} \hat{\boldsymbol{\eta}}_{(t-h)k}\hat{\boldsymbol{\eta}}_{tj}^\top\mathbf{B}_j\,,$$

where $\hat{\boldsymbol{\xi}}_t = (\hat{\xi}_{t1}, \ldots, \hat{\xi}_{t\tilde{d}})^\top$ with $\hat{\xi}_{tm} = \langle Y_t, \hat{\phi}_m\rangle$ for $m \in [\tilde{d}]$ and $\{\hat{\boldsymbol{\eta}}_{tj}\}_{t\in[n],j\in[p]}$ are obtained in Step 1. In Step 2, according to (18), we formulate the block RMD estimator $\widehat{\mathbf{B}}$ by solving the convex optimization problem below:

$$\widehat{\mathbf{B}} = \arg\min_{\mathbf{B}} \sum_{j=1}^{p} \|\mathbf{B}_j\|_{\mathrm{F}} \ \text{ subject to } \max_{k\in[p],h\in[L]} \|\hat{\mathbf{g}}_{hk}(\mathbf{B})\|_{\mathrm{F}} \le \gamma_n\,,$$

where $\gamma_n \ge 0$ is a regularization parameter. In Step 3, we estimate the coefficient functions by

$$\hat{\beta}_j(u, v) = \hat{\boldsymbol{\psi}}_j(u)^\top\widehat{\mathbf{B}}_j\hat{\boldsymbol{\phi}}(v)\,, \quad (u, v) \in \mathcal{U}\times\mathcal{V}\,, j \in [p]\,, \tag{29}$$

where $\{\hat{\boldsymbol{\psi}}_j(u)\}_{j\in[p]}$ and $\hat{\boldsymbol{\phi}}(v) = \{\hat{\phi}_1(v), \ldots, \hat{\phi}_{\tilde{d}}(v)\}^\top$ are obtained in Step 1.

In the following, we investigate the convergence property of $\{\hat{\beta}_j(\cdot,\cdot)\}_{j\in[p]}$ in (29). To simplify the notation, we assume the same truncated dimension $d_j = d$ across $j \in [p]$. We first rewrite (28) in the form of (19) and apply Theorem 2 and Proposition 2 in Appendix A on $\widehat{\mathbf{G}}$ and $\hat{\mathbf{g}}(\mathbf{0})$ to verify Condition 4(i) with the choice of $\epsilon_{n1} \asymp \mathcal{M}_{W,Y} d^{\alpha\vee\tilde{\alpha}+2}(n^{-1}\log p)^{1/2}$, where $\mathcal{M}_{W,Y}$ is specified in Proposition 2. In a similar fashion to α, the parameter $\tilde{\alpha}$ as specified in Condition 10 in

Appendix A determines the tightness of eigengaps of the covariance function of $\{Y_t(\cdot)\}$. We then impose the following smoothness condition on nonzero coefficient functions.

Condition 7. For each $j \in S$, $\beta_{0j}(u,v) = \sum_{l,m=1}^{\infty} a_{jlm}\psi_{jl}(u)\phi_m(v)$ and there exists some positive constant $\tau > \alpha \vee \tilde{\alpha} + 1/2$ such that $|a_{jlm}| \lesssim (l+m)^{-\tau-1/2}$ for $l, m \geq 1$.

We are now ready to present the convergence rate of the FFLR estimate $\hat{\boldsymbol{\beta}}(\cdot,\cdot) = \{\hat{\beta}_1(\cdot,\cdot), \ldots, \hat{\beta}_p(\cdot,\cdot)\}^\top$ under functional ℓ_1 norm in Theorem 5.

Theorem 5. *Suppose that Conditions 1–3, 6(ii), 7 and 9(i), 10 in Appendix A hold, and $\{Y_t(\cdot)\}_{t\in[n]}$ is sub-Gaussian functional linear process. Let $d \asymp \tilde{d}$, and the regularization parameter $\gamma_n \asymp s\{d^{\alpha\vee\tilde{\alpha}+2}\mathcal{M}_{W,Y}(n^{-1}\log p)^{1/2} + d^{-\tau+1/2}\}$, where $\mathcal{M}_{W,Y}$ is specified in Proposition 2 in Appendix A. Then the estimate $\hat{\boldsymbol{\beta}}(\cdot,\cdot)$ satisfies*

$$\sum_{j=1}^{p} \|\hat{\beta}_j - \beta_{0j}\|_{\mathcal{S}} = O_{\mathrm{p}}\left\{\mu^{-2}s^2\left(d^{\alpha+\alpha\vee\tilde{\alpha}+2}\mathcal{M}_{W,Y}\sqrt{\frac{\log p}{n}} + d^{\alpha-\tau+1/2}\right)\right\}. \tag{30}$$

Remark 4. (i) With the same expression of $\mathbf{G}$ for both SFLR and FFLR, Condition 6(ii) is required in both Theorems 4 and 5. Note we can further remove the assumption $d \asymp \tilde{d}$, and establish the general convergence rate as a function of d, $\tilde{d}$ and other parameters.

(ii) The rate for the autocovariance-based estimator in (30) is slightly slower than that for the covariance-based estimator in Fang et al. (2022) by a multiplicative factor $d^{\alpha/2}$.

4.3. High-dimensional VFAR

The high-dimensional VFAR of a fixed lag order H, namely VFAR(H), takes the form of

$$\mathbf{X}_t(v) = \sum_{h'=1}^{H}\int_{\mathcal{U}} \mathbf{A}_0^{(h')}(u,v)\mathbf{X}_{t-h'}(u)\,\mathrm{d}u + \boldsymbol{\varepsilon}_t(v), \quad t = H+1, \ldots, n, \tag{31}$$

where $\{\mathbf{X}_t(\cdot)\}$ satisfy model (1), the errors $\boldsymbol{\varepsilon}_t(\cdot) = \{\varepsilon_{t1}(\cdot), \ldots, \varepsilon_{tp}(\cdot)\}^\top$ are i.i.d. sampled from a p-vector of mean-zero random functions, independent of $\mathbf{X}_{t-1}(\cdot), \mathbf{X}_{t-2}(\cdot), \ldots$, and $\mathbf{A}_0^{(h')} = \{A_{0,jj'}^{(h')}(\cdot,\cdot)\}_{j,j'\in[p]}$ is the unknown functional transition matrix at lag h'. In the special case $H = 1$ with $\mathbf{A}_0 = \mathbf{A}_0^{(1)}$, Theorem 3.1 of Bosq (2000) ensures the stationarity of $\{\mathbf{X}_t(\cdot)\}$ if there exists an integer l_0 such that $\sup_{\|\mathbf{f}\|\leq 1}\|\mathbf{A}_0^{l_0}(\mathbf{f})\| < 1$ for $\mathbf{f} \in \mathbb{H}^p$. According to Guo and Qiao (2023), all VFAR(H) models can be reformulated as a VFAR(1) model and hence it is not hard to adjust the stationarity condition for the general case $H > 1$. To make a feasible fit to (31) under a high-dimensional regime based on observed curves $\{\mathbf{W}_t(\cdot)\}_{t\in[n]}$, we assume $\{\mathbf{A}_0^{(h')}\}_{h'\in[H]}$ is rowwise functional s-sparse with $s = \max_{j\in[p]} s_j \ll p$. To be specific, for the jth row of components in $\{\mathbf{A}_0^{(h')}\}$, we denote the set of nonzero functions by $S_j = \{(j',h') \in [p]\times[H] : \|A_{0,jj'}^{(h')}\|_{\mathcal{S}} \neq 0\}$ and its cardinality by $s_j = |S_j|$ for $j \in [p]$.

For each $j \in [p]$, we approximate $X_{tj}(\cdot)$ based on the expansion in (4) truncated at d_j. With some specific calculations, model (31) can be rowwisely rewritten as

$$\boldsymbol{\eta}_{tj}^\top = \sum_{h'=1}^{H}\sum_{j'=1}^{p} \boldsymbol{\eta}_{(t-h')j'}^\top \boldsymbol{\Omega}_{0,jj'}^{(h')} + \mathbf{r}_{tj}^\top + \boldsymbol{\varepsilon}_{tj}^\top, \quad j \in [p], \tag{32}$$

where $\boldsymbol{\Omega}_{0,jj'}^{(h')} = \int_{\mathcal{U}^2} \boldsymbol{\psi}_{j'}(u)A_{0,jj'}^{(h')}(u,v)\boldsymbol{\psi}_j(v)^\top \mathrm{d}u\mathrm{d}v \in \mathbb{R}^{d_{j'}\times d_j}$ and $\mathbf{r}_{tj} = (r_{tj1}, \ldots, r_{tjd_j})^\top$ is the truncation error with each $r_{tjm} = \sum_{h'=1}^{H}\sum_{j'=1}^{p}\sum_{l=d_{j'}+1}^{\infty}\eta_{(t-h')j'l}\langle\langle\psi_{j'l}, A_{0,jj'}^{(h')}\rangle, \psi_{jm}\rangle$ for $m \in [d_j]$. Let $\boldsymbol{\Omega}_{0j} = \{\boldsymbol{\Omega}_{0,j1}^{(1),\top}, \ldots, \boldsymbol{\Omega}_{0,jp}^{(1),\top}, \ldots, \boldsymbol{\Omega}_{0,j1}^{(H),\top}, \ldots, \boldsymbol{\Omega}_{0,jp}^{(H),\top}\}^\top \in \mathbb{R}^{H\sum_{j'=1}^{p} d_{j'}\times d_j}$. We choose $\{\boldsymbol{\eta}_{(t-H-h)k} : h \in [L], k \in [p]\}$ as the vector-valued instrumental variables, which are assumed to be uncorrelated with the random error $\boldsymbol{\varepsilon}_{tj}$ in (32). Within the framework of (16), we assume that $\boldsymbol{\Omega}_{0j}$ is the unique solution to the following moment equations:

$$\mathbf{0} = \mathbb{E}\{\boldsymbol{\eta}_{(t-H-h)k}\boldsymbol{\varepsilon}_{tj}^\top\} = \mathbf{g}_{j,hk}(\boldsymbol{\Omega}_{0j}) + \mathbf{R}_{j,hk}, \quad h \in [L], k \in [p], \tag{33}$$

where $\mathbf{g}_{j,hk}(\boldsymbol{\Omega}_{0j}) = \mathbb{E}\{\boldsymbol{\eta}_{(t-H-h)k}\boldsymbol{\eta}_{tj}^\top\} - \sum_{h'=1}^{H}\sum_{j'=1}^{p}\mathbb{E}\{\boldsymbol{\eta}_{(t-H-h)k}\boldsymbol{\eta}_{(t-h')j'}^\top\boldsymbol{\Omega}_{0,jj'}^{(h')}\}$ and $\mathbf{R}_{j,hk} = -\mathbb{E}\{\boldsymbol{\eta}_{(t-H-h)k}\mathbf{r}_{tj}^\top\}$.

Given that estimating the functional sparsity in the jth row of $\{\mathbf{A}_0^{(h')}\}_{h'\in[H]}$ is equivalent to estimating the block sparsity in $\boldsymbol{\Omega}_{0j}$ for each j, our goal is to estimate the block sparse matrix $\boldsymbol{\Omega}_{0j}$ using the empirical versions $\boldsymbol{\Omega}_j \mapsto \hat{\mathbf{g}}_{j,hk}(\boldsymbol{\Omega}_j)$ for $h \in [L]$ and $k \in [p]$, where

$$\hat{\mathbf{g}}_{j,hk}(\boldsymbol{\Omega}_j) = \frac{1}{n-H-h}\sum_{t=H+h+1}^{n}\hat{\boldsymbol{\eta}}_{(t-H-h)k}\hat{\boldsymbol{\eta}}_{tj}^\top - \frac{1}{n-H-h}\sum_{t=H+h+1}^{n}\sum_{h'=1}^{H}\sum_{j'=1}^{p}\hat{\boldsymbol{\eta}}_{(t-H-h)k}\hat{\boldsymbol{\eta}}_{(t-h')j'}^\top\boldsymbol{\Omega}_{jj'}^{(h')}$$

and $\{\hat{\boldsymbol{\eta}}_{tj}\}_{t\in[n],j\in[p]}$ are obtained in Step 1. Step 2 follows (18) to formulate the block RMD estimator $\widehat{\boldsymbol{\Omega}}_j$ by solving the following optimization task:

$$\widehat{\boldsymbol{\Omega}}_j = \arg\min_{\boldsymbol{\Omega}_j} \sum_{h'=1}^{H}\sum_{j'=1}^{p} \|\boldsymbol{\Omega}_{jj'}^{(h')}\|_{\mathrm{F}} \ \text{ subject to } \max_{k\in[p],h\in[L]} \|\hat{\mathbf{g}}_{j,hk}(\boldsymbol{\Omega}_j)\|_{\mathrm{F}} \le \gamma_{nj}\,,$$

where $\gamma_{nj} \ge 0$ is a regularization parameter. Step 3 estimates functional transition matrices by

$$\hat{A}_{jj'}^{(h')}(u,v) = \hat{\boldsymbol{\psi}}_{j'}(u)^{\top}\widehat{\boldsymbol{\Omega}}_{jj'}^{(h')}\hat{\boldsymbol{\psi}}_j(v)\,, \quad (u,v)\in\mathcal{U}^2\,,\ j,j'\in[p]\,,\ h'\in[H]\,,$$

where $\{\hat{\boldsymbol{\psi}}_j(\cdot)\}_{j\in[p]}$ are obtained in Step 1.

We next present convergence analysis of $\{\hat{A}_{jj'}^{(h')}(\cdot,\cdot): j,j'\in[p], h'\in[H]\}$. To simplify the notation, we assume the same truncated dimension $d_j = d$ across $j\in[p]$. For each j, we first express (33) as below:

$$\mathbf{g}_j(\boldsymbol{\Omega}_{0j}) + \mathbf{R}_j = \mathbf{G}_j\boldsymbol{\Omega}_{0j} + \mathbf{g}_j(\mathbf{0}) + \mathbf{R}_j = \mathbf{0}\,,$$

where $\mathbf{g}_j = (\mathbf{g}_{j,11}^{\top},\ldots,\mathbf{g}_{j,1p}^{\top},\ldots,\mathbf{g}_{j,L1}^{\top},\ldots,\mathbf{g}_{j,Lp}^{\top})^{\top}$, $\mathbf{R}_j = (\mathbf{R}_{j,11}^{\top},\ldots,\mathbf{R}_{j,1p}^{\top},\ldots,\mathbf{R}_{j,L1}^{\top},\ldots,\mathbf{R}_{j,Lp}^{\top})^{\top}$ and $\mathbf{G}_j = (\mathbf{G}_{j,ii'}) \in \mathbb{R}^{pLd\times pHd}$ whose (i,i')th block is $\mathbf{G}_{j,ii'} = \mathbb{E}\{\boldsymbol{\eta}_{(t-H-h)k}\boldsymbol{\eta}_{(t-h')j'}^{\top}\} \in \mathbb{R}^{d\times d}$ with $i = (h-1)p+k$, $k\in[p]$ for $h\in[L]$ and $i' = (h'-1)p+j'$, $j'\in[p]$ for $h'\in[H]$. Applying Theorem 2 on $\widehat{\mathbf{G}}_j$ and $\hat{\mathbf{g}}_j(\mathbf{0})$, we can verify Condition 4(i) with the choice of $\epsilon_{n1} \asymp \mathcal{M}_1^W d^{\alpha+2}(n^{-1}\log p)^{1/2}$, where $\mathcal{M}_1^W$ is defined in (10). Similar to Condition 6 for SFLR, we then give the following regularity conditions.

Condition 8. (i) For each $j\in[p]$ and $(j',h')\in S_j$, $A_{0,jj'}^{(h')}(u,v) = \sum_{l,m=1}^{\infty} a_{jj'lm}^{(h')}\psi_{j'm}(u)\psi_{jl}(v)$ and there exists some constant $\tau > \alpha + 1/2$ such that $|a_{jj'lm}^{(h')}| \lesssim (l+m)^{-\tau-1/2}$ for $l,m\ge 1$; (ii) For each $j\in[p]$, let $\widetilde{\mathbf{G}}_j = (\widetilde{\mathbf{G}}_{j,ii'})$ be the normalized version of $\mathbf{G}_j = (\mathbf{G}_{j,ii'})$ by replacing each $\mathbf{G}_{j,ii'}$ by $\widetilde{\mathbf{G}}_{j,ii'} = \mathbb{E}\{\mathbf{D}_k\boldsymbol{\eta}_{(t-H-h)k}\boldsymbol{\eta}_{(t-h')j'}^{\top}\mathbf{D}_{j'}\}$ for $i = (h-1)p+k$ and $i' = (h'-1)p+j'$ with $k,j'\in[p]$, $h\in[L]$ and $h'\in[H]$, where $\mathbf{D}_j = \mathrm{diag}(\lambda_{j1}^{-1/2},\ldots,\lambda_{jd}^{-1/2})$. There exist universal constants $\tilde{c}_j > 0$ and $\mu_j > 0$ such that $\sigma_{\max}(m,\widetilde{\mathbf{G}}_j) \ge \tilde{c}_j$ and $\sigma_{\min}(m,\widetilde{\mathbf{G}}_j)/\sigma_{\max}(m,\widetilde{\mathbf{G}}_j) \ge \mu_j$ for $m = 16s_j/\mu_j^2$.

We finally establish the convergence rate of the VFAR estimate $\{\hat{A}_{jj'}^{(h')}(\cdot,\cdot)\}_{j,j'\in[p],h'\in[H]}$ in the sense of functional matrix ℓ_∞ norm as follows.

Theorem 6. *Suppose that Conditions 1–3 and 8 hold. Let the regularization parameters satisfy* $\gamma_{nj} \asymp s_j\{d^{\alpha+2}\mathcal{M}_1^W(n^{-1}\log p)^{1/2} + d^{-\tau+1/2}\}$ *for* $j\in[p]$ *and* $\mu = \min_{j\in[p]}\mu_j$, *where* $\mathcal{M}_1^W$ *is defined in* (10). *Then the estimate* $\{\hat{A}_{jj'}^{(h')}(\cdot,\cdot)\}$ *satisfies*

$$\max_{j\in[p]}\sum_{j'=1}^{p}\sum_{h'=1}^{H}\|\hat{A}_{jj'}^{(h')} - A_{0,jj'}^{(h')}\|_{\mathcal{S}} = O_{\mathrm{p}}\left\{\mu^{-2}s^2\left(d^{2\alpha+2}\mathcal{M}_1^W\sqrt{\frac{\log p}{n}} + d^{\alpha-\tau+1/2}\right)\right\}. \tag{34}$$

Remark 5. Similar to Remarks 3(ii) and 4(ii) for SFLR and FFLR respectively, the rate for $\{\hat{A}_{jj'}^{(h')}(\cdot,\cdot)\}$ in (34) is slightly slower than that for the covariance-based estimator in Guo and Qiao (2023) by the same factor $d^{\alpha/2}$.

5. Empirical studies

5.1. Simulation study

In this section, we conduct a number of simulations to evaluate the finite-sample performance of the proposed autocovariance-based estimators for SFLR, FFLR and VFAR models.

In each simulated scenario, to mimic the infinite-dimensional feature of signal curves, we generate $X_{tj}(u) = \sum_{l=1}^{25}\eta_{tjl}\psi_l(u) = \boldsymbol{\eta}_{tj}^{\top}\boldsymbol{\psi}(u)$ with $\boldsymbol{\eta}_{tj} = (\eta_{tj1},\ldots,\eta_{tj25})^{\top}$ and $\boldsymbol{\psi}(\cdot) = \{\psi_1(\cdot),\ldots,\psi_{25}(\cdot)\}^{\top}$ for $t\in[n]$, $j\in[p]$ and $u\in\mathcal{U} = [0,1]$, where $\{\psi_l(u)\}_{1\le l\le 25}$ is formed by 25-dimensional Fourier basis functions, $1, \sqrt{2}\cos(2\pi lu), \sqrt{2}\sin(2\pi lu)$ for $l = 1,\ldots,12$ and each $\boldsymbol{\eta}_t = (\boldsymbol{\eta}_{t1}^{\top},\ldots,\boldsymbol{\eta}_{tp}^{\top})^{\top} \in \mathbb{R}^{25p}$ is generated from a stationary vector autoregressive (VAR) model, $\boldsymbol{\eta}_t = \boldsymbol{\Omega}\boldsymbol{\eta}_{t-1} + \boldsymbol{\epsilon}_t$, with block transition matrix $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_{jk})_{j,k\in[p]} \in \mathbb{R}^{25p\times 25p}$ and $\boldsymbol{\epsilon}_t = (\boldsymbol{\epsilon}_{t1}^{\top},\ldots,\boldsymbol{\epsilon}_{tp}^{\top})^{\top} \in \mathbb{R}^{25p}$, where the components of each $\boldsymbol{\epsilon}_{tj} = (\epsilon_{tj1},\ldots,\epsilon_{tj25})^{\top}$ are sampled independently according to $\epsilon_{tjl} \sim \mathcal{N}(0, 0.7 - 0.1l)$ for $l = 1,\ldots,5$ and $\mathcal{N}(0, l^{-2})$ for $l = 6,\ldots,25$. Therefore, $\mathbf{X}_t(\cdot)$ follows a VFAR(1) model $\mathbf{X}_t(v) = \int_{\mathcal{U}}\mathbf{A}(u,v)\mathbf{X}_{t-1}(u)\,\mathrm{d}u + \boldsymbol{\varepsilon}_t(v)$, where $\boldsymbol{\varepsilon}_{tj}(v) = \boldsymbol{\psi}(v)^{\top}\boldsymbol{\epsilon}_{tj}$ and autocoefficient functions satisfy $A_{jk}(u,v) = \boldsymbol{\psi}(v)^{\top}\boldsymbol{\Omega}_{jk}\boldsymbol{\psi}(u)$ for $j,k\in[p]$ and $u,v\in\mathcal{U}$. In our simulations, we generate $n = 100, 200, 400$ serially dependent observations of $p = 40, 80$ functional variables. The observed curves are generated from $W_{tj}(u) = X_{tj}(u) + e_{tj}(u)$, where white noise curves $e_{tj}(u) = \sum_{l=1}^{5} z_{tjl}\psi_l(u)$, $\mathbf{z}_{tj} = (z_{tj1},\ldots,z_{tj5})^{\top}$ and $\{\mathbf{z}_{tj}\}_{t\in[n]}$ are sampled independently from multivariate normal distribution with mean zero and covariance matrix $\mathrm{diag}(1, 0.8, 0.3, 1.5, 1.6)$. For each of the three models, the data is generated as follows.

VFAR: We generate block sparse $\boldsymbol{\Omega}$ with 5% or 10% nonzero blocks for $p = 80$ or $p = 40$, respectively. Specifically, for the jth block row, we set the diagonal block $\boldsymbol{\Omega}_{jj} = \text{diag}(0.60, 0.59, 0.58, 0.3, 0.2, 6^{-2}, \ldots, 25^{-2})$ and randomly choose one off-diagonal block being $0.4\boldsymbol{\Omega}_{jj}$ and two off-diagonal blocks being $0.1\boldsymbol{\Omega}_{jj}$. Such block sparse design on $\boldsymbol{\Omega}$ can guarantee the stationarity of the generated VFAR(1) process. It is worth noting that estimating VFAR(1) results in a very high-dimensional task, since, e.g. even under the most 'low-dimensional' setting with $p = 40$, $n = 400$ and truncated dimension $d = 3$, one needs to estimate $(40 \times 3)^2 = 14{,}400$ parameters based on only 400 observations. The p-vector of functional covariates $\{\mathbf{X}_t(\cdot)\}_{t\in[n]}$ for SFLR and FFLR below are generated in the same way as those for VFAR.

SFLR: We generate the scalar responses $\{Y_t\}_{t\in[n]}$ from model (13), where ε_t's are independent $\mathcal{N}(0, 1)$ variables. For each $j \in S = \{1, \ldots, 5\}$, we generate $\beta_j(u) = \sum_{l=1}^{25} b_{jl}\psi_l(u)$ for $u \in \mathcal{U}$, where b_{j1}, b_{j2}, b_{j3} are sampled from the uniform distribution with support $[-1, -0.5] \cup [0.5, 1]$ and $b_{jl} = (-1)^l l^{-2}$ for $l = 4, \ldots, 25$. For $j \in [p] \setminus S$, we let $\beta_j(u) = 0$.

FFLR: We generate the functional responses $\{Y_t(v) : v \in \mathcal{V}\}_{t\in[n]}$ with $\mathcal{V} = [0, 1]$ from model (26), where $\varepsilon_t(v) = \sum_{m=1}^{5} g_{tm}\psi_m(v)$ with g_{tm}'s being independent $\mathcal{N}(0, 1)$ variables. For $j \in S$, we generate $\beta_j(u, v) = \sum_{l,m=1}^{25} b_{jml}\psi_l(u)\psi_m(v)$ for $(u, v) \in \mathcal{U} \times \mathcal{V}$, where components in $\{b_{jlm}\}_{1\le l,m\le 3}$ are sampled from the uniform distribution with support $[-1, -0.5] \cup [0.5, 1]$ and $b_{jlm} = (-1)^{l+m}(l + m)^{-2}$ for l or $m = 4, \ldots, 25$. For $j \in [p] \setminus S$, we let $\beta_j(u, v) = 0$.

Implementing our proposed autocovariance-based learning framework (AUTO) requires choosing L and d_j's. As our simulated results suggest that the estimators are not sensitive to the choice of L, we set $L = 3$ in simulations. To select d_j, we take the standard approach by selecting the largest d_j eigenvalues of $\hat{K}_{jj}$ in (6) such that the cumulative percentage of selected eigenvalues exceeds 90%. To choose the regularization parameter(s) for each model and comparison method, there are several possible methods one could adopt such as AIC, BIC and cross-validation. The AIC and BIC methods require the calculation of the effective degrees of freedom, which leads to a very challenging task given the high-dimensional, functional and dependent nature of the model structure and hence is left for future research. In our simulations, we generate a training sample of size n and a separate validation sample of the same size. Using the training data, we compute a series of estimators with 30 different values of the regularization parameters, i.e. $\{\hat{\mathbf{b}}_j^{(\gamma_n)}\}_{j\in[p]}$ (or $\{\widehat{\mathbf{B}}_j^{(\gamma_n)}\}_{j\in[p]}$) as a function of γ_n for SFLR (or FFLR) and $\{\widehat{\boldsymbol{\Omega}}_{jk}^{(\gamma_{nj})}\}_{k\in[p]}$ as a function of γ_{nj} for VFAR, calculate the squared error between observed and fitted values on the validation set, i.e. $\sum_{t=1}^{n}\{Y_t - \sum_{j=1}^{p}\hat{\mathbf{b}}_j^{(\gamma_n),\top}\hat{\boldsymbol{\eta}}_{tj}\}^2$ for SFLR, $\sum_{t=1}^{n}\|\hat{\boldsymbol{\zeta}}_t - \sum_{j=1}^{p}\widehat{\mathbf{B}}_j^{(\gamma_n),\top}\hat{\boldsymbol{\eta}}_{tj}\|^2$ for FFLR and $\sum_{t=1}^{n}\|\hat{\boldsymbol{\eta}}_{tj} - \sum_{k=1}^{p}\widehat{\boldsymbol{\Omega}}_{jk}^{(\gamma_{nj}),\top}\hat{\boldsymbol{\eta}}_{(t-1)k}\|^2$ for VFAR, and choose the one with the smallest error.

We compare AUTO with the standard covariance-based estimation framework (COV), which proceeds in the following three steps. The first step performs FPCA on $\{W_{tj}(\cdot)\}_{t\in[n]}$ for each $j \in [p]$, where the truncated dimension was selected in the same way as d_j. Therefore, estimating SFLR and FFLR models are transformed into fitting multiple linear regressions with the univariate response (Kong et al., 2016) and the multivariate response (Fang et al., 2022), respectively and the VFAR estimation is converted to the VAR estimation (Guo and Qiao, 2023). The second step considers minimizing the covariance-based criterion, essentially the least squares with the addition of a group lasso type penalty. Such criterion can be optimized using an efficient block fast iterative shrinkage-thresholding algorithm developed in Guo and Qiao (2023), which converges faster than the commonly adopted block coordinate descent algorithm (Fan et al., 2015). The third step recovers functional sparse estimates using estimated eigenfunctions.

We examine the performance of COV and AUTO for three models in terms of relative estimation errors, i.e. $\|\widehat{\mathbf{A}} - \mathbf{A}\|_{\text{F}}/\|\mathbf{A}\|_{\text{F}}$ for VFAR, $(\sum_{j=1}^{p}\|\hat{\beta}_j - \beta_{0j}\|^2)^{1/2}/(\sum_{j=1}^{p}\|\beta_{0j}\|^2)^{1/2}$ for SFLR and $(\sum_{j=1}^{p}\|\hat{\beta}_j - \beta_{0j}\|_{\mathcal{S}}^2)^{1/2}/(\sum_{j=1}^{p}\|\beta_{0j}\|_{\mathcal{S}}^2)^{1/2}$ for FFLR. We ran each simulation 100 times. Fig. 1 displays boxplots of relative estimation errors for three models. Several conclusions can be drawn from Fig. 1. First, AUTO significantly outperforms COV for three models under all scenarios we consider. Second, as discussed in Section 2.1, AUTO provides consistent estimates, while the consistency of COV estimates is jeopardized by the white noise contamination. This can be demonstrated by our empirical results that AUTO provides more substantially improved estimates over COV as n increases from 100 to 400. Third, the performance of AUTO slightly deteriorates as p increases from 40 to 80, providing empirical evidence to support that the rates in (25), (30) and (34) for SFLR, FFLR and VFAR models, respectively, all depend on the $(\log p)^{1/2}$ term.

5.2. Real data analysis

In this section, we illustrate our developed methodology using a public financial dataset, which was obtained from the WRDS database and consists of high-frequency observations of prices for S&P 100 index and component stocks (list available in Table 2 in Appendix C, we removed several stocks for which the data were not available so that $p = 98$ in our analysis) in year 2017 comprising 251 trading days. We obtain one-minute resolution prices by using the last transaction price in each one-minute interval after removing the outliers, and hence convert the trading period (9:30–16:00) to minutes [0, 390]. We construct cumulative intraday return (CIDR) trajectories (Horváth et al., 2014), in percentage, by $W_{tj}(u_k) = 100[\log\{P_{tj}(u_k)\} - \log\{P_{tj}(u_1)\}]$, where $P_{tj}(u_k)$ $(t \in [n], j \in [p], k \in [N])$ denotes the price of the jth stock at the kth minute after the opening time on the tth trading day. We work with mildly smoothed CIDRs obtained by expanding the data with respect to a 45-dimensional B-spline basis. The CIDR curves always start from zero and have nearly the same shape as the original price curves, but make the stationarity assumption more plausible. We performed the functional KPSS test (Horváth et al., 2014) on CIDR curves for each stock using the R package `fsta` (Shang, 2013). The p-values are all larger than 1%, which indicates that there is no overwhelming evidence against the stationarity.

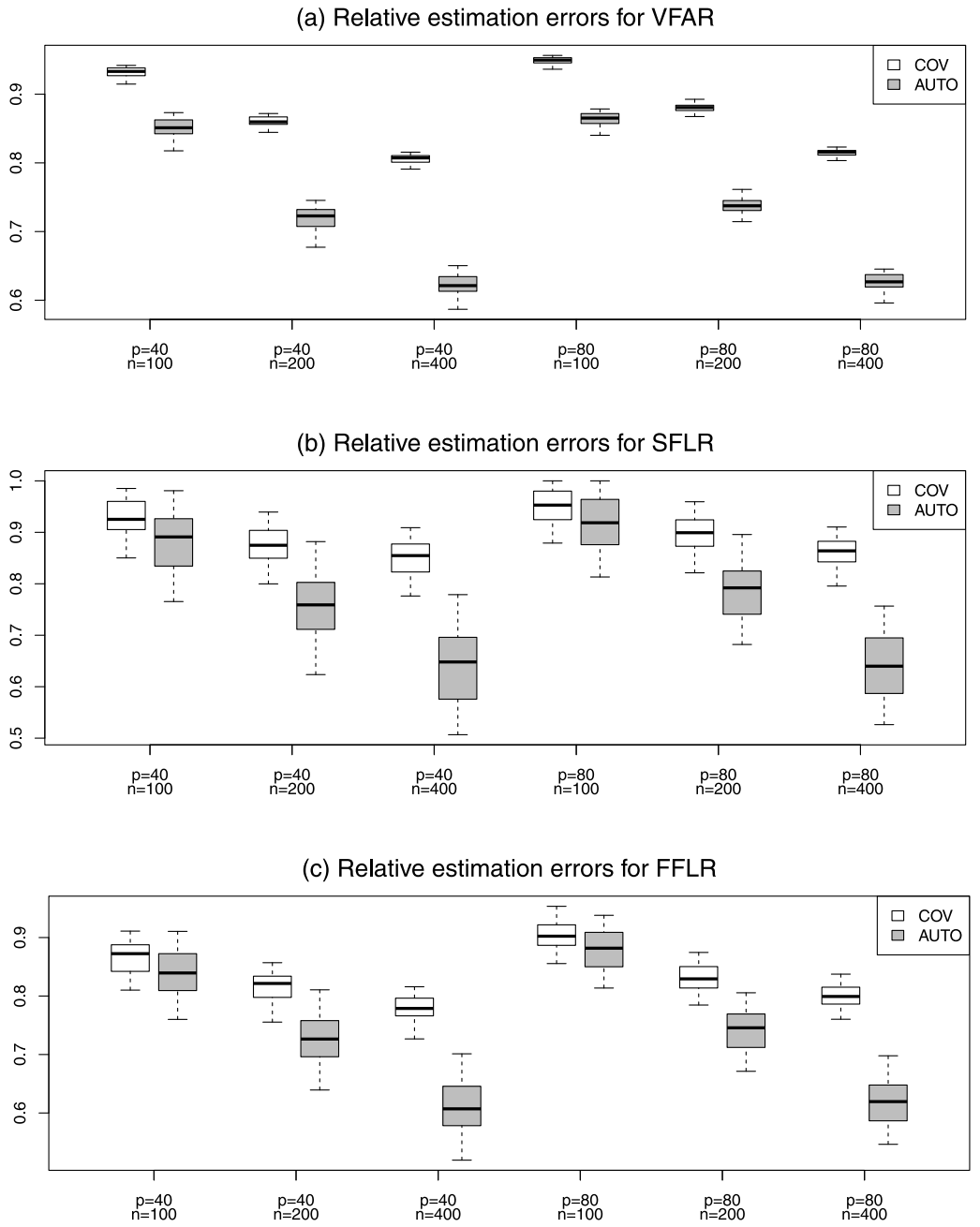

Fig. 1. The boxplots of relative estimation errors for (a) VFAR, (b) SFLR and (c) FFLR.

Our target is to predict the intraday return of the S&P 100 index based on observed CIDR trajectories of component stocks, $W_{tj}(u)$, $u \in \mathcal{U} = [0, N]$ up to time N, where, e.g., $N = 360$ corresponds to 30 min prior to the closing time of the trading day. With this in mind, we construct a sparse SFLR model with erroneous functional covariates as follows

$$Y_t = \sum_{j=1}^{p} \int_{\mathcal{U}} X_{tj}(u)\beta_{0j}(u)\,\mathrm{d}u + \varepsilon_t\,, \quad W_{tj}(u) = X_{tj}(u) + e_{tj}(u)\,, \quad t \in [n]\,,\, j \in [p]\,, \tag{35}$$

where Y_t is the intraday return of the S&P 100 index on the tth trading day, $X_{tj}(\cdot)$ and $e_{tj}(\cdot)$ represent the signal and noise components in $W_{tj}(\cdot)$, respectively. We split the whole dataset into three subsets: training, validation and test sets consisting of the first 171, the subsequent 40 and the last 40 observations, respectively. We apply the validation set approach to select the regularization parameters for AUTO and COV, based on which we estimate sparse functional

Table 1
MSPEs up to different current times, $N = 300, 315, 330, 345, 360, 370$ and 380 min, for AUTO and four competing methods. All entries have been multiplied by 100 for formatting reasons. The lowest MSPE for each value of N is in bold font.

Method	$u \leqslant 300$	$u \leqslant 315$	$u \leqslant 330$	$u \leqslant 345$	$u \leqslant 360$	$u \leqslant 370$	$u \leqslant 380$
AUTO	**5.068**	**4.936**	**4.814**	**4.161**	**3.892**	**3.798**	**3.726**
COV	5.487	5.360	5.222	5.090	4.976	4.927	4.882
AGMM	6.506	6.470	6.454	6.441	6.408	6.385	6.364
CLS	6.859	6.798	6.730	6.655	6.583	6.546	6.507
Mean	8.832	8.832	8.832	8.832	8.832	8.832	8.832

coefficients in (35) and calculate the mean squared prediction errors (MSPEs) on the test set. For comparison, we also implement autocovariance-based generalized method-of-moments (AGMM) (Chen et al., 2022) and covariance-based least squares method (CLS) (Hall and Horowitz, 2007) to fit the univariate version of (35) for each component stock, among which we choose the best models leading to the lowest test MSPEs. Finally, we include the null model using the mean of training responses to predict test responses.

The resulting test MSPEs for different values of N and all comparison approaches are presented in Table 1. We observe a few apparent patterns. First, in all scenarios we consider, AUTO provides the best predictive performance, while the autocovariance-based methods are superior to the covariance-based counterparts. Second, the predictive accuracy for functional regression type of methods improves as N approaches to 390 providing more recent information into the covariates. Third, AUTO and COV significantly outperform AGMM and CLS, while Mean gives the worst results. This indicates that using multiple selected functional covariates from the trading histories indeed improves the prediction results.

Appendix

This appendix contains further non-asymptotic results in Appendix A, all technical proofs in Appendix B and list of S&P 100 stocks in Appendix C.

Appendix A. Further non-asymptotic results

To provide theoretical guarantees for the proposed estimators in Sections 4.1 and 4.2, we present essential non-asymptotic error bounds on the relevant estimated cross-(auto) covariance terms based on the functional cross-spectral stability measure (Fang et al., 2022) between $\{\mathbf{W}_t(\cdot)\}_{t\in\mathbb{Z}}$ and $\tilde{p}$-vector of mean-zero functional time series (or scalar time series) $\{\mathbf{Y}_t(\cdot)\}_{t\in\mathbb{Z}}$ (or $\{\mathbf{Z}_t\}_{t\in\mathbb{Z}}$). Define $\boldsymbol{\Sigma}_h^{W,Y}(u,v) = \mathrm{Cov}\{\mathbf{W}_{t-h}(u), \mathbf{Y}_t(v)\}$ and $\boldsymbol{\Sigma}_h^{W,Z}(u) = \mathrm{Cov}\{\mathbf{W}_{t-h}(u), \mathbf{Z}_t\}$ for $h \in \mathbb{Z}$ and $(u,v) \in \mathcal{U}\times\mathcal{V}$.

Condition 9. (i) For $\{\mathbf{W}_t(\cdot)\}_{t\in\mathbb{Z}}$ and $\{\mathbf{Y}_t(\cdot)\}_{t\in\mathbb{Z}}$, the cross-spectral density function $\mathbf{f}_\theta^{W,Y} = (2\pi)^{-1}\sum_{h\in\mathbb{Z}}\boldsymbol{\Sigma}_h^{W,Y}e^{-ih\theta}$ for $\theta\in[-\pi,\pi]$ exists and the functional cross-spectral stability measure defined in (A.1) is finite, i.e.

$$\mathcal{M}^{W,Y} = 2\pi\cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\,\boldsymbol{\Phi}_1\in\mathbb{H}_0^p,\,\boldsymbol{\Phi}_2\in\mathbb{H}_0^{\tilde{p}}} \frac{|\langle\boldsymbol{\Phi}_1, \mathbf{f}_\theta^{W,Y}(\boldsymbol{\Phi}_2)\rangle|}{\sqrt{\langle\boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_0^W(\boldsymbol{\Phi}_1)\rangle}\sqrt{\langle\boldsymbol{\Phi}_2, \boldsymbol{\Sigma}_0^Y(\boldsymbol{\Phi}_2)\rangle}} < \infty, \tag{A.1}$$

where $\mathbb{H}_0^p = \{\boldsymbol{\Phi}\in\mathbb{H}^p : \langle\boldsymbol{\Phi}, \boldsymbol{\Sigma}_0^W(\boldsymbol{\Phi})\rangle \in (0,\infty)\}$ and $\mathbb{H}_0^{\tilde{p}} = \{\boldsymbol{\Phi}\in\mathbb{H}^{\tilde{p}} : \langle\boldsymbol{\Phi}, \boldsymbol{\Sigma}_0^Y(\boldsymbol{\Phi})\rangle \in (0,\infty)\}$. (ii) For $\{\mathbf{W}_t(\cdot)\}_{t\in\mathbb{Z}}$ and $\{\mathbf{Z}_t\}_{t\in\mathbb{Z}}$, the cross-spectral density function $\mathbf{f}_\theta^{W,Z} = (2\pi)^{-1}\sum_{h\in\mathbb{Z}}\boldsymbol{\Sigma}_h^{W,Z}e^{-ih\theta}$ for $\theta\in[-\pi,\pi]$ exists and the functional cross-spectral stability measure defined in (A.2) is finite, i.e.

$$\mathcal{M}^{W,Z} = 2\pi\cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\,\boldsymbol{\Phi}\in\mathbb{H}_0^p,\mathbf{v}\in\mathbb{R}_0^{\tilde{p}}} \frac{|\langle\boldsymbol{\Phi}, \mathbf{f}_\theta^{W,Z}\mathbf{v}\rangle|}{\sqrt{\langle\boldsymbol{\Phi}, \boldsymbol{\Sigma}_0^X(\boldsymbol{\Phi})\rangle}\sqrt{\mathbf{v}^\top\boldsymbol{\Sigma}_0^Z\mathbf{v}}} < \infty, \tag{A.2}$$

where $\mathbb{R}_0^{\tilde{p}} = \{\boldsymbol{v}\in\mathbb{R}^{\tilde{p}} : \mathbf{v}^\top\boldsymbol{\Sigma}_0^Z\mathbf{v}\in(0,\infty)\}$.

In analogy to (10), we can define the functional cross-spectral stability measure of all k_1-dimensional subsets of $\{\mathbf{W}_t(\cdot)\}$ and k_2-dimensional subsets of $\{\mathbf{Y}_t(\cdot)\}$ (or $\{\mathbf{Z}_t\}$) as $\mathcal{M}_{k_1,k_2}^{W,Y}$ (or $\mathcal{M}_{k_1,k_2}^{W,Z}$). It is easy to verify that $\mathcal{M}_{k_1,k_2}^{W,Y} \leq \mathcal{M}^{W,Y} < \infty$ (or $\mathcal{M}_{k_1,k_2}^{W,Z} \leq \mathcal{M}^{W,Z} < \infty$) for $k_1\in[p]$ and $k_2\in[\tilde{p}]$. For scalar time series $\{\mathbf{Z}_t\}$, the non-functional stability measure degenerates to

$$\mathcal{M}^Z = 2\pi\cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\mathbf{v}\in\mathbb{R}_0^{\tilde{p}}} \frac{\mathbf{v}^\top\mathbf{f}_\theta^Z\mathbf{v}}{\mathbf{v}^\top\boldsymbol{\Sigma}_0^Z\mathbf{v}},$$

which is equivalent to that proposed in Basu and Michailidis (2015). The stability measure of all k-dimensional subsets of $\{\mathbf{Z}_t\}$, i.e. $\mathcal{M}_k^Z$ for $k \in [\tilde{p}]$, can be defined similarly according to (10).

For each $k \in [\tilde{p}]$, we represent $Y_{tk}(\cdot) = \sum_{m=1}^{\infty} \zeta_{tkm}\phi_{km}(\cdot)$ under the Karhunen–Loève expansion, where $\zeta_{tkm} = \langle Y_{tk}, \phi_{km}\rangle$ and $\{(\theta_{km}, \phi_{km})\}_{m\geq 1}$ are the pairs of eigenvalues and eigenfunctions of $\Sigma_{0,kk}^Y$. Let $\{(\hat{\theta}_{km}, \hat{\phi}_{km})\}_{m\geq 1}$ be the estimated eigenpairs of $\widehat{\Sigma}_{0,kk}^Y$ and $\hat{\zeta}_{tkm} = \langle Y_{tk}, \hat{\phi}_{km}\rangle$. We next impose a condition on the eigenvalues $\{\theta_{km}\}_{m\geq 1}$ and then develop the deviation bound in elementwise ℓ_∞-norm on how $\hat{\sigma}_{h,jklm}^{W,Y} = (n-h)^{-1}\sum_{t=h+1}^{n}\hat{\eta}_{(t-h)jl}\hat{\zeta}_{tkm}$ concentrates around $\sigma_{h,jklm}^{W,Y} = \mathrm{Cov}\{\eta_{(t-h)jl}, \zeta_{tkm}\}$, which plays a crucial role in the convergence analysis of the FFLR estimate in Section 4.2.

Condition 10. (i) For each $k \in [\tilde{p}]$, $\theta_{k1} > \theta_{k2} > \cdots > 0$, and there exist some positive constants $\tilde{c}$ and $\tilde{\alpha} > 1$ such that $\theta_{km} - \theta_{k(m+1)} \geq \tilde{c}m^{-\tilde{\alpha}-1}$ for $m \geq 1$; (ii) $\max_{k\in[\tilde{p}]}\sum_{m=1}^{\infty}\theta_{km} = O(1)$.

Proposition 2. *Suppose that Conditions 1–3, 9(i) and 10 hold, $\{\mathbf{Y}_t(\cdot)\}_{t\in[n]}$ is sub-Gaussian functional linear process and h is fixed. Let d and $\tilde{d}$ be positive integers possibly depending on $(n, p, \tilde{p})$ and $\mathcal{M}_{W,Y} = \mathcal{M}_1^W + \mathcal{M}_1^Y + \mathcal{M}_{1,1}^{W,Y}$. For $n \gtrsim (d^{2\alpha+2} \vee \tilde{d}^{2\tilde{\alpha}+2})(\mathcal{M}_{W,Y})^2\log(p\tilde{p})$, there exist some positive constants c_7 and c_8 independent of $(n, p, \tilde{p}, d, \tilde{d})$ such that*

$$\max_{j\in[p],k\in[\tilde{p}],l\in[d],m\in[\tilde{d}]}\frac{|\hat{\sigma}_{h,jklm}^{W,Y} - \sigma_{h,jklm}^{W,Y}|}{l^{\alpha+1}\vee m^{\tilde{\alpha}+1}} \lesssim \mathcal{M}_{W,Y}\sqrt{\frac{\log(p\tilde{p})}{n}} \tag{A.3}$$

holds with probability greater than $1 - c_7(p\tilde{p})^{-c_8}$.

We next consider a mixed process scenario consisting of $\{\mathbf{W}_t(\cdot)\}$ and $\{\mathbf{Z}_t\}$ and establish the deviation bound on how $\hat{\varrho}_{h,jkl}^{X,Z} = (n-h)^{-1}\sum_{t=h+1}^{n}\hat{\eta}_{(t-h)jl}Z_{tk}$ concentrates around $\varrho_{h,jkl}^{X,Z} = \mathrm{Cov}\{\eta_{(t-h)jl}, Z_{tk}\}$, which is essential in deriving the convergence rate of the SFLR estimate in Section 4.1.

Proposition 3. *Suppose that Conditions 1–3 and 9(ii) hold, $\{\mathbf{Z}_t\}_{t\in[n]}$ is sub-Gaussian linear process and h is fixed. Let d be a positive integer possibly depending on $(n, p, \tilde{p})$ and $\mathcal{M}_{W,Z} = \mathcal{M}_1^W + \mathcal{M}_1^Z + \mathcal{M}_{1,1}^{W,Z}$. For $n \gtrsim (\mathcal{M}_{W,Z})^2\log(p\tilde{p})$, there exist some positive constants c_9 and c_{10} independent of $(n, p, \tilde{p}, d)$ such that*

$$\max_{j\in[p],k\in[\tilde{p}],l\in[d]}\frac{|\hat{\varrho}_{h,jkl}^{W,Z} - \varrho_{h,jkl}^{W,Z}|}{l^{\alpha+1}} \lesssim \mathcal{M}_{W,Z}\sqrt{\frac{\log(p\tilde{p})}{n}} \tag{A.4}$$

holds with probability greater than $1 - c_9(p\tilde{p})^{-c_{10}}$.

Appendix B. Technical proofs

Throughout, we use c, $\bar{c}$, $\tilde{c}$, $\check{c}$ and $\dot{c}$ to denote generic positive finite constants that may be different in different uses.

B.1. Auxiliary lemmas

Lemma 1. *Suppose that Condition 4(iii) holds. Then $\|\hat{\boldsymbol{\delta}}_{S^c}\|_1^{(d,\tilde{d})} \leq \|\hat{\boldsymbol{\delta}}_S\|_1^{(d,\tilde{d})}$ with probability at least $1 - \delta_{n2}$.*

Proof. It follows from Condition 4(iii) and $\boldsymbol{\theta}_{0,S^c} = \mathbf{0}$ by definition that with probability at least $1-\delta_{n2}$, $\|\hat{\boldsymbol{\theta}}\|_1^{(d,\tilde{d})} \leq \|\boldsymbol{\theta}_0\|_1^{(d,\tilde{d})} = \|\boldsymbol{\theta}_{0,S}\|_1^{(d,\tilde{d})}$, which implies that $\|\boldsymbol{\theta}_{0,S}\|_1^{(d,\tilde{d})} \geq \|\hat{\boldsymbol{\theta}}_S\|_1^{(d,\tilde{d})} + \|\hat{\boldsymbol{\theta}}_{S^c}\|_1^{(d,\tilde{d})} \geq \|\boldsymbol{\theta}_{0,S}\|_1^{(d,\tilde{d})} - \|\hat{\boldsymbol{\theta}}_S - \boldsymbol{\theta}_{0,S}\|_1^{(d,\tilde{d})} + \|\hat{\boldsymbol{\theta}}_{S^c}\|_1^{(d,\tilde{d})}$. By cancelling $\|\boldsymbol{\theta}_{0,S}\|_1^{(d,\tilde{d})}$ on both sides above, we obtain $\|\hat{\boldsymbol{\theta}}_{S^c} - \boldsymbol{\theta}_{0,S^c}\|_1^{(d,\tilde{d})} \leq \|\hat{\boldsymbol{\theta}}_S - \boldsymbol{\theta}_{0,S}\|_1^{(d,\tilde{d})}$. □

Lemma 2. *For $\mathbf{A} \in \mathbb{R}^{q\times p}$ with $\mathrm{rank}(\mathbf{A}) \leq \min(p, q)$ and $\mathbf{x} \in \mathbb{R}^{p\times d}$, let $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^\top$ be the singular value decomposition of $\mathbf{A}$ with $\boldsymbol{\Lambda} = \mathrm{diag}\{\sigma_1, \ldots, \sigma_r\}$ and $\sigma_1 \geq \cdots \geq \sigma_r > 0$. Then we have $\sigma_r\|\mathbf{x}\|_{\mathrm{F}} \leq \|\mathbf{A}\mathbf{x}\|_{\mathrm{F}} \leq \sigma_1\|\mathbf{x}\|_{\mathrm{F}}$.*

Proof. Let $\mathbf{v}_j$ denotes the jth row of $\mathbf{V}^\top\mathbf{x}$ for $j \in [r]$. Write $\sigma_r^2\|\mathbf{x}\|_{\mathrm{F}}^2 \leq \|\mathbf{A}\mathbf{x}\|_{\mathrm{F}}^2 = \mathrm{tr}(\mathbf{x}^\top\mathbf{A}^\top\mathbf{A}\mathbf{x}) = \mathrm{tr}(\mathbf{x}^\top\mathbf{V}\boldsymbol{\Lambda}^2\mathbf{V}^\top\mathbf{x}) = (\sum_{j=1}^{r}\sigma_j^2\mathbf{v}_j^\top\mathbf{v}_j)^{1/2} \leq \sigma_1^2\|\mathbf{x}\|_{\mathrm{F}}^2$, where, in the inequalities above, we have used $\|\mathbf{V}^\top\mathbf{x}\|_{\mathrm{F}} = \|\mathbf{x}\|_{\mathrm{F}}$ due to the orthonormality of $\mathbf{V}$. Taking the squared root on both sides completes the proof of this lemma. □

To simplify the notation, we will use $\sigma_{\min}(m)$ and $\sigma_{\max}(m)$ to represent $\sigma_{\min}(m, \mathbf{G})$ and $\sigma_{\max}(m, \mathbf{G})$, respectively.

Lemma 3. *It holds that*

$$\kappa(\boldsymbol{\theta}_0) \geq \max_{m\geq s}\left\{\frac{\sigma_{\min}(m)}{\sqrt{m}} - \frac{2\sigma_{\max}(m)}{\sqrt{m}}\sqrt{\frac{s}{m}}\right\}\frac{s^{-1/2}}{2(1+2\sqrt{s/m})}.$$

Proof. Let $T \subset [p]$ and $\|\boldsymbol{\delta}_{T^c}\|_1^{(d,\tilde{d})} \leq \|\boldsymbol{\delta}_T\|_1^{(d,\tilde{d})}$ by (21). Let T_1 denote the largest m components of $\{\|\boldsymbol{\delta}_i\|_F\}_{i\in[p]}$, and T_2 be the subsequent m-largest, etc. Let $\mathbf{V}_{\boldsymbol{\mu}} = \mathrm{diag}(\boldsymbol{\mu} \otimes \mathbf{1}_d)$ where $\boldsymbol{\mu} \in \mathbb{R}^q$ with $\sum_{i=1}^q I(|\mu_i| \neq 0) \leq m$ and $\mathbf{1}_d = (1, \ldots, 1)^\top \in \mathbb{R}^d$. We let $\|\boldsymbol{\mu}\| = (\sum_{i=1}^q \mu_i^2)^{1/2}$ and $\|\boldsymbol{\mu}\|_\infty = \max_{i\in[q]} |\mu_i|$. Then, we have

$$\begin{aligned}\|\mathbf{G}\boldsymbol{\delta}\|_{\max}^{(d,\tilde{d})} &= \max_{\boldsymbol{\mu}} \left\| \frac{1}{\|\boldsymbol{\mu}\|} \mathbf{V}_{\boldsymbol{\mu}} \mathbf{G}\boldsymbol{\delta} \right\|_F \geq \left\| \frac{1}{\sqrt{m}\|\boldsymbol{\mu}\|_\infty} \mathbf{V}_{\boldsymbol{\mu}} \mathbf{G}\boldsymbol{\delta} \right\|_F \\ &\geq \left\| \frac{1}{\sqrt{m}\|\boldsymbol{\mu}\|_\infty} \mathbf{V}_{\boldsymbol{\mu}} \mathbf{G}_{\cdot,T_1} \boldsymbol{\delta}_{T_1} \right\|_F - \sum_{j\geq 2} \left\| \frac{1}{\sqrt{m}\|\boldsymbol{\mu}\|_\infty} \mathbf{V}_{\boldsymbol{\mu}} \mathbf{G}_{\cdot,T_j} \boldsymbol{\delta}_{T_j} \right\|_F, \end{aligned} \tag{B.1}$$

where $\mathbf{G}_{\cdot,T_j}$ is the block submatrix of $\mathbf{G}$ consisting of all rows and all block columns in T_j of $\mathbf{G}$ for $j \geq 1$.

Define $\tilde{J}_1 = \arg\max_{|J|\leq m} \sigma_{\min}(\mathbf{G}_{J,T_1})$. We can let $\boldsymbol{\mu} = (\mu_i)$ with $\mu_i = 1$ if $i \in \tilde{J}_1$ and 0 otherwise, so that $\|\boldsymbol{\mu}\|_\infty = 1$. Then the first term in (B.1) becomes

$$\begin{aligned}\left\| \frac{1}{\sqrt{m}\|\boldsymbol{\mu}\|_\infty} \mathbf{V}_{\boldsymbol{\mu}} \mathbf{G}_{\cdot,T_1} \boldsymbol{\delta}_{T_1} \right\|_F &= \left\| \frac{1}{\sqrt{m}} \mathbf{G}_{\tilde{J}_1,T_1} \boldsymbol{\delta}_{T_1} \right\|_F \\ &\geq \frac{\sigma_{\min}(\mathbf{G}_{\tilde{J}_1,T_1})}{\sqrt{m}} \|\boldsymbol{\delta}_{T_1}\|_F = \frac{1}{\sqrt{m}} \max_{|J|\leq m} \sigma_{\min}(\mathbf{G}_{J,T_1}) \|\boldsymbol{\delta}_{T_1}\|_F \\ &\geq \frac{1}{\sqrt{m}} \min_{|M|\leq m} \max_{|J|\leq m} \sigma_{\min}\left(\mathbf{G}_{J,M}\right) \|\boldsymbol{\delta}_{T_1}\|_F = \frac{\sigma_{\min}(m)}{\sqrt{m}} \|\boldsymbol{\delta}_{T_1}\|_F, \end{aligned} \tag{B.2}$$

where the first inequality comes from Lemma 2. Define $\tilde{J}_j = \arg\max_{|J|\leq m} \sigma_{\max}(\mathbf{G}_{J,T_j})$ for each $j \geq 2$. By the similar arguments as above, the second term in (B.1) becomes

$$\begin{aligned}\sum_{j\geq 2} \left\| \frac{1}{\sqrt{m}\|\boldsymbol{\mu}\|_\infty} \mathbf{V}_{\boldsymbol{\mu}} \mathbf{G}_{\cdot,T_j} \boldsymbol{\delta}_{T_j} \right\|_F &= \frac{1}{\sqrt{m}} \sum_{j\geq 2} \|\mathbf{G}_{\tilde{J}_j,T_j} \boldsymbol{\delta}_{T_j}\|_F \\ &\leq \frac{1}{\sqrt{m}} \sum_{j\geq 2} \sigma_{\max}(\mathbf{G}_{\tilde{J}_j,T_j}) \|\boldsymbol{\delta}_{T_j}\|_F = \frac{1}{\sqrt{m}} \sum_{j\geq 2} \max_{|J|\leq m} \sigma_{\max}(\mathbf{G}_{J,T_j}) \|\boldsymbol{\delta}_{T_j}\|_F \\ &\leq \frac{1}{\sqrt{m}} \max_{|M|\leq m} \max_{|J|\leq m} \sigma_{\max}\left(\mathbf{G}_{J,M}\right) \sum_{j\geq 2} \|\boldsymbol{\delta}_{T_j}\|_F = \frac{\sigma_{\max}(m)}{\sqrt{m}} \sum_{j\geq 2} \|\boldsymbol{\delta}_{T_j}\|_F. \end{aligned} \tag{B.3}$$

By the construction of sets $\{T_j\}_{j\geq 1}$, we have $\|\boldsymbol{\delta}_{T_j}\|_1^{(d,\tilde{d})} = \sum_{l\in T_j} \|\boldsymbol{\delta}_l\|_F \geq m\|\boldsymbol{\delta}_{T_{j+1}}\|_{\max}^{(d,\tilde{d})} \geq \sqrt{m}\|\boldsymbol{\delta}_{T_{j+1}}\|_F$, which implies that

$$\sum_{j\geq 2} \|\boldsymbol{\delta}_{T_j}\|_F \leq \frac{1}{\sqrt{m}} \sum_{j\geq 1} \|\boldsymbol{\delta}_{T_j}\|_1^{(d,\tilde{d})} \leq \frac{\|\boldsymbol{\delta}\|_1^{(d,\tilde{d})}}{\sqrt{m}}. \tag{B.4}$$

Combining (B.2)–(B.4) yields

$$\begin{aligned}\|\mathbf{G}\boldsymbol{\delta}\|_{\max}^{(d,\tilde{d})} &\geq \frac{\sigma_{\min}(m)}{\sqrt{m}} \|\boldsymbol{\delta}_{T_1}\|_F - \frac{\sigma_{\max}(m)}{\sqrt{m}} \|\boldsymbol{\delta}\|_1^{(d,\tilde{d})}/\sqrt{m} \\ &\geq \frac{\sigma_{\min}(m)}{\sqrt{m}} \|\boldsymbol{\delta}_{T_1}\|_F - \frac{\sigma_{\max}(m)}{\sqrt{m}} 2\sqrt{\frac{s}{m}} \|\boldsymbol{\delta}_T\|_F \\ &= \left\{ \frac{\sigma_{\min}(m)}{\sqrt{m}} - 2\frac{\sigma_{\max}(m)}{\sqrt{m}} \sqrt{\frac{s}{m}} \frac{\|\boldsymbol{\delta}_T\|_F}{\|\boldsymbol{\delta}_{T_1}\|_F} \right\} \|\boldsymbol{\delta}_{T_1}\|_F, \end{aligned} \tag{B.5}$$

where the second inequality comes from $\|\boldsymbol{\delta}\|_1^{(d,\tilde{d})} \leq 2\|\boldsymbol{\delta}_T\|_1^{(d,\tilde{d})} \leq 2\sqrt{s}\|\boldsymbol{\delta}_T\|_F$ with $|T| \leq s$. This fact together with (B.4) implies that

$$\|\boldsymbol{\delta}\|_F \leq \|\boldsymbol{\delta}_{T_1}\|_F + \sum_{j\geq 2} \|\boldsymbol{\delta}_{T_j}\|_F \leq \|\boldsymbol{\delta}_{T_1}\|_F + 2\sqrt{s/m}\|\boldsymbol{\delta}_T\|_F \leq (1 + 2\sqrt{s/m})\|\boldsymbol{\delta}_{T_1}\|_F. \tag{B.6}$$

Combining (B.5) and (B.6) yields that

$$\begin{aligned}\|\mathbf{G}\boldsymbol{\delta}\|_{\max}^{(d,\tilde{d})} &\geq \left\{ \frac{\sigma_{\min}(m)}{\sqrt{m}} - 2\frac{\sigma_{\max}(m)}{\sqrt{m}} \sqrt{\frac{s}{m}} \right\} \frac{\|\boldsymbol{\delta}\|_F}{(1 + 2\sqrt{s/m})} \\ &\geq \left\{ \frac{\sigma_{\min}(m)}{\sqrt{m}} - 2\frac{\sigma_{\max}(m)}{\sqrt{m}} \sqrt{\frac{s}{m}} \right\} \frac{\|\boldsymbol{\delta}\|_1^{(d,\tilde{d})}/\sqrt{s}}{2(1 + 2\sqrt{s/m})}, \end{aligned} \tag{B.7}$$

where the second inequality comes from $\|\boldsymbol{\delta}\|_F \geq \|\boldsymbol{\delta}_T\|_F \geq \|\boldsymbol{\delta}_T\|_1^{(d,\tilde{d})}/\sqrt{s} \geq \|\boldsymbol{\delta}\|_1^{(d,\tilde{d})}/\sqrt{4s}$. We complete our proof by (21) and dividing $\|\boldsymbol{\delta}\|_1^{(d,\tilde{d})}$ on both sides of (B.7). □

Lemma 4. *Suppose that Condition 5 holds. Then there exists some positive constant c such that $\kappa(\boldsymbol{\theta}_0) \geq c\mu^2/(24s)$.*

Proof. Applying Lemma 3 and choosing $m = 16s/\mu^2$ yields that

$$\begin{aligned}\kappa(\boldsymbol{\theta}_0) &\geq \max_{m\geq s} \frac{\sigma_{\max}(m,\mathbf{G})}{\sqrt{m}}\left\{\frac{\sigma_{\min}(m,\mathbf{G})}{\sigma_{\max}(m,\mathbf{G})} - \frac{\mu}{2}\right\}\frac{s^{-1/2}}{2(1+\mu/2)} \\ &\geq \frac{c\mu}{4\sqrt{s}}\left(\mu - \frac{\mu}{2}\right)\left\{2\left(1+\frac{\mu}{2}\right)\right\}^{-1} s^{-1/2} \geq \frac{c\mu^2}{24s},\end{aligned}$$

which completes our proof. □

For each $j \in [p]$, let $\omega_{j1} \geq \omega_{j2} \geq \cdots > 0$ be the eigenvalues of $\Sigma^X_{0,jj}$ with the corresponding eigenfunctions $\nu_{j1}(\cdot), \nu_{j2}(\cdot), \ldots$. Similarly, let $\{(\omega^W_{jl}, \nu^W_{jl}(\cdot))\}_{l=1}^{\infty}$ be the eigenpairs of $\Sigma^W_{0,jj}$.

Lemma 5. *Suppose that Condition 2 holds. Then we have $\omega^W_0 = \max_j \sum_{l=1}^{\infty} \omega^W_{jl} = O(1)$.*

Proof. This lemma follows directly from Lemma 2 of Fang et al. (2022) and hence the proof is omitted here. □

Lemma 6. *For $p \times p$ lag-h autocovariance function of $\{\mathbf{W}_t(\cdot)\}$, $\{\Sigma^W_{h,jk}(\cdot,\cdot)\}_{j,k\in[p]}$, we have $\|\Sigma^W_{h,jk}\|_{\mathcal{S}} \leq \omega^W_0$ and $\|\Sigma^W_{h,jk}(\psi_{km})\|_{\mathcal{S}} \leq \omega^{1/2}_{km}(\omega^W_0)^{1/2}$ for $m \geq 1$.*

Proof. This lemma follows directly from Lemma 8 of Guo and Qiao (2023) and hence the proof is omitted here. □

B.2. Proof of Theorem 1

Along the line of the proofs of Theorem 1 in Fang et al. (2022) and Proposition 1 in Guo and Qiao (2023), we can obtain that for $h \geq 1$

$$\mathbb{P}\left\{\left|\frac{\langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}^W_h - \boldsymbol{\Sigma}^W_h)(\boldsymbol{\Phi}_2)\rangle}{\langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}^W_0(\boldsymbol{\Phi}_1)\rangle + \langle \boldsymbol{\Phi}_2, \boldsymbol{\Sigma}^W_0(\boldsymbol{\Phi}_2)\rangle}\right| > 2\mathcal{M}^W_k\delta\right\} \leq 8\exp\{-cn\min(\delta^2,\delta)\}. \tag{B.8}$$

For each $j \in [p]$, consider the spectral decomposition $\Sigma^W_{0,jj}(u,v) = \sum_{l=1}^{\infty}\omega^W_{jl}\nu^W_{jl}(u)\nu^W_{jl}(v)$ and $\omega_0 = \max_j\sum_{l=1}^{\infty}\omega^W_{jl} = O(1)$, implied from Lemma 5. For each (j,k,l,m), choosing $\boldsymbol{\Phi}_1 = \{0,\ldots,0,(\omega^W_{jl})^{-1/2}\nu^W_{jl},0,\ldots,0\}^\top$ and $\boldsymbol{\Phi}_2 = \{0,\ldots,0,(\omega^W_{km})^{-1/2}\nu^W_{km},0,\ldots,0\}^\top$ on (B.8) and following the same procedure to prove Theorem 2 of Guo and Qiao (2023) with the choice of suitable constant $\bar{c}$, we can obtain that

$$\mathbb{P}\left\{\|\widehat{\Sigma}^W_{h,jk} - \Sigma^W_{h,jk}\|_{\mathcal{S}} > \mathcal{M}^W_1\delta\right\} \leq 8\exp\{-\bar{c}n\min(\delta^2,\delta)\}. \tag{B.9}$$

By (5), (6) and Cauchy–Schwarz inequality, we have $\|\hat{K}_{jj} - K_{jj}\|^2_{\mathcal{S}} \leq 2L\sum_{h=1}^{L}\|\widehat{\Sigma}^W_{h,jj} - \Sigma^W_{h,jj}\|^2_{\mathcal{S}}\|\Sigma^W_{h,jj}\|^2_{\mathcal{S}} + L\sum_{h=1}^{L}\|\widehat{\Sigma}^W_{h,jj} - \Sigma^W_{h,jj}\|^4_{\mathcal{S}}$. Let $\Omega^{(h)}_{\omega,jk} = \{\|\widehat{\Sigma}^W_{h,jk} - \Sigma^W_{h,jk}\|_{\mathcal{S}} \leq \omega_0\}$ and $\Omega^{(h)}_{jk} = \{\|\widehat{\Sigma}^W_{h,jk} - \Sigma^W_{h,jk}\|_{\mathcal{S}} \leq \mathcal{M}^W_1\delta\}$. On the event $\Lambda_j = \Omega^{(1)}_{\omega,jj} \cap \cdots \cap \Omega^{(L)}_{\omega,jj} \cap \Omega^{(1)}_{jj} \cap \cdots \cap \Omega^{(L)}_{jj}$, it follows from the above results and Lemma 6 that

$$\|\hat{K}_{jj} - K_{jj}\|_{\mathcal{S}} \leq \sqrt{3}L\omega_0\mathcal{M}^W_1\delta. \tag{B.10}$$

Applying (B.9) and choosing $\delta = (\mathcal{M}^W_1)^{-1}\omega_0$ for $\Omega^{(1)}_{\omega,jj},\ldots,\Omega^{(L)}_{\omega,jj}$ yields $\mathbb{P}(\Lambda^c_j) \leq 8L\exp\{-cn\min(\delta^2,\delta)\} + 8L\exp[-cn\min\{(\mathcal{M}^W_1)^{-2}\omega^2_0, (\mathcal{M}^W_1)^{-1}\omega_0\}]$. Combining the above results, we obtain

$$\mathbb{P}\left(\|\hat{K}_{jj} - K_{jj}\|_{\mathcal{S}} > \mathcal{M}^W_1\delta\right) \leq \bar{c}\exp\{-cn\min(\delta^2,\delta)\} + \bar{c}\exp(-cn). \tag{B.11}$$

For each $j \in [p]$, it follows from Lemma 4.3 of Bosq (2000) and Condition 3 with $\min_{k\in[l]}\{\lambda_{jk} - \lambda_{j(k+1)}\} \geq c_0 l^{-\alpha-1}$ that

$$\max_{l\in[d]}|\hat{\lambda}_{jl} - \lambda_{jl}| \leq \|\hat{K}_{jj} - K_{jj}\|_{\mathcal{S}} \quad \text{and} \quad \max_{l\in[d]}\left(\|\hat{\psi}_{jl} - \psi_{jl}\|/l^{\alpha+1}\right) \leq 2\sqrt{2}c_0^{-1}\|\hat{K}_{jj} - K_{jj}\|_{\mathcal{S}}. \tag{B.12}$$

Combining (B.11), (B.12) and the union bound of probability yields that

$$\begin{aligned}&\mathbb{P}\left(\max_{j\in[p],l\in[d]}|\hat{\lambda}_{jl} - \lambda_{jl}| > \mathcal{M}^W_1\delta\right) \vee \mathbb{P}\left\{\max_{j\in[p],l\in[d]}(\|\hat{\psi}_{jl} - \psi_{jl}\|/l^{\alpha+1}) > 2\sqrt{2}c_0^{-1}\mathcal{M}^W_1\delta\right\} \\ &\qquad \leq \bar{c}p\exp\{-cn\min(\delta^2,\delta)\} + \bar{c}p\exp(-cn).\end{aligned}$$

Let $\delta = \rho\sqrt{n^{-1}\log p} \leq 1$. Choosing suitable positive constants $\tilde{c}$ and $\check{c} = c\rho^2 - 1$, we obtain that (11) holds with probability greater than $1 - \tilde{c}p^{-\check{c}}$, which completes the proof of Theorem 1. □

B.3. Proof of Theorem 2

For each (j,k,l,m) and $h \geq 1$, we write

$$\begin{aligned}\hat{\sigma}^{(h)}_{jklm} - \sigma^{(h)}_{jklm} &= \langle \hat{\psi}_{jl}, \widehat{\Sigma}^{W}_{h,jk}(\hat{\psi}_{km})\rangle - \langle \psi_{jl}, \Sigma^{W}_{h,jk}(\psi_{km})\rangle \\ &= \langle (\hat{\psi}_{jl} - \psi_{jl}), \widehat{\Sigma}^{W}_{h,jk}(\hat{\psi}_{km} - \psi_{km})\rangle + \langle \psi_{jl}, (\widehat{\Sigma}^{W}_{h,jk} - \Sigma^{W}_{h,jk})(\psi_{km})\rangle \\ &\quad + \{\langle (\hat{\psi}_{jl} - \psi_{jl}), (\widehat{\Sigma}^{W}_{h,jk} - \Sigma^{W}_{h,jk})(\psi_{km})\rangle + \langle \psi_{jl}, (\widehat{\Sigma}^{W}_{h,jk} - \Sigma^{W}_{h,jk})(\hat{\psi}_{km} - \psi_{km})\rangle\} \\ &\quad + \{\langle (\hat{\psi}_{jl} - \psi_{jl}), \Sigma^{W}_{h,jk}(\psi_{km})\rangle + \langle \psi_{jl}, \Sigma^{W}_{h,jk}(\hat{\psi}_{km} - \psi_{km})\rangle\} \\ &=: J_1 + J_2 + J_3 + J_4 .\end{aligned}$$

On the event $\widetilde{\Omega}_{jk} = \Omega^{(h)}_{\omega,jk} \cap \Omega^{(h)}_{jk} \cap \Lambda_j \cap \Lambda_k$, it follows from Lemma 6, (B.10), (B.12), the orthonormality of $\{\psi_{jl}\}$, $\{\psi_{km}\}$ that $\max_{l,m\in[d]}\{|J_1|/(l\vee m)^{2(\alpha+1)}\} \lesssim (\mathcal{M}^W_1)^2\delta^2$, $\max_{l,m\in[d]}|J_2| \leq \mathcal{M}^W_1\delta$, $\max_{l,m\in[d]}\{|J_3|/(l\vee m)^{\alpha+1}\} \lesssim \mathcal{M}^W_1\delta$ and $\max_{l,m\in[d]}\{|J_4|/(l\vee m)^{\alpha+1}\} \lesssim \mathcal{M}^W_1\delta$. Then $\max_{l,m\in[d]}\{\sum_{i=1}^4 |J_i|/(l\vee m)^{\alpha+1}\} \leq c\mathcal{M}^W_1\delta + \tilde{c}d^{\alpha+1}(\mathcal{M}^W_1)^2\delta^2$. Applying (B.9) and choosing $\delta = (\mathcal{M}^W_1)^{-1}\omega_0$ for $\Omega^{(h)}_{\omega,jk}$, $\Omega^{(1)}_{\omega,jj}, \ldots, \Omega^{(L)}_{\omega,jj}$, $\Omega^{(1)}_{\omega,kk}, \ldots, \Omega^{(L)}_{\omega,kk}$ yields that $\mathbb{P}(\widetilde{\Omega}^c_{jk}) \leq (16L+8)\exp\{-cn\min(\delta^2,\delta)\} + (16L+8)\exp[-cn\min\{(\mathcal{M}^W_1)^{-2}\omega_0^2, (\mathcal{M}^W_1)^{-1}\omega_0\}]$. Combining the above results, choosing suitable positive constants $\bar{c}, \tilde{c}, \check{c}$, and applying the union bound of probability yields that

$$\mathbb{P}\left\{\max_{j,k\in[p],l,m\in[d]}\left|\frac{\hat{\sigma}^{(h)}_{jklm} - \sigma^{(h)}_{jklm}}{(l\vee m)^{\alpha+1}}\right| > \mathcal{M}^W_1\delta + \bar{c}d^{\alpha+1}(\mathcal{M}^W_1)^2\delta^2\right\} \leq \tilde{c}p^2[\exp\{-\check{c}n\min(\delta^2,\delta)\} + \exp(-\check{c}n)] . \tag{B.13}$$

Choosing $\delta = \rho_1\sqrt{n^{-1}\log p} \leq 1$ and $1 + \bar{c}d^{\alpha+1}\mathcal{M}^W_1\delta \leq \rho_2$ for some positive constants ρ_1, ρ_2, which can be achieved for sufficiently large $n \gtrsim d^{2\alpha+2}(\mathcal{M}^W_1)^2\log p$, it follows from (B.13) that there exist positive constants $c, \dot{c}$ such that, with probability greater than $1 - cp^{-\dot{c}}$,

$$\max_{j,k\in[p],l,m\in[d]}\left|\frac{\hat{\sigma}^{(h)}_{jklm} - \sigma^{(h)}_{jklm}}{(l\vee m)^{\alpha+1}}\right| \leq \rho_1\rho_2\mathcal{M}^W_1\sqrt{\frac{\log p}{n}} ,$$

which completes the proof of Theorem 2. □

B.4. Proof of Proposition 2

For each (h,j,k,l,m), we write

$$\begin{aligned}\hat{\sigma}^{W,Y}_{h,jklm} - \sigma^{W,Y}_{h,jklm} &= \langle (\hat{\psi}_{jl} - \psi_{jl}), \widehat{\Sigma}^{W,Y}_{h,jk}(\hat{\phi}_{km} - \phi_{km})\rangle + \langle \psi_{jl}, (\widehat{\Sigma}^{W,Y}_{h,jk} - \Sigma^{W,Y}_{h,jk})(\phi_{km})\rangle \\ &\quad + \{\langle (\hat{\psi}_{jl} - \psi_{jl}), (\widehat{\Sigma}^{W,Y}_{h,jk} - \Sigma^{W,Y}_{h,jk})(\phi_{km})\rangle + \langle \psi_{jl}, (\widehat{\Sigma}^{W,Y}_{h,jk} - \Sigma^{W,Y}_{h,jk})(\hat{\phi}_{km} - \phi_{km})\rangle\} \\ &\quad + \{\langle (\hat{\psi}_{jl} - \psi_{jl}), \Sigma^{W,Y}_{h,jk}(\phi_{km})\rangle + \langle \psi_{jl}, \Sigma^{W,Y}_{h,jk}(\hat{\phi}_{km} - \phi_{km})\rangle\} \\ &=: I_1 + I_2 + I_3 + I_4 .\end{aligned}$$

Let $\Omega^Y_{0kk} = \{\|\widehat{\Sigma}^Y_{0,kk} - \Sigma^Y_{0,kk}\|_{\mathcal{S}} \leq \mathcal{M}^Y_1\delta\}$ and $\Omega^{W,Y}_{hjk} = \{\|\widehat{\Sigma}^{W,Y}_{h,jk} - \Sigma^{W,Y}_{h,jk}\|_{\mathcal{S}} \leq \mathcal{M}_{W,Y}\delta\}$. On the event $\Lambda_j \cap \Omega^Y_{0,kk} \cap \Omega^{W,Y}_{h,jk}$, it follows from $\|\langle \Sigma^{W,Y}_{h,jk}, \phi_{km}\rangle\| \leq \omega_0^{1/2}\theta_{km}^{1/2}$ and $\|\langle \psi_{jl}, \Sigma^{W,Y}_{h,jk}\rangle\| \leq \omega_0^{1/2}\theta_0^{1/2}$, derived by the similar techniques to prove Lemma 6, together with Lemma 5, (B.10), (B.12), the orthonormality of $\{\psi_{jl}\}$, $\{\phi_{km}\}$ and Condition 10 that $\max_{l\in[d],m\in[\tilde{d}]}\{|I_1|/(l^{2(\alpha+1)}\vee m^{2(\tilde{\alpha}+1)})\} \lesssim (\mathcal{M}^W_1)^2\delta^2 + (\mathcal{M}^Y_1)^2\delta^2$, $\max_{l\in[d],m\in[\tilde{d}]}|I_2| \leq \mathcal{M}_{W,Y}\delta$, $\max_{l\in[d],m\in[\tilde{d}]}\{|I_3|/(l^{\alpha+1}\vee m^{\tilde{\alpha}+1})\} \lesssim \mathcal{M}^W_1\mathcal{M}_{W,Y}\delta^2 + \mathcal{M}^Y_1\mathcal{M}_{W,Y}\delta^2$ and $\max_{l\in[d],m\in[\tilde{d}]}\{|I_4|/(l^{\alpha+1}\vee m^{\tilde{\alpha}+1})\} \lesssim \mathcal{M}^W_1\delta + \mathcal{M}^Y_1\delta$. Combining the above results and $\mathcal{M}_{W,Y} = \mathcal{M}^W_1 + \mathcal{M}^Y_1 + \mathcal{M}^{W,Y}_{1,1}$ yields that $\max_{l\in[d],m\in[\tilde{d}]}\{\sum_{i=1}^4 |I_i|/(l^{\alpha+1}\vee m^{\tilde{\alpha}+1})\} \leq c\mathcal{M}_{W,Y}\delta + \bar{c}(d^{\alpha+1}\vee\tilde{d}^{\tilde{\alpha}+1})(\mathcal{M}_{W,Y})^2\delta^2$. Following the same developments to prove (B.13), we apply (B.11), Theorem 2, Lemma 24 of Fang et al. (2022) and the union bound of probability, choose suitable positive constants $\tilde{c}, \check{c}, \dot{c}$ and hence obtain that

$$\begin{aligned}&\mathbb{P}\left\{\max_{j\in[p],k\in[\tilde{p}],l\in[d],m\in[\tilde{d}]}\frac{|\hat{\sigma}^{W,Y}_{h,jklm} - \sigma^{W,Y}_{h,jklm}|}{l^{\alpha+1}\vee m^{\tilde{\alpha}+1}} > \mathcal{M}_{W,Y}\delta + \tilde{c}(d^{\alpha+1}\vee\tilde{d}^{\tilde{\alpha}+1})(\mathcal{M}_{W,Y})^2\delta^2\right\} \\ &\quad \leq \check{c}p\tilde{p}[\exp\{-\dot{c}n\min(\delta^2,\delta)\} + \exp(-\dot{c}n)] .\end{aligned} \tag{B.14}$$

Choosing $\delta = \rho_3\sqrt{n^{-1}\log(p\tilde{p})} \leq 1$ and $1 + \tilde{c}(d^{\alpha+1}\vee\tilde{d}^{\tilde{\alpha}+1})\mathcal{M}_{W,Y}\delta \leq \rho_4$ for some positive constants ρ_3, ρ_4, which can be achieved for sufficiently large $n \gtrsim (d^{2\alpha+2}\vee\tilde{d}^{2\tilde{\alpha}+2})(\mathcal{M}_{W,Y})^2\log(p\tilde{p})$, it follows from (B.14) that there exist positive constants $c, \bar{c}$ such that, with probability greater than $1 - c(p\tilde{p})^{-\bar{c}}$,

$$\max_{j\in[p],k\in[\tilde{p}],l\in[d],m\in[\tilde{d}]}\frac{|\hat{\sigma}^{W,Y}_{h,jklm} - \sigma^{W,Y}_{h,jklm}|}{l^{\alpha+1}\vee m^{\tilde{\alpha}+1}} \leq \rho_3\rho_4\mathcal{M}_{W,Y}\sqrt{\frac{\log(p\tilde{p})}{n}} ,$$

which completes the proof of Proposition 2. □

B.5. Proof of Proposition 3

For each (h, j, k, l), we write $\hat{\varrho}^{W,Z}_{h,jkl} - \varrho^{W,Z}_{h,jkl} = \langle(\hat{\psi}_{jl} - \psi_{jl}), (\widehat{\Sigma}^{W,Z}_{h,jk} - \Sigma^{W,Z}_{h,jk})\rangle + \langle\psi_{jl}, (\widehat{\Sigma}^{W,Z}_{h,jk} - \Sigma^{W,Z}_{h,jk})\rangle + \langle(\hat{\psi}_{jl} - \psi_{jl}), \Sigma^{W,Z}_{h,jk}\rangle =: \tilde{T}_1 + \tilde{T}_2 + \tilde{T}_3$. Let $\Omega^{W,Z}_{hjk} = \{\|\widehat{\Sigma}^{W,Z}_{h,jk} - \Sigma^{W,Z}_{h,jk}\|_S \le \mathcal{M}_{W,Z}\delta\}$. On the event $\Lambda_j \cap \Omega^{W,Z}_{hjk}$, it follows from (B.10), (B.12), the orthonormality of $\{\psi_{jl}\}$ and $\|\Sigma^{WZ}_{h,jk}\| \le \omega_0^{1/2}\sigma^Z_{0,kk}$ that $\max_{l\in[d]}(|\tilde{T}_1|/l^{\alpha+1}) \lesssim \mathcal{M}_1^W\delta\mathcal{M}_{W,Z}\delta$, $\max_{l\in[d]}|\tilde{T}_2| \le \mathcal{M}_{W,Z}\delta$ and $\max_{l\in[d]}(|\tilde{T}_3|/l^{\alpha+1}) \lesssim \mathcal{M}_1^W\delta$. Combining the above results and $\mathcal{M}_{W,Z} = \mathcal{M}_1^W + \mathcal{M}_1^Z + \mathcal{M}_{1,1}^{W,Z}$ implies that $\max_{l\in[d]}(\sum_{i=1}^3 |\tilde{T}_i|/l^{\alpha+1}) \le c\mathcal{M}_{W,Z}\delta + \bar{c}(\mathcal{M}_{W,Z})^2\delta^2$. Following the same developments to prove (B.13), we apply (B.11), Remark 3 and Lemma 28 of Fang et al. (2022) and the union bound of probability, choose suitable positive constants $\tilde{c}, \check{c}, \dot{c}$ and hence obtain that

$$\mathbb{P}\left\{\max_{j\in[p],k\in[\tilde{p}],l\in[d]} \frac{|\hat{\varrho}^{W,Z}_{h,jkl} - \varrho^{W,Z}_{h,jkl}|}{l^{\alpha+1}} > \mathcal{M}_{W,Z}\delta + \tilde{c}(\mathcal{M}_{W,Z})^2\delta^2\right\} \le \check{c}p\tilde{p}[\exp\{-\dot{c}n\min(\delta^2,\delta)\} + \exp(-\dot{c}n)]\,. \tag{B.15}$$

Choosing $\delta = \rho_5\sqrt{n^{-1}\log(p\tilde{p})} \le 1$ and $1 + \tilde{c}\mathcal{M}_{W,Z}\delta \le \rho_6$ for some positive constants ρ_5, ρ_6, which can be achieved for sufficiently large $n \gtrsim (\mathcal{M}_{W,Z})^2\log(p\tilde{p})$, it follows from (B.15) that there exist positive constants $c, \bar{c}$ such that, with probability greater than $1 - c(p\tilde{p})^{-\bar{c}}$,

$$\max_{j\in[p],k\in[\tilde{p}],l\in[d]} \frac{|\hat{\varrho}^{W,Z}_{h,jkl} - \varrho^{W,Z}_{h,jkl}|}{l^{\alpha+1}} \le \rho_5\rho_6\mathcal{M}_{W,Z}\sqrt{\frac{\log(p\tilde{p})}{n}}\,,$$

which completes the proof of Proposition 3. □

B.6. Proof of Theorem 3

By $\mathbf{g}(\boldsymbol{\theta}) = \mathbf{G}\boldsymbol{\theta} + \mathbf{g}(\mathbf{0})$ and (19), we have $\mathbf{g}(\hat{\boldsymbol{\theta}}) = \mathbf{G}\hat{\boldsymbol{\theta}} + \mathbf{g}(\mathbf{0})$, $\mathbf{G}\boldsymbol{\theta}_0 + \mathbf{g}(\mathbf{0}) + \mathbf{R} = \mathbf{0}$ and $\hat{\mathbf{g}}(\hat{\boldsymbol{\theta}}) = \widehat{\mathbf{G}}\hat{\boldsymbol{\theta}} + \hat{\mathbf{g}}(\mathbf{0})$. Consider the event $A = \{\|\widehat{\mathbf{G}} - \mathbf{G}\|^{(d,d)}_{\max} \vee \|\hat{\mathbf{g}}(\mathbf{0}) - \mathbf{g}(\mathbf{0})\|^{(d,\tilde{d})}_{\max} \le \epsilon_{n1}\} \cap \{\|\hat{\mathbf{g}}(\boldsymbol{\theta}_0)\|^{(d,\tilde{d})}_{\max} \le \gamma_n\}$. By the union bound of probability and Conditions 4(i) and 4(iii), this event occurs with probability at least $1 - \delta_{n1} - \delta_{n2}$. On event A, we have

$$\begin{aligned}
\|\mathbf{G}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\|^{(d,\tilde{d})}_{\max} &\le \|\mathbf{g}(\hat{\boldsymbol{\theta}})\|^{(d,\tilde{d})}_{\max} + \|\mathbf{R}\|^{(d,\tilde{d})}_{\max} \\
&\le \|\hat{\mathbf{g}}(\hat{\boldsymbol{\theta}}) - \mathbf{g}(\hat{\boldsymbol{\theta}})\|^{(d,\tilde{d})}_{\max} + \|\hat{\mathbf{g}}(\hat{\boldsymbol{\theta}})\|^{(d,\tilde{d})}_{\max} + \|\mathbf{R}\|^{(d,\tilde{d})}_{\max} \\
&\le \|(\widehat{\mathbf{G}} - \mathbf{G})\hat{\boldsymbol{\theta}}\|^{(d,\tilde{d})}_{\max} + \|\hat{\mathbf{g}}(\mathbf{0}) - \mathbf{g}(\mathbf{0})\|^{(d,\tilde{d})}_{\max} + \|\hat{\mathbf{g}}(\hat{\boldsymbol{\theta}})\|^{(d,\tilde{d})}_{\max} + \|\mathbf{R}\|^{(d,\tilde{d})}_{\max} \\
&\le \|\widehat{\mathbf{G}} - \mathbf{G}\|^{(d,d)}_{\max}\|\boldsymbol{\theta}_0\|^{(d,\tilde{d})}_1 + \|\hat{\mathbf{g}}(\mathbf{0}) - \mathbf{g}(\mathbf{0})\|^{(d,\tilde{d})}_{\max} + \|\hat{\mathbf{g}}(\hat{\boldsymbol{\theta}})\|^{(d,\tilde{d})}_{\max} + \|\mathbf{R}\|^{(d,\tilde{d})}_{\max} \\
&\le K\epsilon_{n1} + \epsilon_{n1} + \gamma_n + \epsilon_2\,,
\end{aligned} \tag{B.16}$$

where, in the last two inequalities, we have used facts that $\|(\widehat{\mathbf{G}} - \mathbf{G})\hat{\boldsymbol{\theta}}\|^{(d,\tilde{d})}_{\max} = \max_{i\in[q]}\sum_{j=1}^p \|(\widehat{\mathbf{G}} - \mathbf{G})_{ij}\hat{\boldsymbol{\theta}}_j\|_{\mathrm{F}} \le \max_{i,j}\|(\widehat{\mathbf{G}} - \mathbf{G})_{ij}\|_{\mathrm{F}}\sum_j\|\hat{\boldsymbol{\theta}}_j\|_{\mathrm{F}} = \|\widehat{\mathbf{G}} - \mathbf{G}\|^{(d,d)}_{\max}\|\hat{\boldsymbol{\theta}}\|^{(d,\tilde{d})}_1$, $\|\hat{\boldsymbol{\theta}}\|^{(d,\tilde{d})}_1 \le \|\boldsymbol{\theta}_0\|^{(d,\tilde{d})}_1 \le K$ and $\|\hat{\mathbf{g}}(\hat{\boldsymbol{\theta}})\|^{(d,\tilde{d})}_{\max} \le \gamma_n$ by the definition of the block RMD estimator in (20) and $\|\mathbf{R}\|^{(d,\tilde{d})}_{\max} \le \epsilon_2$ by Condition 4(ii).

On event A, choosing the set $T = S$ in (21) and applying Lemma 1 under Condition 4(iii) yields $\|\hat{\boldsymbol{\delta}}_{S^c}\|^{(d,\tilde{d})}_1 \le \|\hat{\boldsymbol{\delta}}_S\|^{(d,\tilde{d})}_1$ and hence $\hat{\boldsymbol{\delta}} \in C_S$. Then by (21), (B.16) and Lemma 4 under Condition 5, we have $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^{(d,\tilde{d})}_1 \le \kappa(\boldsymbol{\theta}_0)^{-1}\cdot\|\mathbf{G}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\|^{(d,\tilde{d})}_{\max} \lesssim s\mu^{-2}\{(K+1)\epsilon_{n1} + \gamma_n + \epsilon_2\}$, which completes the proof. □

B.7. Proof of Proposition 1

Define $\tilde{\kappa}(\boldsymbol{\theta}_0)$ by substituting $\mathbf{G}$ in (21) by $\widetilde{\mathbf{G}}$. By $\widetilde{\mathbf{G}} = \mathbf{D}_x\mathbf{G}\mathbf{D}_y$ with $\mathbf{D}_x$ and $\mathbf{D}_y$ being diagonal matrices, we have $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^{(d,\tilde{d})}_1 \le \tilde{\kappa}(\boldsymbol{\theta}_0)^{-1}\cdot\|\widetilde{\mathbf{G}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\|^{(d,\tilde{d})}_{\max} \le \tilde{\kappa}(\boldsymbol{\theta}_0)^{-1}\cdot\|\mathbf{D}_x\|_{\max}\|\mathbf{D}_y\|_{\max}\|\mathbf{G}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\|^{(d,\tilde{d})}_{\max}$. Following the same procedure to prove Theorem 3, we can obtain (23). □

B.8. Proof of Theorem 4

We first verify Condition 4(i) for SFLR. For sufficiently large positive constants $c, \bar{c}$, define two events

$$I_1 = \left\{\max_{j,k\in[p],h\in[L],l,m\in[d]}\left|\hat{\sigma}^{(h)}_{jklm} - \sigma^{(h)}_{jklm}\right| \le cd^{\alpha+1}\mathcal{M}_1^W\sqrt{\frac{\log p}{n}}\right\}, \tag{B.17}$$

$$I_2 = \left\{\max_{k\in[p],h\in[L],m\in[d]}\left|\frac{1}{n-h}\sum_{t=h+1}^n \hat{\eta}_{(t-h)km}Y_t - \mathbb{E}\{\eta_{(t-h)km}Y_t\}\right| \le \bar{c}d^{\alpha+1}\mathcal{M}_{W,Y}\sqrt{\frac{\log p}{n}}\right\}.$$

On event $I_1 \cap I_2$, we have

$$\|\widehat{\mathbf{G}} - \mathbf{G}\|_{\max}^{(d,d)} = \max_{j,k\in[p],h\in[L]} \left\| \frac{1}{n-h} \sum_{t=h+1}^{n} \hat{\boldsymbol{\eta}}_{(t-h)k} \hat{\boldsymbol{\eta}}_{tj}^{\top} - \mathbb{E}\{\boldsymbol{\eta}_{(t-h)k} \boldsymbol{\eta}_{tj}^{\top}\} \right\|_{\mathrm{F}} \le c d^{\alpha+2} \mathcal{M}_1^W \sqrt{\frac{\log p}{n}}, \tag{B.18}$$

$$\|\hat{\mathbf{g}}(\mathbf{0}) - \mathbf{g}(\mathbf{0})\|_{\max}^{(d,1)} = \max_{k\in[p],h\in[L]} \left\| \frac{1}{n-h} \sum_{t=h+1}^{n} \hat{\boldsymbol{\eta}}_{(t-h)k} Y_t - \mathbb{E}\{\boldsymbol{\eta}_{(t-h)k} Y_t\} \right\| \le \bar{c} d^{\alpha+3/2} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}}. \tag{B.19}$$

By Theorem 2, Proposition 3 and the union bound of probability, $\mathbb{P}(I_1 \cap I_2) \ge 1 - \tilde{c} p^{-\check{c}}$ for some positive constants $\tilde{c}, \check{c}$. By (B.18) and (B.19), Condition 4(i) can be verified by choosing $\delta_{n1} = \tilde{c} p^{-\check{c}}$ (p depends on n) and

$$\epsilon_{n1} = (c \vee \bar{c}) d^{\alpha+2} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}}. \tag{B.20}$$

We next verify Condition 4(ii) for SFLR. If follows from $r_t = \sum_{j=1}^{p} \sum_{l=d+1}^{\infty} \eta_{tjl} \langle \psi_{jl}, \beta_{0j} \rangle$, orthonormality of $\{\psi_{jl}\}$, Cauchy–Schwarz inequality and Condition 6(i) that

$$\begin{aligned}
\left\{\|\mathbf{R}\|_{\max}^{(d,1)}\right\}^2 &= \max_{k\in[p],h\in[L]} \|\mathbb{E}\{\boldsymbol{\eta}_{(t-h)k} r_t\}\|^2 = \max_{k,h} \sum_{m=1}^{d} \left\{ \mathbb{E}\left(\eta_{(t-h)km} \sum_{j=1}^{p} \sum_{l=d+1}^{\infty} \eta_{tjl} a_{jl} \right) \right\}^2 \\
&\le \max_{k,h} \sum_{m=1}^{d} \left[\sum_{j\in S} \sum_{l=d+1}^{\infty} \sqrt{\mathbb{E}\{\eta_{(t-h)km}^2\} \mathbb{E}(\eta_{tjl}^2)} a_{jl} \right]^2 \\
&\le s^2 \max_{k,j} \sum_{m=1}^{d} \left(\sum_{l=d+1}^{\infty} \lambda_{km}^{1/2} \lambda_{jl}^{1/2} a_{jl} \right)^2 \\
&\le s^2 \max_{k} \sum_{m=1}^{d} \lambda_{km} \max_{j} \left\{ \sum_{l=d+1}^{\infty} \lambda_{jl} \sum_{l=d+1}^{\infty} a_{jl}^2 \right\} \lesssim \lambda_0^2 s^2 \sum_{l=d+1}^{\infty} l^{-2\tau} = O(s^2 d^{-2\tau+1}),
\end{aligned}$$

where the asymptotic inequality comes from Condition 6(i) and $\lambda_0 = \max_j \sum_{l=1}^{\infty} \lambda_{jl} = O(1)$ implied by some calculations based on (5) and Lemma 5. Therefore

$$\|\mathbf{R}\|_{\max}^{(d,1)} \le \dot{c} s d^{-\tau+1/2} = \epsilon_2. \tag{B.21}$$

By the similar technique above and Condition 6(i),

$$\|\mathbf{b}_0\|_1^{(d,1)} = \sum_{j\in S} \left(\sum_{l=1}^{d} a_{jl}^2 \right)^{1/2} \lesssim s \max_{j\in S} \left(\sum_{l=1}^{d} l^{-2\tau} \right)^{1/2} = O(s). \tag{B.22}$$

Finally, we verify Condition 4(iii) for SFLR. On event $I_1 \cap I_2$, combining (B.20)–(B.22) yields that

$$\begin{aligned}
\|\hat{\mathbf{g}}(\mathbf{b}_0)\|_{\max}^{(d,1)} &\le \|\hat{\mathbf{g}}(\mathbf{b}_0) - \mathbf{g}(\mathbf{b}_0)\|_{\max}^{(d,1)} + \|\mathbf{R}\|_{\max}^{(d,1)} \\
&\le \|(\widehat{\mathbf{G}} - \mathbf{G})\mathbf{b}_0\|^{(d,1)} + \|\hat{\mathbf{g}}(\mathbf{0}) - \mathbf{g}(\mathbf{0})\|_{\max}^{(d,1)} + \|\mathbf{R}\|_{\max}^{(d,1)} \\
&\le \|\widehat{\mathbf{G}} - \mathbf{G}\|_{\max}^{(d,d)} \|\mathbf{b}_0\|_1^{(d,1)} + \|\hat{\mathbf{g}}(\mathbf{0}) - \mathbf{g}(\mathbf{0})\|_{\max}^{(d,1)} + \|\mathbf{R}\|_{\max}^{(d,1)} \\
&\le cs\left(d^{\alpha+2} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}} + d^{-\tau+1/2} \right) = \gamma_n.
\end{aligned}$$

By Condition 3 with $\max_j \|\mathbf{D}_j\|_{\max} \le \max_j \lambda_{jd}^{-1/2} = O(d^{\alpha/2})$ and Proposition 1 under Condition 6(ii), we have

$$\|\hat{\mathbf{b}} - \mathbf{b}_0\|_1^{(d,1)} = O_{\mathrm{p}}\left\{ \mu^{-2} s^2 d^{\alpha} \left(d^{\alpha+2} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}} + d^{-\tau+1/2} \right) \right\}. \tag{B.23}$$

For each $j \in [p]$, let $R_j(u) = \sum_{l=d+1}^{\infty} a_{jl} \psi_{jl}(u)$. By the orthonormality of $\{\psi_{jl}\}$ and $\|R_j\|^2 = \|\sum_{l=d+1}^{\infty} a_{jl}\psi_{jl}\|^2 = \sum_{l=d+1}^{\infty} a_{jl}^2 \lesssim d^{-2\tau+1}$ for $j \in S$ under Condition 6(i), we have

$$\begin{aligned}
\|\hat{\beta}_j - \beta_{0j}\| = \|\hat{\boldsymbol{\psi}}_j^{\top} \hat{\mathbf{b}}_j - \boldsymbol{\psi}_j^{\top} \mathbf{b}_{0j} - R_j\| &\le \|(\hat{\boldsymbol{\psi}}_j - \boldsymbol{\psi}_j)^{\top} \hat{\mathbf{b}}_j\| + \|\boldsymbol{\psi}_j^{\top} (\hat{\mathbf{b}}_j - \mathbf{b}_{0j})\| + \|R_j\| \\
&\le d^{1/2} \max_{l\in[d]} \|\hat{\psi}_{jl} - \psi_{jl}\| \|\hat{\mathbf{b}}_j\| + \|\hat{\mathbf{b}}_j - \mathbf{b}_{0j}\| + O(d^{-\tau+1/2}),
\end{aligned}$$

which implies that $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1 \le d^{1/2} \max_{j\in[p],l\in[d]} \|\hat{\psi}_{jl} - \psi_{jl}\| \|\hat{\mathbf{b}}\|_1^{(d,1)} + \|\hat{\mathbf{b}} - \mathbf{b}_0\|_1^{(d,1)} + O(sd^{-\tau+1/2})$, where the third term above is of a smaller order of the second term due to (B.23). By $\|\hat{\mathbf{b}}\|_1^{(d,1)} \le \|\hat{\mathbf{b}} - \mathbf{b}_0\|_1^{(d,1)} + \|\mathbf{b}_0\|_1^{(d,1)}$, (B.22) and Theorem 1, the first term above is of a smaller order of the second term. Hence, we obtain (25) from (B.23), which completes the proof. □

B.9. Proof of Theorem 5

We first verify Condition 4(i) for FFLR. In addition to event I_1 in (B.17), we define event

$$I_3 = \left\{ \max_{k\in[p],h\in[L],m\in[d],l\in[\tilde{d}]} \left| \frac{1}{n-h} \sum_{t=h+1}^{n} \hat{\eta}_{(t-h)km}\hat{\zeta}_{tl} - \mathbb{E}\{\eta_{(t-h)km}\zeta_{tl}\} \right| \le \bar{c} d^{\alpha\vee\tilde{\alpha}+1} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}} \right\}$$

for some sufficiently large $\bar{c}$. On event $I_1 \cap I_3$, we have

$$\|\hat{\mathbf{g}}(\mathbf{0}) - \mathbf{g}(\mathbf{0})\|_{\max}^{(d,\tilde{d})} = \max_{k\in[p],h\in L} \left\| \frac{1}{n-h} \sum_{t=h+1}^{n} \hat{\boldsymbol{\eta}}_{(t-h)k} \hat{\boldsymbol{\zeta}}_t^\top - \mathbb{E}\{\boldsymbol{\eta}_{(t-h)k}\boldsymbol{\zeta}_t^\top\} \right\|_{\mathrm{F}} \le \bar{c} d^{\alpha\vee\tilde{\alpha}+2} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}} . \tag{B.24}$$

By Theorem 2, Proposition 2 and the union bound probability, $\mathbb{P}(I_1 \cap I_3) \ge 1 - \tilde{c}p^{-\check{c}}$ for some positive constants $\tilde{c}, \check{c}$. By (B.18) and (B.24), Condition 4(i) can be verified with the choice of

$$\epsilon_{n1} = (c \vee \bar{c}) d^{\alpha\vee\tilde{\alpha}+2} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}} . \tag{B.25}$$

We next verify Condition 4(ii) for FFLR. If follows from $\mathbf{r}_t = (r_{t1}, \ldots, r_{t\tilde{d}})^\top$ with each $r_{tm'} = \sum_{j=1}^{p}\sum_{l=d+1}^{\infty} \eta_{tjl}\langle\langle\psi_{jl}, \beta_{0j}\rangle, \phi_{m'}\rangle$, orthonormality of $\{\psi_{jl}\}$, $\{\phi_{m'}\}$, Cauchy–Schwarz inequality and Condition 7 that

$$\begin{aligned}
\left\{\|\mathbf{R}\|_{\max}^{(d,\tilde{d})}\right\}^2 &= \max_{k\in[p],h\in[L]} \|\mathbb{E}\{\boldsymbol{\eta}_{(t-h)k}\mathbf{r}_t^\top\}\|_{\mathrm{F}}^2 = \max_{k,h} \sum_{m=1}^{d}\sum_{m'=1}^{\tilde{d}} \left\{ \mathbb{E}\left(\eta_{(t-h)km} \sum_{j=1}^{p}\sum_{l=d+1}^{\infty} \eta_{tjl} a_{jlm'} \right) \right\}^2 \\
&\le \max_{k,h} \sum_{m=1}^{d}\sum_{m'=1}^{\tilde{d}} \left[\sum_{j\in S}\sum_{l=d+1}^{\infty} \sqrt{\mathbb{E}\{\eta_{(t-h)km}^2\}\mathbb{E}(\eta_{tjl}^2)} a_{jlm'} \right]^2 \\
&\le s^2 \max_{k,j} \sum_{m=1}^{d}\sum_{m'=1}^{\tilde{d}} \left(\sum_{l=d+1}^{\infty} \lambda_{km}^{1/2}\lambda_{jl}^{1/2} a_{jlm'} \right)^2 \\
&\le s^2 \max_{k} \sum_{m=1}^{d} \lambda_{km} \max_{j} \left\{ \sum_{l=d+1}^{\infty} \lambda_{jl} \sum_{m'=1}^{\tilde{d}}\sum_{l=d+1}^{\infty} a_{jlm'}^2 \right\} \\
&\lesssim \lambda_0^2 s^2 \sum_{m'=1}^{\tilde{d}}\sum_{l=d+1}^{\infty} (l+m')^{-2\tau-1} = O(s^2 d^{-2\tau+1}) ,
\end{aligned}$$

which implies that

$$\|\mathbf{R}\|_{\max}^{(d,\tilde{d})} \le \dot{c} s d^{-\tau+1/2} = \epsilon_2 . \tag{B.26}$$

By the similar technique above and Condition 7,

$$\|\mathbf{B}_0\|_1^{(d,\tilde{d})} = \sum_{j\in S} \left(\sum_{l=1}^{d}\sum_{m=1}^{\tilde{d}} a_{jlm}^2 \right)^{1/2} \lesssim s \max_{j\in S} \left\{ \sum_{l=1}^{d}\sum_{m=1}^{\tilde{d}} (l+m)^{-2\tau-1} \right\}^{1/2} = O(s) . \tag{B.27}$$

Finally, we verify Condition 4(iii) for FFLR. On event $I_1 \cap I_3$, combining (B.25)–(B.27) and applying the similar techniques for SFLR, we have

$$\begin{aligned}
\|\hat{\mathbf{g}}(\mathbf{B}_0)\|_{\max}^{(d,\tilde{d})} &\le \|\widehat{\mathbf{G}} - \mathbf{G}\|_{\max}^{(d,d)} \|\mathbf{B}_0\|_1^{(d,\tilde{d})} + \|\hat{\mathbf{g}}(\mathbf{0}) - \mathbf{g}(\mathbf{0})\|_{\max}^{(d,\tilde{d})} + \|\mathbf{R}\|_{\max}^{(d,\tilde{d})} \\
&\le cs \left(d^{\alpha\vee\tilde{\alpha}+2} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}} + d^{-\tau+1/2} \right) = \gamma_n .
\end{aligned}$$

By Condition 3 and Proposition 1 under Condition 6(ii), we have

$$\|\widehat{\mathbf{B}} - \mathbf{B}_0\|_1^{(d,\tilde{d})} = O_{\mathrm{p}}\left\{ \mu^{-2} s^2 d^{\alpha} \left(d^{\alpha\vee\tilde{\alpha}+2} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}} + d^{-\tau+1/2} \right) \right\} . \tag{B.28}$$

For each $j \in [p]$, let $R_j(u,v) = (\sum_{l=1}^{d}\sum_{m=1}^{\tilde{d}} - \sum_{l,m=1}^{\infty}) a_{jlm}\psi_{jl}(u)\phi_m(v)$ and write

$$\begin{aligned}
\hat{\beta}_j(u,v) - \beta_{0j}(u,v) &= \hat{\boldsymbol{\psi}}_j(u)^\top \widehat{\mathbf{B}}_j \hat{\boldsymbol{\phi}}(v) - \boldsymbol{\psi}_j(u)^\top \mathbf{B}_{0j} \boldsymbol{\phi}(v) + R_j(u,v) \\
&= \hat{\boldsymbol{\psi}}_j(u)^\top \widehat{\mathbf{B}}_j \{\hat{\boldsymbol{\phi}}(v) - \boldsymbol{\phi}(v)\} + \{\hat{\boldsymbol{\psi}}_j(u) - \boldsymbol{\psi}_j(u)\}^\top \widehat{\mathbf{B}}_j \boldsymbol{\phi}(v) \\
&\quad + \boldsymbol{\psi}_j(u)^\top (\widehat{\mathbf{B}}_j - \mathbf{B}_{0j}) \boldsymbol{\phi}(v) + R_j(u,v) .
\end{aligned}$$

By Lemma 9 of Guo and Qiao (2023), we bound the first three terms by

$$\begin{aligned}
&\big\|\hat{\boldsymbol{\psi}}_j^\top \widehat{\mathbf{B}}_j(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi})\big\|_{\mathcal{S}} \le \tilde{d}^{1/2}\max_{m\in[\tilde{d}]}\|\hat{\phi}_m-\phi_m\|\|\widehat{\mathbf{B}}_j\|_{\mathrm{F}}\,,\\
&\big\|(\hat{\boldsymbol{\psi}}_j-\boldsymbol{\psi}_j)^\top\widehat{\mathbf{B}}_j\boldsymbol{\phi}\big\|_{\mathcal{S}} \le d^{1/2}\max_{l\in[d]}\|\hat{\psi}_{jl}-\psi_{jl}\|\|\widehat{\mathbf{B}}_j\|_{\mathrm{F}}\,,\\
&\big\|\boldsymbol{\psi}_j^\top(\widehat{\mathbf{B}}_j-\mathbf{B}_{0j})\boldsymbol{\phi}\big\|_{\mathcal{S}} = \|\widehat{\mathbf{B}}_j-\mathbf{B}_{0j}\|_{\mathrm{F}}\,.
\end{aligned} \tag{B.29}$$

We next bound the fourth term. For $j\in S$, by the orthonormality of $\{\psi_{jl}\}$ and $\{\phi_m\}$,

$$\begin{aligned}
\|R_j\|_{\mathcal{S}}^2 &= \left\|\left(\sum_{l=1}^{d}\sum_{m=1}^{\tilde{d}}-\sum_{l,m=1}^{\infty}\right)a_{jlm}\psi_{jl}\phi_m\right\|_{\mathcal{S}}^2\\
&= O(1)\cdot\left(\sum_{l=1}^{d}\sum_{m=\tilde{d}+1}^{\infty}a_{jlm}^2+\sum_{l=1}^{\infty}\sum_{m=1}^{\tilde{d}}a_{jlm}^2\right)\\
&= O(1)\cdot\left\{\sum_{l=1}^{d}\sum_{m=\tilde{d}+1}^{\infty}(l+m)^{-2\tau-1}+\sum_{l=1}^{\infty}\sum_{m=1}^{\tilde{d}}(l+m)^{-2\tau-1}\right\}=O(d^{-2\tau+1})\,.
\end{aligned} \tag{B.30}$$

Combining (B.29) and (B.30), we obtain $\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_1 \le \|\widehat{\mathbf{B}}\|_1^{(d,\tilde{d})}\{\tilde{d}^{1/2}\max_{m\in[\tilde{d}]}\|\hat{\phi}_m-\phi_m\|+d^{1/2}\max_{j\in[p],l\in[d]}\|\hat{\psi}_{jl}-\psi_{jl}\|\}+\|\widehat{\mathbf{B}}-\mathbf{B}_0\|_1^{(d,\tilde{d})}+O(sd^{-\tau+1/2})$, where the third term above is of a smaller order of the second term due to (B.28). By $\|\widehat{\mathbf{B}}\|_1^{(d,\tilde{d})}\le\|\widehat{\mathbf{B}}-\mathbf{B}_0\|_1^{(d,\tilde{d})}+\|\mathbf{B}_0\|_1^{(d,\tilde{d})}$, (B.27) and Theorem 1, the first term is of a smaller order of the second term. According to (B.28), we complete the proof. □

B.10. Proof of Theorem 6

For each $j\in[p]$, we first verify Condition 4(i) for VFAR. On event I_1 in (B.17),

$$\begin{aligned}
\|\widehat{\mathbf{G}}_j-\mathbf{G}_j\|_{\max}^{(d,d)} &= \max_{j',k\in[p],h\in[L],h'\in[H]}\left\|\frac{1}{n-H-h}\sum_{t=H+h+1}^{n}\hat{\boldsymbol{\eta}}_{(t-H-h)k}\hat{\boldsymbol{\eta}}_{(t-h')j'}^\top-\mathbb{E}\{\boldsymbol{\eta}_{(t-H-h)k}\boldsymbol{\eta}_{(t-h')j'}^\top\}\right\|_{\mathrm{F}}\\
&\le cd^{\alpha+2}\mathcal{M}_1^W\sqrt{\frac{\log p}{n}}\,,
\end{aligned} \tag{B.31}$$

$$\begin{aligned}
\|\hat{\mathbf{g}}_j(\mathbf{0})-\mathbf{g}_j(\mathbf{0})\|_{\max}^{(d,d)} &= \max_{k\in[p],h\in[L]}\left\|\frac{1}{n-H-h}\sum_{t=H+h+1}^{n}\hat{\boldsymbol{\eta}}_{(t-H-h)k}\hat{\boldsymbol{\eta}}_{tj}^\top-\mathbb{E}\{\boldsymbol{\eta}_{(t-H-h)k}\boldsymbol{\eta}_{tj}^\top\}\right\|_{\mathrm{F}}\\
&\le cd^{\alpha+2}\mathcal{M}_1^W\sqrt{\frac{\log p}{n}}\,.
\end{aligned} \tag{B.32}$$

It follows from Theorem 2 that $\mathbb{P}(I_1)\ge 1-\tilde{c}p^{-\check{c}}$ for some positive constants $\tilde{c},\check{c}$. By (B.31) and (B.32), Condition 4(i) can be verified by choosing

$$\epsilon_{n1}=cd^{\alpha+2}\mathcal{M}_1^W\sqrt{\frac{\log p}{n}}\,. \tag{B.33}$$

We next verify Condition 4(ii) for VFAR. It follows from $\mathbf{r}_{tj}=(r_{tj1},\ldots,r_{tjd})^\top$ with each $r_{tjm'}=\sum_{h'=1}^{H}\sum_{j'=1}^{p}\sum_{l=d+1}^{\infty}\eta_{(t-h')j'l}\langle\langle\psi_{j'l},A_{0,jj'}^{(h')}\rangle,\psi_{jm'}\rangle$, orthonormality of $\{\psi_{jl}\}$, Cauchy–Schwarz inequality and Condition 8(i) that

$$\begin{aligned}
\big\{\|\mathbf{R}_j\|_{\max}^{(d,d)}\big\}^2 &= \max_{k\in[p],h\in[L]}\|\mathbb{E}\{\boldsymbol{\eta}_{(t-H-h)k}\mathbf{r}_{tj}^\top\}\|_{\mathrm{F}}^2\\
&= \max_{k,h}\sum_{m=1}^{d}\sum_{m'=1}^{d}\left[\mathbb{E}\left\{\eta_{(t-H-h)km}\sum_{h'=1}^{H}\sum_{j'=1}^{p}\sum_{l=d+1}^{\infty}\eta_{(t-h')j'l}a_{jj'lm'}^{(h')}\right\}\right]^2\\
&\le \max_{k,h}\sum_{m=1}^{d}\sum_{m'=1}^{d}\left[\sum_{(j',h')\in S_j}\sum_{l=d+1}^{\infty}\sqrt{\mathbb{E}\{\eta_{(t-H-h)km}^2\}\mathbb{E}\{\eta_{(t-h')j'l}^2\}}a_{jj'lm'}^{(h')}\right]^2\\
&\le s_j^2\max_{k,j',h'}\sum_{m=1}^{d}\sum_{m'=1}^{d}\left\{\sum_{l=d+1}^{\infty}\lambda_{km}^{1/2}\lambda_{j'l}^{1/2}a_{jj'lm'}^{(h')}\right\}^2
\end{aligned}$$

$$\le s_j^2 \max_k \sum_{m=1}^{d} \lambda_{km} \max_{j',h'} \left[\sum_{l=d+1}^{\infty} \lambda_{j'l} \sum_{m'=1}^{d} \sum_{l=d+1}^{\infty} \{a_{jj'lm'}^{(h')}\}^2 \right]$$
$$\lesssim \lambda_0^2 s_j^2 \sum_{m'=1}^{d} \sum_{l=d+1}^{\infty} (l+m')^{-2\tau-1} = O(s_j^2 d^{-2\tau+1}),$$

which implies that

$$\|\mathbf{R}_j\|_{\max}^{(d,d)} \le \dot{c} s_j d^{-\tau+1/2} = \epsilon_2 . \tag{B.34}$$

By the similar technique above and Condition 8(i), we have

$$\|\boldsymbol{\Omega}_{0j}\|_1^{(d,d)} = \sum_{(j',h')\in S_j} \left[\sum_{l=1}^{d} \sum_{m=1}^{d} \{a_{jj'lm}^{(h')}\}^2 \right]^{1/2} \lesssim s_j \max_{(j',h')\in S_j} \left\{ \sum_{l=1}^{d} \sum_{m=1}^{d} (l+m)^{-2\tau-1} \right\}^{1/2} = O(s_j) . \tag{B.35}$$

Finally, we verify Condition 4(iii) for VFAR. On event I_1, combining (B.33), (B.34), (B.35) and applying the similar techniques, we have

$$\|\hat{\mathbf{g}}_j(\boldsymbol{\Omega}_{0j})\|_{\max}^{(d,d)} \le \|\widehat{\mathbf{G}}_j - \mathbf{G}_j\|_{\max}^{(d,d)} \|\boldsymbol{\Omega}_{0j}\|_1^{(d,d)} + \|\hat{\mathbf{g}}_j(\mathbf{0}) - \mathbf{g}_j(\mathbf{0})\|_{\max}^{(d,d)} + \|\mathbf{R}_j\|_{\max}^{(d,d)}$$
$$\le c s_j \left(d^{\alpha+2} \mathcal{M}_1^W \sqrt{\frac{\log p}{n}} + d^{-\tau+1/2} \right) = \gamma_{nj} .$$

By Condition 3 and Proposition 1 under Condition 8(ii), we have

$$\|\widehat{\boldsymbol{\Omega}}_j - \boldsymbol{\Omega}_{0j}\|_1^{(d,d)} = O_{\mathrm{p}} \left\{ \mu_j^{-2} s_j^2 d^{\alpha} \left(d^{\alpha+2} \mathcal{M}_1^W \sqrt{\frac{\log p}{n}} + d^{-\tau+1/2} \right) \right\} . \tag{B.36}$$

For each $j' \in [p]$, let $R_{jj'}^{(h')}(u,v) = (\sum_{l=1}^{d} \sum_{m=1}^{d} - \sum_{l,m=1}^{\infty}) a_{jj'lm}^{(h')} \psi_{j'm}(u) \psi_{jl}(v)$ and write

$$\hat{A}_{jj'}^{(h')}(u,v) - A_{0,jj'}^{(h')}(u,v) = \hat{\boldsymbol{\psi}}_{j'}(u)^{\top} \widehat{\boldsymbol{\Omega}}_{jj'}^{(h')} \hat{\boldsymbol{\psi}}_j(v) - \boldsymbol{\psi}_{j'}(u)^{\top} \boldsymbol{\Omega}_{0,jj'}^{(h')} \boldsymbol{\psi}_j(v) + R_{jj'}^{(h')}(u,v)$$
$$= \hat{\boldsymbol{\psi}}_{j'}(u)^{\top} \widehat{\boldsymbol{\Omega}}_{jj'}^{(h')} \{\hat{\boldsymbol{\psi}}_j(v) - \boldsymbol{\psi}_j(v)\} + \{\hat{\boldsymbol{\psi}}_{j'}(u) - \boldsymbol{\psi}_{j'}(u)\}^{\top} \widehat{\boldsymbol{\Omega}}_{jj'}^{(h')} \boldsymbol{\psi}_j(v)$$
$$+ \boldsymbol{\psi}_{j'}(u)^{\top} \{\widehat{\boldsymbol{\Omega}}_{jj'}^{(h')} - \boldsymbol{\Omega}_{0,jj'}^{(h')}\} \boldsymbol{\psi}_j(v) + R_{jj'}^{(h')}(u,v) .$$

By the same techniques to prove (B.29), we bound the first three terms

$$\left\| \hat{\boldsymbol{\psi}}_{j'}^{\top} \widehat{\boldsymbol{\Omega}}_{jj'}^{(h')} (\hat{\boldsymbol{\psi}}_j - \boldsymbol{\psi}_j) \right\|_{\mathcal{S}} \le d^{1/2} \max_{l\in[d]} \|\hat{\psi}_{jl} - \psi_{jl}\| \|\widehat{\boldsymbol{\Omega}}_{jj'}^{(h')}\|_{\mathrm{F}} ,$$
$$\left\| (\hat{\boldsymbol{\psi}}_{j'} - \boldsymbol{\psi}_{j'})^{\top} \widehat{\boldsymbol{\Omega}}_{jj'}^{(h')} \boldsymbol{\psi}_j \right\|_{\mathcal{S}} \le d^{1/2} \max_{m\in[d]} \|\hat{\psi}_{j'm} - \psi_{j'm}\| \|\widehat{\boldsymbol{\Omega}}_{jj'}^{(h')}\|_{\mathrm{F}} , \tag{B.37}$$
$$\left\| \boldsymbol{\psi}_{j'}^{\top} \{\widehat{\boldsymbol{\Omega}}_{jj'}^{(h')} - \boldsymbol{\Omega}_{0,jj'}^{(h')}\} \boldsymbol{\psi}_j \right\|_{\mathcal{S}} = \|\widehat{\boldsymbol{\Omega}}_{jj'}^{(h')} - \boldsymbol{\Omega}_{0,jj'}^{(h')}\|_{\mathrm{F}} .$$

We next bound the fourth term. For $(j',h') \in S_j$, by the orthonormality of $\{\psi_{jl}\}$,

$$\left\| R_{jj'}^{(h')} \right\|_{\mathcal{S}}^2 = \left\| \left(\sum_{l=1}^{d} \sum_{m=1}^{d} - \sum_{l,m=1}^{\infty} \right) a_{jj'lm}^{(h')} \psi_{jl} \psi_{j'm} \right\|_{\mathcal{S}}^2$$
$$= O(1) \left[\sum_{l=1}^{d} \sum_{m=d+1}^{\infty} \{a_{jj'lm}^{(h')}\}^2 \right] = O(1) \left\{ \sum_{l=1}^{d} \sum_{m=d+1}^{\infty} (l+m)^{-2\tau-1} \right\} = O(d^{-2\tau+1}) . \tag{B.38}$$

Combining (B.37) and (B.38), we obtain

$$\max_{j\in[p]} \sum_{j'=1}^{p} \sum_{h'=1}^{H} \|\hat{A}_{jj'}^{(h')} - A_{0,jj'}^{(h')}\|_{\mathcal{S}} \le \max_j \|\widehat{\boldsymbol{\Omega}}_j\|_1^{(d,d)} \left\{ d^{1/2} \max_{j\in[p],l\in[d]} \|\hat{\psi}_{jl} - \psi_{jl}\| + d^{1/2} \max_{j'\in[p],m\in[d]} \|\hat{\psi}_{j'm} - \psi_{j'm}\| \right\}$$
$$+ \max_j \|\widehat{\boldsymbol{\Omega}}_j - \boldsymbol{\Omega}_{0j}\|_1^{(d,d)} + O(s_j d^{-\tau+1/2}) ,$$

where the third term above is of a smaller order of the second term due to (B.36). By $\max_j \|\widehat{\boldsymbol{\Omega}}_j\|_1^{(d,d)} \le \max_j \|\widehat{\boldsymbol{\Omega}}_j - \boldsymbol{\Omega}_{0j}\|_1^{(d,d)} + \max_j \|\boldsymbol{\Omega}_{0j}\|_1^{(d,d)}$, (B.35) and Theorem 1, the first term is of a smaller order of the second term. Applying (B.36) with $\mu = \min_j \mu_j$ and $s = \max_j s_j$ completes our proof. □

Table 2
List of S&P 100 stocks.

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

Appendix C. List of S&P 100 component stocks used in Section 5.2

See Table 2.

References

Aue, A., Norinho, D., Hörmann, S., 2015. On the prediction of stationary functional time series. J. Amer. Statist. Assoc. 110 (509), 378–392.
Basu, S., Michailidis, G., 2015. Regularized estimation in sparse high-dimensional time series models. Ann. Statist. 43 (4), 1535–1567.
Bathia, N., Yao, Q., Ziegelmann, F., 2010. Identifying the finite dimensionality of curve time series. Ann. Statist. 38 (6), 3352–3386.
Belloni, A., Chernozhukov, V., Chetverikov, D., Hansen, C., Kato, K., 2018. High-dimensional econometrics and regularized GMM. arXiv:1806.01888.
Bosq, D., 2000. Linear Processes in Function Spaces: Theory and Applications, Vol. 149. Springer Science & Business Media.
Chen, C., Guo, S., Qiao, X., 2022. Functional linear regression: Dependence and error contamination. J. Bus. Econom. Statist. 40 (1), 444–457.
Cho, H., Goude, Y., Brossat, X., Yao, Q., 2013. Modeling and forecasting daily electricity load curves: A hybrid approach. J. Amer. Statist. Assoc. 108 (501), 7–21.
Descary, M.-H., Panaretos, V.M., 2019. Functional data analysis by matrix completion. Ann. Statist. 47 (1), 1–38.
Fan, Y., Foutz, N., James, G.M., Jank, W., 2014. Functional response additive model estimation with online virtual stock markets. Ann. Appl. Stat. 8 (4), 2435–2460.

Fan, Y., James, G.M., Radchenko, P., 2015. Functional additive regression. Ann. Statist. 43 (5), 2296–2325.
Fang, Q., Guo, S., Qiao, X., 2022. Finite sample theory for high-dimensional functional/scalar time series with applications. Electron. J. Stat. 16 (1), 527–591.
Fu, A., Narasimhan, B., Boyd, S., 2020. CVXR: An R package for disciplined convex optimization. J. Stat. Softw. 94 (14), 1–34.
Gautier, E., Rose, C., 2019. High-dimensional instrumental variables regression and confidence sets. arXiv:1105.2454.
Guo, S., Qiao, X., 2023. On consistency and sparsity for high-dimensional functional time series with application to autoregressions. Bernoulli 29 (1), 451–472.
Hall, P., Horowitz, J.L., 2007. Methodology and convergence rates for functional linear regression. Ann. Statist. 35 (1), 70–91.
Hall, P., Vial, C., 2006. Assessing the finite dimensionality of functional data. J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (4), 689–705.
Hamilton, J.D., 1994. Time Series Analysis, Vol. 2. Princeton, New Jersey.
Hörmann, S., Kidziński, Ł., Hallin, M., 2015. Dynamic functional principal components. J. R. Stat. Soc. Ser. B Stat. Methodol. 77 (2), 319–348.
Hörmann, S., Kokoszka, P., 2010. Weakly dependent functional data. Ann. Statist. 38 (3), 1845–1884.
Horváth, L., Kokoszka, P., Rice, G., 2014. Testing stationarity of functional time series. J. Econometrics 179 (1), 66–82.
Kong, D., Xue, K., Yao, F., Zhang, H.H., 2016. Partially functional linear regression in high dimensions. Biometrika 103 (1), 147–159.
Li, D., Robinson, P.M., Shang, H.L., 2020. Long-range dependent curve time series. J. Amer. Statist. Assoc. 115 (530), 957–971.
Luo, R., Qi, X., 2017. Function-on-function linear regression by signal compression. J. Amer. Statist. Assoc. 112 (518), 690–705.
Müller, H.-G., Sen, R., Stadtmüller, U., 2011. Functional data analysis for volatility. J. Econometrics 165 (2), 233–245.
Panaretos, V.M., Tavakoli, S., 2013. Fourier analysis of stationary time series in function space. Ann. Statist. 41 (2), 568–603.
Rudelson, M., Vershynin, R., 2013. Hanson-Wright inequality and sub-Gaussian concentration. Electron. Commun. Probab. 18, 1–9.
Shang, H.L., 2013. Ftsa: An R package for analyzing functional time series. R J. 5 (1), 64–72.
Xue, K., Yao, F., 2021. Hypothesis testing in large-scale functional linear regression. Statist. Sinica 31, 1101–1123.
Yao, F., Müller, H.-G., Wang, J.-L., 2005. Functional data analysis for sparse longitudinal data. J. Amer. Statist. Assoc. 100 (470), 577–590.
Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (1), 49–67.