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

Jinyuan Chang¹, Cheng Chen¹, Xinghao Qiao², and Qiwei Yao²

¹Joint Laboratory of Data Science and Business Intelligence, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China
²Department of Statistics, London School of Economics, London, WC2A 2AE, UK

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.

Key words: Block regularized minimum distance estimation; Dimension reduction; Functional time series; High-dimensional data; Non-asymptotics; Sparsity.

JEL code: C13, C32, C55

*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: Xinghao Qiao and Qiwei Yao. Email 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)
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 of Yao et al. (2005) and banded 0 e of Descary and Panaretos (2019). In contrast, we do not impose any structures on Σ X. To separate e version of model (1). When Σ structures were imposed; see, e.g., diagonal Σ of Yao et al. (2005) and banded Σ of Descary and Panaretos (2019). In contrast, we do not impose any structures on Σ in this paper, and our estimation filters out the impact of e 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, 2022), 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 \( W_t(\cdot) \) are subject to the error contamination as in model (1). Both FPCA and penalized least squares estimation based on the estimated covariance function \( \hat{\Sigma}_W^W \) are inappropriate since \( \hat{\Sigma}_W^W = \Sigma_X^X + \Sigma_e^e \) and, hence, \( \hat{\Sigma}_W^W \) is no longer a consistent estimator for \( \Sigma_X^X \). Motivated from the fact that \( \Sigma_h^W = \Sigma_h^X \) for any \( h \neq 0 \), which automatically removes the impact from the noise \( e_t(\cdot) \) and ensures that the estimator for \( \Sigma_h^W \) is also legitimate for \( \Sigma_h^X \), 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 \( X_{ij}(\cdot) \) based on the eigenanalysis of a positive-definite operator defined in terms of the autocovariance functions of \( W_{ij}(\cdot) \). 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
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 $\ell_1$ norm objective function subject to the block $\ell_\infty$ 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 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 denote $[q] = \{1, \ldots, q\}$. 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 $\langle f, g \rangle = \int_\mathcal{U} f(u)g(u) \, du$. 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 $\mathbf{f} = (f_1, \ldots, f_{p})^T$ and $\mathbf{g} = (g_1, \ldots, g_{p})^T$ 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 $\ell_0$-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 $K_{ij} \in \mathbb{S}$, $\mathbf{K}(\mathbf{f})(u) = \{\sum_{j=1}^{p} \langle K_{ij}(u, \cdot), f_j(\cdot) \rangle, \ldots, \sum_{j=1}^{p} \langle K_{ij}(u, \cdot), f_j(\cdot) \rangle\}^T \in \mathbb{H}^q$ for any $\mathbf{f} = (f_1, \ldots, f_{p})^T \in \mathbb{H}^p$. 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 $\ell_\infty$-norms by $\|\mathbf{K}\|_F = (\sum_{i=1}^{q} \sum_{j=1}^{p} \|K_{ij}\|_{\mathbb{S}}^2)^{1/2}$ and $\|\mathbf{K}\|_\infty = \max_{i \in [q]} \sum_{j=1}^{p} \|K_{ij}\|_{\mathbb{S}}$, respectively, where $\|K_{ij}\|_{\mathbb{S}} = \left(\int_\mathcal{U} \int_{\mathcal{U}} K_{ij}^2(u, v) \, du \, dv\right)^{1/2}$ denotes the Hilbert–Schmidt norm of $K_{ij}$. For any real matrix $\mathbf{B} = (b_{ij})_{q \times p}$, we write $\|\mathbf{B}\|_{\text{max}} = \max_{i \in [q], j \in [p]} |b_{ij}|$ and use $\|\mathbf{B}\|_F = (\sum_{i=1}^{q} \sum_{j=1}^{p} |b_{ij}|^2)^{1/2}$ and $\|\mathbf{B}\|_2 = \lambda_{\max}^{1/2}((\mathbf{B}^T \mathbf{B}))$ to denote its Frobenius norm and $\ell_2$-norm, respectively. For two sequences of positive numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n \preceq b_n$ or $b_n \succeq a_n$ if there exist a positive constant $c$ such that $a_n / b_n \leq c$. We write $a_n \asymp b_n$ if and only if $a_n \preceq b_n$ and $b_n \preceq a_n$ 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 \(\{\psi_{j}(\cdot)\}_{j=1}^{\infty}\), and truncate the expansion to the first \(d_{j}\) (to be specified in Section 5) terms:

\[
X_{tj}(\cdot) = \sum_{l=1}^{\infty} \eta_{tjl} \psi_{j}(\cdot) \approx \eta_{tj}^{j} \psi_{j}(\cdot) , \quad j \in [p],
\]

where \(\eta_{tjl} = \langle X_{tj}, \psi_{j} \rangle\), \(\eta_{tj} = (\eta_{tj1}, \ldots, \eta_{tjd_{j}})^{T} \in \mathbb{R}^{d_{j}}\) and \(\psi_{j}(\cdot) = \{\psi_{j1}(\cdot), \ldots, \psi_{jd_{j}}(\cdot)\}^{T}\). 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

\[
K_{jj}(u, v) = \sum_{h=1}^{L} \int_{\mathcal{U}} \Sigma_{h,jh}^{X}(u, z) \Sigma_{h,jh}^{X}(v, z) \, dz ,
\]

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

\[
K_{jj}(u, v) = \sum_{h=1}^{L} \int_{\mathcal{U}} \Sigma_{h,jh}^{W}(u, z) \Sigma_{h,jh}^{W}(v, z) \, dz ,
\]

since \(\Sigma_{h,jh}^{X}(u, z) = \Sigma_{h,jh}^{W}(u, z)\) for all \((u, z) \in \mathcal{U}^{2}\) and \(h \neq 0\). This paves the way to estimate \(K_{jj}\), and therefore also \(\psi_{j}(\cdot)\), 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 \(\Sigma_{h,jh}^{W}\) 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 \(H\). Hence the estimation is not sensitive to the choice of \(L\). In practice, we choose small \(L\) such as \(1 \leq L \leq 5\), as the most significant autocorrelations typically occur at small lags.

In the standard Karhunen–Loève expansion, \(\{\psi_{j}(\cdot)\}_{j=1}^{\infty}\) is deduced from the spectral decomposition of \(\Sigma_{0,jj}^{X}\). Since

\[
\Sigma_{0,jj}^{X}(u, v) = \Sigma_{0,jj}^{W}(u, v) - \Sigma_{0,jj}^{e}(u, v) ,
\]

some strong assumptions have to be imposed to eliminate the impact of \(\Sigma_{0,jj}^{e}(u, v)\) in order to obtain consistent estimates for \(\psi_{j}(\cdot)\). 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 \(\infty\). Note that the dimension reduction via FPCA can also be performed based on the spectral decomposition of \(\Sigma_{0,jj}^{W}\) instead of \(\Sigma_{0,jj}^{X}\), as any basis could be used for expanding the data. However, because of \(\Sigma_{0,jj}^{W} = \Sigma_{0,jj}^{X} + \Sigma_{0,jj}^{e}\), using \(\Sigma_{0,jj}^{W}\) 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 \(\Sigma_{0,jk}^{X}(u, v) = \Sigma_{0,jk}^{W}(u, v) - \Sigma_{0,jk}^{e}(u, v)\) and hence is inappropriate under model \([1]\).
With the available observations \( \{W_t(\cdot)\}_{t \in [n]} \), a natural estimator for \( K_{jj} \) in (5) is defined as

\[
\hat{K}_{jj}(u, v) = \sum_{h=1}^{L} \int dt \hat{\Sigma}_h^{W}(u, z) \hat{\Sigma}_h^{W}(v, z) dz = \frac{1}{(n - L)^2} \sum_{h=1}^{L} \sum_{t, s=0}^{n} W_{t-h}(u) W_{s-h}(v) \langle W_{tj}, W_{sj} \rangle ,
\]

(6)

where

\[
\hat{\Sigma}_h^{W}(u, v) = \frac{1}{n-h} \sum_{t=h+1}^{n} W_{t-h}(u) W_t(v)^T = \left\{ \hat{\Sigma}_j^{W}(u, v) \right\}_{j, k} \in [p] , \quad (u, v) \in \mathcal{U}^2 , \quad h \geq 0 .
\]

(7)

Performing the spectral decomposition

\[
\hat{K}_{jj}(u, v) = \sum_{l=1}^{\infty} \hat{\lambda}_l \hat{\psi}_l(u) \hat{\psi}_l(v) ,
\]

(8)

where \( \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots > 0 \) are the eigenvalues, and \( \hat{\psi}_1(\cdot), \hat{\psi}_2(\cdot), \cdots \) are the corresponding eigenfunctions.

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

2.2 Rates in elementwise \( \ell_\infty \)-norm

To characterize the effect of serial dependence on the relevant estimated terms, we will use the functional stability measure of \( \{W_t(\cdot)\}_{t \in \mathbb{Z}} \) (Guo and Qiao, 2022).

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

\[
\mathcal{M}^W = 2\pi \cdot \text{ess sup}_{\theta \in [-\pi, \pi], \Phi \in \mathcal{H}_p^0} \frac{\langle \Phi, f_\theta^W(\Phi) \rangle}{\langle \Phi, \Sigma_0^W(\Phi) \rangle} < \infty ,
\]

(9)

where \( \mathcal{H}_p^0 = \{\Phi \in \mathcal{H}_p : \langle \Phi, \Sigma_0^W(\Phi) \rangle \in (0, \infty)\} \).

The quantity \( \mathcal{M}^W \) in (9) is expressed proportional to functional Rayleigh quotients of \( f_\theta^W \) relative to \( \Sigma_0^W \). Hence it can more precisely capture the effect of small decaying eigenvalues of \( \Sigma_0^W \) 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 \( \{W_t(\cdot)\}_{t \in \mathbb{Z}} \), i.e. \( \{(W_{tj}(\cdot) : j \in J)^T\}_{t \in \mathbb{Z}} \) for \( J \subset [p] \) with cardinality \( |J| \leq k \), by

\[
\mathcal{M}_k^W = 2\pi \cdot \text{ess sup}_{\theta \in [-\pi, \pi], \|\Phi\|_0 \leq k, \Phi \in \mathcal{H}_p^0} \frac{\langle \Phi, f_\theta^W(\Phi) \rangle}{\langle \Phi, \Sigma_0^W(\Phi) \rangle} , \quad k \in [p] .
\]

(10)
Under Condition 1 it is easy to verify that $\mathcal{M}_k^W \leq \mathcal{M}^W < \infty$.

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 $\Sigma_0 : \mathbb{H} \rightarrow \mathbb{H}$ be a covariance operator. Then $Z_t(\cdot)$ is a sub-Gaussian process if there exists a constant $c > 0$ such that $\mathbb{E}(e^{\langle x,Z_t(\cdot) \rangle}) \leq e^{c^2 \langle x,\Sigma_0(x) \rangle/2}$ for all $x \in \mathbb{H}$.

**Condition 2.** (i) $\{W_t(\cdot)\}_{t \in \mathbb{Z}}$ is a sequence of multivariate functional linear processes with sub-Gaussian errors, namely sub-Gaussian functional linear processes, $W_t(\cdot) = \sum_{l=0}^{\infty} B_l(\varepsilon_{t-l})$ for any $t \in \mathbb{Z}$, where $B_l = (B_{l,jk})_{p \times p}$ with each $B_{l,jk} \in \mathbb{S}$, $\varepsilon_t(\cdot) = \{\varepsilon_{t1}(\cdot), \ldots, \varepsilon_{tp}(\cdot)\}^T \in \mathbb{H}^p$ and the components in $\{\varepsilon_t(\cdot)\}_{t \in \mathbb{Z}}$ are independent sub-Gaussian processes satisfying Definition 1; (ii) The coefficient functions satisfy Definition 1(ii) ensures the stationarity of $\{W_t(\cdot)\}_{t \in \mathbb{Z}}$ and, together with Condition 2(ii), implies that $\omega^W_0 = \max_{j \in [p]} \int_{\mathbb{R}} \Sigma^\varepsilon_{0,jj}(u,u) \, du = O(1)$ (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 frameworks.

**Condition 3.** (i) For each $j \in [p]$, $\lambda_{j1} > \lambda_{j2} > \cdots > 0$, and there exist some positive constants $c_0$ and $\alpha > 1$ such that $\lambda_{jl} - \lambda_{j(l+1)} \geq c_0 l^{-\alpha-1}$ for $l \geq 1$; (ii) For each $j \in [p]$, the linear space spanned by $\{\nu_{jl}(\cdot)\}_{l=1}^{\infty}$ (i.e. eigenfunctions of $\Sigma_{0,jj}$) 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 $\alpha$ 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 $\alpha$ across $j$, but this condition can be relaxed by allowing $\alpha$ 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 $\ell_\infty$-norm.

**Theorem 1.** Let Conditions 1, 2, 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

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

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 $h \geq 1$ be fixed. For $n \geq d^{2\alpha+2}(\mathcal{M}_1^W)^2 \log p$, there exist some positive constants $c_3$ and $c_4$ independent of $(n, p, d)$ such that
\begin{equation}
\max_{j, k \in [p], \ell, m \in [d]} \frac{|\sigma_{jklm}^{(h)} - \sigma_{jklm}^{(h)}|}{(l \vee m)^{\alpha+1}} \leq \mathcal{M}_1^W \sqrt{\frac{\log p}{n}}
\end{equation}
holds with probability greater than $1 - c_3p^{-c_4}$, where $\mathcal{M}_1^W$ is defined in (10).

Remark 2.1. (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 $j, k \in [p], \ell, m \in [d_j]$, and the corresponding right-sides remain the same.

(ii) Compared with the normalized deviation bounds under FPCA framework established in Guo and Qiao (2022), we obtain slower rates in (11) and (12) for decaying eigenvalues. Note that $\{\nu_{jl}(\cdot)\}_{l \geq 1}$ provides the unique basis with respect to which $X_{lj}(\cdot)$ can be expressed as Karhunen–Loève expansion with uncorrelated coefficients. It gives the most rapidly convergent representation of $X_{lj}(\cdot)$ in the $L_2$ sense. By comparison, the expansion of $X_{lj}(\cdot)$ through $\{\psi_{jl}(\cdot)\}_{l \geq 1}$ 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 $\{\tilde{\eta}_{(t-h)k} : h \in [L], k \in [p]\}$ as 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:
\begin{equation}
Y_t = \sum_{j=1}^p \int_{\mathcal{U}} X_{lj}(u)\beta_{0j}(u) \, du + \varepsilon_t, \quad t \in [n],
\end{equation}
where $\{X_{lj}(\cdot)\}_{t \in [n], j \in [p]}$ satisfy model (1), $\{\varepsilon_t\}_{t \in [n]}$ are i.i.d. and mean-zero random errors, and $\{\nu_{jl}(\cdot)\}$ and $\{\varepsilon_t\}$ are independent. Given observations $\{(W_t(\cdot), Y_t)\}_{t \in [n]}$, our goal is to estimate $p$ functional coefficients $\beta_0(\cdot) = (\beta_{01}(\cdot), \ldots, \beta_{0p}(\cdot))^\top$. To guarantee a feasible solution under high-dimensional scaling, we assume that $\beta_{0j}(\cdot)$ is functional $s$-sparse, i.e. $s$ components in $\beta_{0j}(\cdot)$ are nonzero with $s \ll p$.

Resulting from the truncated expansion of $X_{lj}(\cdot)$ via (4) in Step 1, (13) can be rewritten as
\begin{equation}
Y_t = \sum_{j=1}^p \eta_{lj}^\top b_{0j} + r_t + \varepsilon_t,
\end{equation}
where $b_{0j} = \int_{\mathcal{U}} \psi_j(u)\beta_{0j}(u) \, du \in \mathbb{R}^{d_j}$ and $r_t = \sum_{j=1}^p \sum_{l=d_j+1}^\infty \eta_{jll}^\top \psi_{jl}(\beta_{0j})$ is the truncation error. Given some prescribed positive integer $L$, in Step 2, we choose $\{\tilde{\eta}_{(t-h)k} : h \in [L], k \in [p]\}$ as vector-valued...
instrumental variables. Then $\mathbf{b}_0 = (\mathbf{b}_{01}^T, \ldots, \mathbf{b}_{0p}^T)^T \in \mathbb{R}^{\sum_{j=1}^{d_j}}$ can be identified by the following moment equations:

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

where $\mathbf{g}_{hk}(\mathbf{b}_0) = \mathbb{E}\{\eta_{(t-h)k}Y_t\} - \sum_{j=1}^{p} \mathbb{E}\{\eta_{(t-h)k}\eta_{i,j}^{T} \mathbf{b}_{0j}\}$ and the bias term $\mathbf{R}_{hk} = -\mathbb{E}\{\eta_{(t-h)k}\varepsilon_t\}$.

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

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

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

### 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, $\mathbf{\theta}_0 = (\mathbf{\theta}_1^T, \ldots, \mathbf{\theta}_p^T)^T \in \mathbb{R}^{\sum_{j=1}^{d_j} \times \bar{d}}$ with each $\mathbf{\theta}_{0j} \in \mathbb{R}^{d_j \times \bar{d}}$. For SFLR with a scalar response, $\bar{d} = 1$. Given some prescribed positive integer $L$ and $q = pL$ target moment functions $\mathbf{\theta} \mapsto \mathbf{g}_i(\mathbf{\theta})$ mapping $\mathbf{\theta} \in \mathbb{R}^{\sum_{j=1}^{d_j} \times \bar{d}}$ to $\mathbf{g}_i(\mathbf{\theta}) \in \mathbb{R}^{d_j \times \bar{d}}$ with $i = (h-1)p + k$ and $k \in [p]$ for $h \in [L]$, where both $p$ and $q$ are large, we assume that $\mathbf{\theta}_0$ can be identified by the following moment equations:

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

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 $\mathbf{\theta}_0$ based on empirical mappings $\mathbf{\theta} \mapsto \hat{\mathbf{g}}_i(\mathbf{\theta})$ of $\mathbf{\theta} \mapsto \mathbf{g}_i(\mathbf{\theta})$ for $i \in [q]$. See Sections 3.1 and 4 for detailed expressions of $\mathbf{g}_i(\cdot)$ and $\hat{\mathbf{g}}_i(\cdot)$ in some exemplified models.

It follows from (16) that

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

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

$$
\hat{\mathbf{\theta}} = \arg\min_{\mathbf{\theta}} \sum_{j=1}^{p} \|\mathbf{\theta}_j\|_F \quad \text{subject to} \quad \max_{i \in [q]} \|\hat{\mathbf{g}}_i(\mathbf{\theta})\|_F \leq \gamma_n,
$$

where $\gamma_n \geq 0$ is a regularization parameter. The group information is encoded in the objective function, which forces the elements of $\hat{\mathbf{\theta}}_j$ to either all be zero or nonzero, thus producing the block sparsity in $\hat{\mathbf{\theta}}$. 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 $\|\cdot\|_F$. 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, 4.2 and 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 $B = (B_{ij})_{i \in [N_1], j \in [N_2]} \in \mathbb{R}^{N_1 \times m_1 \times N_2 \times m_2}$ with the $(i, j)$-th block $B_{ij} \in \mathbb{R}^{m_1 \times m_2}$, let $\|B\|_{\max}^{(m_1, m_2)} = \max_{i \in [N_1], j \in [N_2]} \|B_{ij}\|_F$, and $\|B\|_1^{(m_1, m_2)} = \sum_{i=1}^{N_1} \|B_i\|_F$ when $N_2 = 1$. To simplify notation in this section and theoretical analysis in Section 3, 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 $g(\theta) = \{g_1(\theta)^T, \ldots, g_q(\theta)^T\}^T$ and $R = (R_1^T, \ldots, R_q^T)^T \in \mathbb{R}^{qd \times \hat{d}}$. We focus on the case of which the moment function $\theta \rightarrow g(\theta)$ mapping from $\mathbb{R}^{pd \times \hat{d}}$ to $\mathbb{R}^{qd \times \hat{d}}$ is linear with respect to $\theta$ in the form of $g(\theta) = G\theta + g(0)$ for some $G \in \mathbb{R}^{qd \times pd}$. This together with (16) implies that

$$G\theta_0 + g(0) + R = 0,$$

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

$$\hat{\theta} = \arg\min_{\theta} \|\theta\|_1^{(d, \hat{d})} \text{ subject to } \|\hat{g}(\theta)\|_{\max}^{(d, \hat{d})} \leq \gamma_n,$$

where $\hat{g}(\theta) = \hat{G}\theta + \hat{g}(0)$ is the empirical version of $g(\theta)$. It is worth noting that $\theta_0$ is block $s$-sparse with support $S = \{j \in [p] : \|\theta_{0j}\|_F \neq 0\}$ and its cardinality $s = |S|$.

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

**Condition 4.** (i) There exist $\epsilon_{n1}, \delta_{n1} > 0$ such that $\|\hat{G} - G\|_{\max}^{(d, \hat{d})} \lor \|\hat{g}(0) - g(0)\|_{\max}^{(d, \hat{d})} \leq \epsilon_{n1}$ with probability at least $1 - \delta_{n1}$; (ii) There exists $\epsilon_2 > 0$ such that $\|R\|_{\max}^{(d, \hat{d})} \leq \epsilon_2$; (iii) There exists $\delta_{n2} > 0$ such that $\|\hat{g}(\theta_0)\|_{\max}^{(d, \hat{d})} \leq \gamma_n$ with probability at least $1 - \delta_{n2}$.

Conditions (i) and (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 (i) for SFLR, FFLR and VFAR. With further imposed smoothness conditions on coefficient functions, Condition (ii) can also be verified. Condition (iii) indicates that $\theta_0$ is feasible in the optimization problem (20) with high probability, in which case a solution $\hat{\theta}$ of (20) exists and satisfies $\|\hat{\theta}\|_1^{(d, \hat{d})} \leq \|\theta_0\|_1^{(d, \hat{d})}$. The non-block version of such property typically plays a crucial role to tackle high-dimensional models in the literature.

Let $\delta = \theta - \theta_0$. We define a block $\ell_1$-sensitivity coefficient

$$\kappa(\theta_0) = \inf_{T:|T| \leq s} \inf_{\delta \in C_T: \|\delta\|_1^{(d, \hat{d})} > 0} \frac{\|G\delta\|_{\max}^{(d, \hat{d})}}{\|\delta\|_1^{(d, \hat{d})}}, \tag{21}$$

where $C_T = \{\delta \in \mathbb{R}^{pd \times \hat{d}} : \|\delta_T\|_1^{(d, \hat{d})} \leq \|\delta_T\|_1^{(d, \hat{d})}\}$ for $T \subset [p]$. Provided that $\hat{\delta} = \hat{\theta} - \theta_0 \in C_S$ under Condition (iii) as justified in Lemma 1 in Appendix B, the lower bound of $\kappa(\theta_0)$ is useful to establish the error bound for $\|\hat{\delta}\|_1^{(d, \hat{d})}$. See also Gautier and Rose (2019) for non-block $\ell_q$-sensitivity quantities to handle high-dimensional instruments. We then need Condition 5 below to determine such lower bound. Note that $G$ can be divided into $q \times p$ blocks of the size $d \times d$. Let $G_{j,M}$ be the submatrix of $G$ consisting
of all the \((j, k)\)-blocks with \(j \in J \subseteq [q]\) and \(k \in M \subseteq [p]\). For an integer \(m \geq s\), let

\[
\sigma_{\text{min}}(m, G) = \min_{|M| \leq m} \max_{|J| \leq m} \sigma_{\text{min}}(G_{J,M}) \quad \text{and} \quad \sigma_{\text{max}}(m, G) = \max_{|M| \leq m} \max_{|J| \leq m} \sigma_{\text{max}}(G_{J,M}),
\]

where \(\sigma_{\text{min}}(G_{J,M})\) and \(\sigma_{\text{max}}(G_{J,M})\) are the smallest and largest singular values of \(G_{J,M}\).

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

In Condition 3, the quantity \(\mu\) serves as a key factor to determine the lower bound of \(\kappa(\theta_0)\), which is justified in Lemma 4 in Appendix B. When \(\mu\) 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 \(\|\theta_0\|_{1,\bar{d}} \leq K\) for some \(K > 0\) and the regularization parameter \(\gamma_n \leq (K + 1)\epsilon_{n1} + \epsilon_2\), then with probability at least \(1 - (\delta_{n1} + \delta_{n2})\), the block RMD estimator \(\hat{\theta}\) satisfies

\[
\|\hat{\theta} - \theta_0\|_{1,\bar{d}} \leq s\mu^{-2}\{(K + 1)\epsilon_{n1} + \epsilon_2\}.
\]

**Remark 3.1.** (i) The error bound in (22) has the familiar variance-bias tradeoff as commonly considered in nonparametrics 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, \(G\) is formed by \(\{\sigma^{(h)}_{jklm} : j, k \in [p], l, m \in [d], h \in [L]\}\) with the components \(\sigma^{(h)}_{jklm}\) satisfying \(\sigma^{(h)}_{jklm} \leq \{E(\eta^2_{(l-h)ji})\}^{1/2}\{E(\eta^2_{tkmn})\}^{1/2} = \lambda^{1/2}_{j} \lambda^{1/2}_{km} \to 0\) as \(l, m \to \infty\). Consider a general cross-covariance matrix \(G = E(xy^T) \in \mathbb{R}^{qd \times pd}\) with entries decaying to zero as \(d \to \infty\), where \(x = (x_1, \ldots, x_{pd})^T\) with \(E(x) = 0\) and \(y = (y_1, \ldots, y_{pd})^T\) with \(E(y) = 0\), it is more sensible to impose Condition 5 on its normalized version \(\bar{G} = D_x GD_y\) instead of \(G\) itself, where \(D_x = \text{diag}\{\text{Var}(x_1)^{-1/2}, \ldots, \text{Var}(x_{pd})^{-1/2}\}\) and \(D_y = \text{diag}\{\text{Var}(y_1)^{-1/2}, \ldots, \text{Var}(y_{pd})^{-1/2}\}\). For three exemplified models, \(D_x\) and \(D_y\) are formed by \(\{\lambda^{-1/2}_{jl} : j \in [p], l \in [d]\}\).

Remark 3.1(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 \(\bar{G}\), then with probability at least \(1 - (\delta_{n1} + \delta_{n2})\), the block RMD estimator \(\hat{\theta}\) satisfies

\[
\|\hat{\theta} - \theta_0\|_{1,\bar{d}} \leq s\mu^{-2}\|D_x\|_{\text{max}}\|D_y\|_{\text{max}}\{(K + 1)\epsilon_{n1} + \epsilon_2\}.
\]

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 \( \{W_{ij}(\cdot)\}_{t \in [n]} \) for each \( j \in [p] \). According to Section 3.1 and following the optimization framework in [18], we then develop the block RMD estimator \( \hat{b} \) as a solution to the constrained optimization problem:

\[
\hat{b} = \arg \min_b \sum_{j=1}^{p} \| b_j \|_2 \text{ subject to } \max_{k \in [p], h \in [L]} \| \hat{g}_{hk}(b) \|_2 \leq \gamma_n,
\]

where \( \gamma_n \geq 0 \) is a regularization parameter and \( \hat{g}_{hk}(b) \) is defined in (15). Given that the recovery of functional sparsity in \( \beta_0(\cdot) \) is equivalent to estimating the block sparsity in \( b_0 \), in Step 3, we estimate functional sparse coefficients by

\[
\hat{\beta}_j(\cdot) = \hat{\psi}_j(\cdot)^T \hat{b}_j, \quad j \in [p].
\]

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 \( g = (g_{11}^T, \ldots, g_{1p}^T, \ldots, g_{L1}^T, \ldots, g_{Lp}^T)^T \), \( R = (R_{11}^T, \ldots, R_{1p}^T, \ldots, R_{L1}^T, \ldots, R_{Lp}^T)^T \) and \( G = (G_{ij}) \in \mathbb{R}^{pL \times pd} \) whose \((i, j)\)-th block is \( G_{ij} = E\{\eta_{(t-h)k} \eta_{ij}^T\} \in \mathbb{R}^{d \times d} \) with \( i = (h-1)p + k \) and \( k \in [p] \) for \( h \in [L] \). Applying Theorem 2 and Proposition 3 in Appendix A on \( \hat{G} \) and \( \hat{g}(0) \), respectively, we can verify Condition 4(i) with the choice of \( \epsilon_{n1} = \mathcal{M}_{W,Y}d^{\alpha+2}(n^{-1} \log p)^{1/2} \), where \( \mathcal{M}_{W,Y} \) is specified in Proposition 3. Before presenting the main theorem, we list the regularity conditions below.

**Condition 6.** (i) For each \( j \in S = \{ j \in [p] : \| \beta_0(j) \| \neq 0 \} \), \( \beta_0(j)(\cdot) = \sum_{l=1}^{\infty} a_{jl} \psi_{jl}(\cdot) \) and there exists some positive constant \( \tau > \alpha + 1/2 \) such that \( |a_{jl}| \leq l^{-\tau} \) for \( l \geq 1 \); (ii) Let \( \tilde{G} = (\tilde{G}_{ij}) \) be the normalized version of \( G = (G_{ij}) \) by replacing each \( G_{ij} \) by \( \tilde{G}_{ij} = E\{D_k \eta_{(t-h)k} \eta_{ij}^T D_j\} \), \( i = (h-1)p + k \), \( k \in [p] \) for \( h \in [L] \) and \( j \in [p] \), where \( D_j = \text{diag}(\lambda_{j1}^{-1/2}, \ldots, \lambda_{jd}^{-1/2}) \). There exist universal constants \( c_6 > 0 \) and \( \mu > 0 \) such that \( \sigma_{\max}(m, \tilde{G}) \geq c_6 \) and \( \sigma_{\min}(m, \tilde{G})/\sigma_{\max}(m, \tilde{G}) \geq \mu \) for \( m = 16s/\mu^2 \).

Condition 6(i) restricts each component in \( \{\beta_0(j)(\cdot) : j \in S\} \) based on its expansion through basis \{\psi_{jl}(\cdot)\}_{l \geq 1}. The parameter \( \tau \) determines the decay rate of basis coefficients and hence controls the level of smoothness with large values yielding smoother functions in \{\beta_0(j)(\cdot) : j \in S\}. See similar conditions in [Hall and Horowitz (2007)] and [Kong et al. (2016)]. Noting that components of \( G \) decay to zero as \( d \) grows to infinity, we impose Condition 6(ii) on \( \tilde{G} \), 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 \( \hat{\beta}(\cdot) = (\hat{\beta}_1(\cdot), \ldots, \hat{\beta}_p(\cdot))^T \) under functional \( \ell_1 \) 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. If the regularization parameter \( \gamma_n \gg s\{d^{\alpha+2} M_{W,Y}(n^{-1} \log p)^{1/2} + d^{-\tau+1/2}\} \), then the estimate \( \hat{\beta}(\cdot) \) satisfies

\[
\sum_{j=1}^{p} \| \hat{\beta}_j - \beta_0(j) \|_2 = O_p\left( s^2 \left( d^{2\alpha+2} M_{W,Y} \left( \frac{\log p}{n} + d^{1-\tau+1/2} \right) \right) \right).
\]

**Remark 4.1.** (i) The rate of convergence in (25) is governed by both dimensionality parameters \( (n, p, s) \) and internal parameters \( (M_{W,Y}, d, \alpha, \tau, \mu) \). Typically, the rate is better when \( \tau, \mu \) are large and \( M_{W,Y}, \alpha \) positive.
are small. To balance the variance and bias terms in \((25)\) for the optimal estimation, we can choose the optimal truncated dimension \(d \approx (M^2_{y,v}n^{-1} \log p)^{-1/(2r+2\alpha+3)}\).

(ii) Note that our convergence analysis relies on \((12)\) rather than the normalized deviation bounds in Guo and Qiao (2022), the rate in \((25)\) is slightly slower than that in Fang et al. (2022) by a multiplicative factor \(d^{3/2}\). For univariate functional linear regression, we similarly observe a slower rate for the autocovariance-based generalized methods-of-moments estimator \((\text{Chen et al. } 2022)\) compared to the covariance-based least squares estimator \((\text{Hall and Horowitz } 2007)\).

4.2 High-dimensional FFLR

Consider high-dimensional FFLR in the form of

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

where \(\{X_t(\cdot)\}_{t \in [n]}\) 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 \(\{(W_t(u), Y_t(v)) : (u, v) \in U \times V, t \in [n]\}\), we target to estimate \(\beta_0 = (\beta_{01}(\cdot, \cdot), \ldots, \beta_{0p}(\cdot, \cdot))^T\) under a functional sparsity constraint when \(p\) is large. Specifically, we assume \(\beta_0\) is functional \(s\)-sparse with support \(S = \{j \in [p] : \|\beta_{0j}\|_S \neq 0\}\) 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 \(d\), i.e. \(Y_t(\cdot) \approx \zeta_t^T \phi(\cdot)\), where \(\zeta_t = (\zeta_{t1}, \ldots, \zeta_{td})^T\) and \(\phi(\cdot) = (\phi_1(\cdot), \ldots, \phi_d(\cdot))^T\). Note that we can relax the independence assumption for \(\{\varepsilon_t(\cdot)\}_{t \in [n]}\) and model observed responses via \(\tilde{Y}_t(\cdot) = Y_t(\cdot) + e_T^T(\cdot)\), where \(Y_t(\cdot)\) and \(e_T(\cdot)\) 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

\[
\zeta_t^T = \sum_{j=1}^{P} \eta_{tj}^T \phi_{0j} + r_t^T + \varepsilon_t^T, \tag{27}
\]

where \(\phi_{0j} = \int_{U \times V} \phi_j(u)\beta_{0j}(u,v) \phi(v)^T \, du \in \mathbb{R}^{d_j \times d}\) and \(r_t = (r_{t1}, \ldots, r_{td})^T\) is the truncation error with \(r_{tm} = \sum_{j=1}^{P} \sum_{l=d_j+1}^{d} \eta_{lj} \phi_{l}\beta_{0j}, \phi_m)\) for \(m \in [d]\). Let \(B_0 = (B_{01}, \ldots, B_{0p})^T \in \mathbb{R}^{p \times d_j \times d}\). We choose \(\{\eta_{(l-h)k} : h \in [L], k \in [p]\}\) as vector-valued instrumental variables, which are assumed to be uncorrelated with the random error \(\varepsilon_t\) in \((27)\). Within the framework of \((16)\), we assume that \(B_0\) is the unique solution to the following moment equations:

\[
0 = \mathbb{E}\{\eta_{(l-h)k} \varepsilon_t^T\} = g_{hk}(B_0) + R_{hk}, \quad h \in [L], \ k \in [p], \tag{28}
\]

where \(g_{hk}(B_0) = \mathbb{E}\{\eta_{(l-h)k} \zeta_t^T\} - \sum_{j=1}^{P} \mathbb{E}\{\eta_{(l-h)k} \eta_{tj}^T \phi_{0j}\}\) and \(R_{hk} = -\mathbb{E}\{\eta_{(l-h)k} r_t^T\}\).

Given the recovery equivalence between functional sparsity in \(\beta_0\) and the block sparsity in \(B_0\), we aim
to estimate the block sparse matrix $B_0$ using the empirical versions $B \mapsto \hat{g}_{hk}(B)$ for $h \in [L]$ and $k \in [p]$, 

$$
\hat{g}_{hk}(B) = \frac{1}{n-h} \sum_{t=h+1}^{n} \hat{\eta}_{(t-h)k} \hat{\zeta}_{t}^T - \frac{1}{n-h} \sum_{t=h+1}^{n} \sum_{j=1}^{p} \hat{\eta}_{(t-h)k} \hat{\eta}_{tj} B_j,
$$

where $\hat{\zeta}_t = (\hat{\zeta}_{t1}, \ldots, \hat{\zeta}_{td})^T$ with $\hat{\zeta}_{tm} = \langle Y_t, \hat{\phi}_m \rangle$ for $m \in [d]$ and $(\hat{\eta}_{tj})_{t \in [n], j \in [p]}$ are obtained in Step 1. In Step 2, according to (18), we formulate the block RMD estimator of eigengaps of the covariance function of $\alpha, \epsilon$ with the choice of $\gamma_0$ where $\gamma_0 = \langle \alpha, \epsilon \rangle$ such that $|A_{jm}| \leq (l + m)^{-\tau + 1/2}$ for $l, m \geq 1$.

$$
\hat{\beta}_j(u, v) = \hat{\psi}_j(u)^T \hat{B}_j \hat{\phi}(v), \ (u, v) \in U \times V, j \in [p],
$$

where $\{\hat{\psi}_j(u)\}_{j \in [p]}$ and $\hat{\phi}(v) = (\hat{\phi}_1(v), \ldots, \hat{\phi}_d(v))^T$ 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 $G$ and $\hat{g}(0)$ to verify Condition 4(i) with the choice of $\epsilon_{n1} = \mathcal{M}_{W,Y} d^{\alpha_1 + 1/2}$, where $\mathcal{M}_{W,Y}$ is specified in Proposition 2. In a similar fashion to $\alpha$, the parameter $\alpha$ as specified in Condition 4(i) 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 $j \in S$, $\beta_{0j}(u, v) = \sum_{l,m=1}^{r} a_{jm} \psi_{jl}(u) \phi_m(v)$ and there exists some positive constant $\tau > \alpha \land \alpha_1 + 1/2$ such that $|a_{jm}| \leq (l + m)^{-\tau + 1/2}$ for $l, m \geq 1$.

We are now ready to present the convergence rate of the FFLR estimate $\hat{\beta}(\cdot, \cdot) = (\hat{\beta}_1(\cdot, \cdot), \ldots, \hat{\beta}_p(\cdot, \cdot))^T$ under functional $\ell_1$ 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 $d = \tilde{d}$. If the regularization parameter $\gamma_n \asymp s \{d^{\alpha_1 + 1/2} \mathcal{M}_{W,Y} (n^{-1} \log p)^{1/2} + d^{\alpha + 1/2}\}$, then the estimate $\hat{\beta}(\cdot, \cdot)$ satisfies

$$
\sum_{j=1}^{p} \|\hat{\beta}_j - \beta_{0j}\|_S = O_p \left\{ \mu^{-2} s^2 \left( d^{\alpha_1 + 1/2} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}} + d^{\alpha + 1/2} \right) \right\},
$$

(30)

**Remark 4.2.** (i) With the same expression of $G$ for both SFLR and FFLR, Condition 6(ii) is required in both Theorems 4 and 5. Note we can further remove the assumption $d = \tilde{d}$, 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 $d^{\alpha/2}$. 

14
4.3 High-dimensional VFAR

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

$$
X_t(v) = \sum_{h' = 1}^{H} \int_{\mathcal{U}} A_0^{(h')}(u, v) X_{t-h'}(u) \, du + \varepsilon_t(v), \quad t = H + 1, \ldots, n, \tag{31}
$$

where $\{X_t(\cdot)\}$ satisfy model [1], the errors $\varepsilon_t(\cdot) = (\varepsilon_{t1}(\cdot), \ldots, \varepsilon_{tp}(\cdot))^T$ are i.i.d. sampled from a $p$-vector of mean-zero random functions, independent of $X_{t-1}(\cdot), X_{t-2}(\cdot), \ldots$, and $A_0^{(h')} = \{A_{0, j,j'}^{(h')} (\cdot, \cdot)\}_{j,j' \in [p]}$ is the unknown functional transition matrix at lag $h'$. In the special case $H = 1$ with $A_0 = A_0^{(1)}$, Theorem 3.1 of Bosq (2000) ensures the stationarity of $\{X_t(\cdot)\}$ if there exists an integer $l_0$ such that $\sup_{|f| \leq 1} \|A_0^{(h')} (f)\| < 1$ for $f \in \mathbb{H}^p$. According to Guo and Qiao (2022), 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 $\{W_t(\cdot)\}_{t \in [n]}$, we assume $\{A_0^{(h')}\}_{h' \in [H]}$ is rowwise functional $s$-sparse with $s = \max_{j \in [p]} s_j < p$. To be specific, for the $j$-th row of components in $\{A_0^{(h')}\}$, we denote the set of nonzero functions by $S_j = \{(j', h') \in [p] \times [H] : \|A_{0, j,j'}^{(h')}\|_S \neq 0\}$ 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

$$
\eta_{tj}^\top = \sum_{h' = 1}^{H} \sum_{j' = 1}^{p} \eta_{(t-h')j'}^\top \Omega_{0, j,j'}^{(h')} + r_{tj}^\top + \varepsilon_{tj}^\top, \quad j \in [p], \tag{32}
$$

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

$$
0 = \mathbb{E}\{\eta_{(t-H-h)j} \varepsilon_{tj}^\top\} = g_{j, hk}(\Omega_{0,j}) + R_{j, hk}, \quad h \in [L], k \in [p], \tag{33}
$$

where $g_{j, hk}(\Omega_{0,j}) = \mathbb{E}\{\eta_{(t-H-h)j} \eta_{tj}^\top\} - \sum_{h' = 1}^{H} \sum_{j' = 1}^{p}\mathbb{E}\{\eta_{(t-H-h)j} \eta_{tj}^\top \Omega_{0,j,j'}^{(h')}\}$ and $R_{j, hk} = -\mathbb{E}\{\eta_{(t-H-h)j} r_{tj}^\top\}$.

Given that estimating the functional sparsity in the $j$-th row of $\{A_0^{(h')}\}_{h' \in [H]}$ is equivalent to estimating the block sparsity in $\Omega_{0,j}$ for each $j$, our goal is to estimate the block sparse matrix $\Omega_{0,j}$ using the empirical versions $\hat{\Omega}_j \to \hat{g}_{j, hh}(\Omega_j)$ for $h \in [L]$ and $k \in [p]$, where

$$
\hat{g}_{j, hh}(\Omega_j) = \frac{1}{n - H - h} \sum_{t = H + h + 1}^{n} \hat{\eta}_{(t-H-h)k} \hat{\eta}_{tj} - \frac{1}{n - H - h} \sum_{t = H + h + 1}^{n} \sum_{h' = 1}^{H} \sum_{j' = 1}^{p} \hat{\eta}_{(t-H-h)k} \hat{\eta}_{(t-h')j}^{\top} \Omega_{jj'}^{(h')},
$$

and $\{\hat{\eta}_{tj}\}_{t \in [n], j \in [p]}$ are obtained in Step 1. Step 2 follows [18] to formulate the block RMD estimator $\hat{\Omega}_j$.
Theorem 6. Suppose that Conditions (i) For each \( j, j' \in [p], h' \in [H] \), we then give the following regularity conditions.

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

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

We next present convergence analysis of \( \hat{A}_{jj'}^{(h')} (\cdot, \cdot) \). 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:

\[
g_j(\Omega_{0j}) + R_j = G_j\Omega_{0j} + g_j(0) + R_j = 0,
\]

where \( g_j = (g_{j,11}^{T}, \ldots, g_{j,1p}^{T}, \ldots, g_{j,L1}^{T}, \ldots, g_{j,Lp}^{T})^{T} \). For each \( j \in [p] \), let \( \hat{G}_j = (\hat{G}_{j,ii'} = \mathbb{E}D_{j,ii'}) \) be the normalized version of \( G_j = (G_{j,ii'}) \) by replacing each \( G_{j,ii'} \) by \( \hat{G}_{j,ii'} = \mathbb{E}D_{j,ii'} = \mathbb{E}(\mathbf{D}_{j,ii'}) \) for \( i = (h - 1)p + k, \ i' = (h' - 1)p + j' \) with \( k, j' \in [p], h \in [L] \) and \( h' \in [H] \).}

Conditions 8. (i) For each \( j \in [p] \) and \((j', h') \in S_j, A_{0,jj'}^{(h')} (u, v) = \sum_{l,m=1}^{\infty} a_{l,m}^{(h')} \psi_{j,l}(u)\psi_{m}(v) \) and there exists some constant \( \gamma > \alpha + 1/2 \) such that \( |A_{l,m}^{(h')}| \leq (l + m)^{-\gamma - 1/2} \) for \( l, m \geq 1 \); (ii) For each \( j \in [p] \), let \( \hat{G}_j = (\hat{G}_{j,ii'}) \) be the normalized version of \( G_j = (G_{j,ii'}) \) by replacing each \( G_{j,ii'} \) by \( \hat{G}_{j,ii'} = \mathbb{E}D_{j,ii'} = \mathbb{E}(\mathbf{D}_{j,ii'}) \) for \( i = (h - 1)p + k, i' = (h' - 1)p + j' \) with \( k, j' \in [p], h \in [L] \) and \( h' \in [H] \), where \( \mathbf{D}_{j,ii'} = \text{diag}(\lambda_{1}^{j,ii'}, \ldots, \lambda_{1}^{j,ii'}) \). There exist universal constants \( \bar{c}_j > 0 \) and \( \mu_j > 0 \) such that

\[
\sigma_{\max}(\hat{G}_j) \geq \bar{c}_j \text{ and } \sigma_{\min}(\hat{G}_j)/\sigma_{\max}(\hat{G}_j) \geq \mu_j \text{ for } m = 16s_j/\mu_j^2.
\]

We finally establish the convergence rate of the VFAR estimate \( \hat{A}_{jj'}^{(h')} (\cdot, \cdot) \) in the sense of functional matrix \( \ell_{\infty} \) norm as follows.

Theorem 6. Suppose that Conditions (i) \( H \) and (ii) \( \mathbb{E} \) hold. If the regularization parameters satisfy \( \gamma_m \geq s_j \{d^{\alpha+2}M_1 W(n^{-1} \log p)^{1/2} + d^{-\gamma+1/2} \} \) for \( j \in [p] \) and \( \mu = \min_{j \in [p]} \mu_j \), the estimate \( \hat{A}_{jj'}^{(h')} (\cdot, \cdot) \) satisfies

\[
\max_{j \in [p]} \sum_{j' = 1}^{H} \sum_{h' = 1}^{H} \|A_{jj'}^{(h')} - A_{0,jj'}^{(h')}\|_{\infty} = O_P \left\{ \mu^{-2} s^2 \left( d^{\alpha+2} M_1 W \sqrt{\frac{\log p}{n}} + d^{-\gamma+1/2} \right) \right\}.
\]

Remark 4.3. Similar to Remarks 4.1(ii) and 4.2(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 (2022) 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 \( X_{tj}(u) = \sum_{l=1}^{25} \eta_{jl} \psi_l(u) = \eta_{jt}^T \psi(u) \) with \( \eta_{jt} = (\eta_{j1}, \ldots, \eta_{j25})^T \) and \( \psi(\cdot) = \{\psi_1(\cdot), \ldots, \psi_{25}(\cdot)\}^T \) for \( t \in [n], j \in [p] \) and \( u \in \mathcal{U} = [0, 1] \), where \( \{\psi_l(u)\}_{1 \leq l \leq 25} \) 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 \( \eta_{jt} = (\eta_{jt1}, \ldots, \eta_{jt25})^T \in \mathbb{R}^{25p} \) is generated from a stationary vector autoregressive (VAR) model, \( \eta_t = \Omega \eta_{t-1} + \epsilon_t \), with block transition matrix \( \Omega = (\Omega_{jk})_{j,k \in [p]} \in \mathbb{R}^{25p \times 25p} \) and \( \epsilon_t = (\epsilon_{t1}, \ldots, \epsilon_{t25})^T \), whose components are sampled independently according to \( \epsilon_{tj} \sim \mathcal{N}(0, 0.7 - 0.1j) \) for \( j = 1, \ldots, 5 \) and \( \mathcal{N}(0, j^{-2}) \) for \( j = 6, \ldots, 25 \). Therefore, \( X_t(\cdot) \) follows a VFAR(1) model \( X_t(\cdot) = \sum_{l=1}^{5} z_{tl} \psi_l(\cdot) \) and \( x_t(\cdot) = \psi(\cdot)^T \Omega x_{t-1}(\cdot) + \epsilon_t \), where \( \epsilon_t(v) = \psi(v)^T \epsilon_t \) and autocoefficient functions satisfy \( A_{jk}(u, v) = \psi(v)^T \Omega_{jk} \psi(u) \) 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 \( N \)-noise curves \( e_{tj}(u) = \sum_{l=1}^{5} z_{tlj} \psi_l(u) \) and \( z_{tj} = (z_{tj1}, \ldots, z_{tj25})^T \) and \( \{z_{tj}\}_{j \in [n]} \) are sampled independently from multivariate normal distribution with mean zero and covariance matrix \( \text{diag}(1, 0.8, 0.3, 1.5, 1.6) \). For each of the three models, the data is generated as follows.

**VFAR:** We generate block sparse \( \Omega \) with 5% or 10% nonzero blocks for \( p = 80 \) or \( p = 40 \), respectively. Specifically, for the \( j \)-th block row, we set the diagonal block \( \Omega_{jj} = \text{diag}(0.60, 0.59, 0.58, 0.3, 0.2, 6^{-2}, \ldots, 25^{-2}) \) and randomly choose one off-diagonal block being \( 0.4 \Omega_{jj} \) and two off-diagonal blocks being \( 0.1 \Omega_{jj} \). Such block sparse design on \( \Omega \) 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 \( \{X_t(\cdot)\}_{t \in [n]} \) 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 \( \epsilon_t \)'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 \( b_{jl} = (-1)^{l-2} \) 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 \( \epsilon_t(v) = \sum_{m=1}^{5} g_{tm} \psi_m(v) \) with \( g_{tm} \)'s being independent \( \mathcal{N}(0, 1) \) variables. For \( j \in S \), we generate \( \beta_j(u, v) = \sum_{l=1}^{25} b_{jlm} \psi_l(u) \psi_m(v) \) for \( (u, v) \in \mathcal{U} \times \mathcal{V} \), where components in \( \{b_{jlm}\}_{1 \leq l, m \leq 3} \) 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. \( \{ \hat{B}_j(\gamma) \}_{j \in [p]} \) (or \( \{ \hat{B}_j^{(\gamma)} \}_{j \in [p]} \) as a function of \( \gamma \)) for SFLR (or FFLR) and \( \{ \hat{\Omega}_{jk}^{(\gamma)} \} \) as a function of \( \gamma_{nj} \) for VFAR, calculate the squared error between observed and fitted values on the validation set, i.e. \( \sum_{t=1}^{n} [Y_t - \sum_{j=1}^{p} \{ \hat{B}_j(\gamma) \}^T \hat{\eta}_{tj}]^2 \) for SFLR, \( \sum_{t=1}^{n} \| \hat{c}_t - \sum_{j=1}^{p} \{ \hat{B}_j^{(\gamma)} \}^T \hat{\eta}_{tj} \|^2 \) for FFLR and \( \sum_{t=1}^{n} \| \hat{\eta}_{tj} - \sum_{k=1}^{p} \{ \hat{\Omega}_{jk}(\gamma) \}^T \hat{\eta}_{(t-1)k} \|^2 \) 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 \( \{ W_{i,j}(\cdot) \}_{i \in [n]} \) 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 \( \textit{Kong et al. 2016} \) and the multivariate response \( \textit{Fang et al. 2022} \), respectively and the VFAR estimation is converted to the VAR estimation \( \textit{Guo and Qiao 2022} \). 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 \( \textit{Guo and Qiao 2022} \), which converges faster than the commonly adopted block coordinate descent algorithm \( \textit{Fan et al. 2015} \). 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. \( \| \hat{A} - A \|_F / \| A \|_F \) for VFAR, \( \sum_{j=1}^{p} \| \hat{\beta}_j - \beta_0 \|^2 / (\sum_{j=1}^{p} \| \beta_0 \|^2)^{1/2} \) for SFLR and \( \sum_{j=1}^{p} \| \hat{\beta}_j - \beta_0 \|^2 / (\sum_{j=1}^{p} \| \beta_0 \|^2)^{1/2} \) for FFLR. We ran each simulation 100 times. Figure 1 displays boxplots of relative estimation errors for three models. Several conclusions can be drawn from Figure 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 \( \textit{Horváth et al. 2014} \), 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 \( j \)-th stock at the \( k \)-th minute after the opening time on the \( t \)-th trading day. We work with mildly smoothed CIDRs obtained by expanding the data with respect to a 45-dimensional B-spline basis. The
Figure 1: The boxplots of relative estimation errors for (a) VFAR, (b) SFLR and (c) FFLR.
Table 1: MSPEs up to different current times, $N = 300, 315, 330, 345, 360, 370$ and $380$ minutes, 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  | $u \leq 300$ | $u \leq 315$ | $u \leq 330$ | $u \leq 345$ | $u \leq 360$ | $u \leq 370$ | $u \leq 380$ |
|---------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 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    | $8.832$       | $8.832$       | $8.832$       | $8.832$       | $8.832$       | $8.832$       | $8.832$       |

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.

Our target is to predict the intraday return of the S&P 100 index based on observed CIDR trajectories of component stocks, $W_{ij}(u), u \in \mathcal{U} = [0, N]$ up to time $N$, where, e.g., $N = 360$ corresponds to 30 minutes 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

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

(35)

where $Y_t$ is the intraday return of the S&P 100 index on the $t$-th trading day, $X_{ij}(\cdot)$ and $e_{ij}(\cdot)$ represent the signal and noise components in $W_{ij}(\cdot)$, 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 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 Section A, all technical proofs in Section B and list of S&P 100 stocks in Section C.

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 \( \{W_t(\cdot)\}_{t \in \mathbb{Z}} \) and \( \tilde{p} \)-vector of mean-zero functional time series (or scalar time series) \( \{Y_t(\cdot)\}_{t \in \mathbb{Z}} \) (or \( \{Z_t\}_{t \in \mathbb{Z}} \)). Define \( \Sigma_{h}^{W,Y}(u,v) = \text{Cov}\{W_{t-h}(u),Y_t(v)\} \) and \( \Sigma_h^{W,Z}(u) = \text{Cov}\{W_{t-h}(u),Z_t\} \) for \( h \in \mathbb{Z} \) and \( (u,v) \in U \times V \).

**Condition 9.** (i) For \( \{W_t(\cdot)\}_{t \in \mathbb{Z}} \) and \( \{Y_t(\cdot)\}_{t \in \mathbb{Z}} \), the cross-spectral density function \( f_{\theta}^{W,Y} = (2\pi)^{-1} \sum_{h \in \mathbb{Z}} \Sigma_{h}^{W,Y} e^{-i\theta h} \) for \( \theta \in [-\pi,\pi] \) exists and the functional cross-spectral stability measure defined in (A.1) is finite, i.e.

\[
\mathcal{M}^{W,Y} = 2\pi \cdot \text{ess sup}_{\theta \in [-\pi,\pi]} \frac{|\langle \Phi_1, f_{\theta}^{W,Y}(\Phi_2) \rangle|}{\sqrt{\langle \Phi_1, \Sigma_0^{W} (\Phi_1) \rangle \sqrt{\langle \Phi_2, \Sigma_0^{Y} (\Phi_2) \rangle}}} < \infty, \tag{A.1}
\]

where \( \mathbb{H}_0^p = \{ \Phi \in \mathbb{H}^p : \langle \Phi, \Sigma_0^{W} (\Phi) \rangle \in (0,\infty) \} \) and \( \mathbb{H}_0^y = \{ \Phi \in \mathbb{H}^y : \langle \Phi, \Sigma_0^{Y} (\Phi) \rangle \in (0,\infty) \} \).

(ii) For \( \{W_t(\cdot)\}_{t \in \mathbb{Z}} \) and \( \{Z_t\}_{t \in \mathbb{Z}} \), the cross-spectral density function \( f_{\theta}^{W,Z} = (2\pi)^{-1} \sum_{h \in \mathbb{Z}} \Sigma_{h}^{W,Z} e^{-i\theta h} \) for \( \theta \in [-\pi,\pi] \) exists and the functional cross-spectral stability measure defined in (A.2) is finite, i.e.

\[
\mathcal{M}^{W,Z} = 2\pi \cdot \text{ess sup}_{\theta \in [-\pi,\pi]} \frac{|\langle \Phi, f_{\theta}^{W,Z}(\nu) \rangle|}{\sqrt{\langle \Phi, \Sigma_0^{W} (\Phi) \rangle \sqrt{\langle \nu, \Sigma_0^{Z} \nu \rangle}}} < \infty, \tag{A.2}
\]

where \( \mathbb{R}_0^p = \{ \nu \in \mathbb{R}^2 : \nu^T \Sigma_0^p \nu \in (0,\infty) \} \).

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

\[
\mathcal{M}^{Z} = 2\pi \cdot \text{ess sup}_{\theta \in [-\pi,\pi]} \frac{\nu^T f_{\theta}^{Z} \nu}{\sqrt{\nu^T \Sigma_0^p \nu}},
\]

which is equivalent to that proposed in Basu and Michailidis (2015). The stability measure of all \( k \)-dimensional subsets of \( \{Z_t\} \), i.e., \( \mathcal{M}^{Z}_k \) 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 \( \{(\theta_{km},\hat{\phi}_{km})\}_{m \geq 1} \) are the pairs of eigenvalues and eigenfunctions of \( \Sigma_{Y,kk} \). Let \( \{(\hat{\theta}_{km},\hat{\phi}_{km})\}_{m \geq 1} \) be the estimated eigenpairs of \( \hat{\Sigma}_{Y,kk} \) 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 \( \ell_2 \)-norm on how \( \hat{\sigma}_{h,jklm}^{W,Y} = (n-h)^{-1} \sum_{t=h+1}^{n} \hat{\eta}_{(t-h)j} \hat{\zeta}_{tkm} \) concentrates around \( \sigma_{h,jklm}^{W,Y} = \text{Cov}\{\eta_{(t-h)j}, \zeta_{tkm}\} \), 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 $\tilde{d} > 1$ such that $\theta_{kn} - \theta_{k(n+1)} \geq \tilde{c}m^{-\tilde{d}^{-1}}$ for $m \geq 1$; (ii) $\max_{k \in [\tilde{p}]} \sum_{m=1}^{\infty} \theta_{km} = O(1)$.

Proposition 2. Suppose that Conditions \[ \text{(i) and (ii)} \] hold, $\{Y_t(\cdot)\}_{t \in [n]}$ is sub-Gaussian functional linear process and $h$ is fixed. Let $\tilde{d}$ and $\tilde{d}$ be positive integers possibly depending on $(n,p,\tilde{p})$ and $M_{W,Y} = M_{W,Y}^{W_1} + M_{W,Y}^{W_1,1}$. For $n \geq (d^{2\alpha+2} + d^{2\tilde{d}+2})(M_{W,Y}^{W})^{2} \log(p\tilde{p})$, there exist some positive constants $c_7$ and $c_8$ independent of $(n,p,\tilde{p},d,\tilde{d})$ such that

$$\max_{j \in [\tilde{p}], k \in [\tilde{p}], l \in [d], m \in [d]} \left| \frac{\sigma_{k,jl} - \sigma_{k,lj}}{l^{\alpha+1} + m^{\tilde{d}+1}} \right| \leq M_{W,Y} \sqrt{\frac{\log(p\tilde{p})}{n}}$$

(A.3)

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

We next consider a mixed process scenario consisting of $\{W_t(\cdot)\}$ and $\{Z_t\}$ and establish the deviation bound on how $\hat{\theta}_{h,jkl} = (n-h)^{-1} \sum_{t=1}^{n} \tilde{W}_{h,jkl} Z_{tk}$ concentrates around $\hat{\theta}_{h,jkl} = \text{Cov}\{\eta_{(t-h)j}, Z_{tk}\}$, which is essential in deriving the convergence rate of the SFLR estimate in Section 4.1.

Proposition 3. Suppose that Conditions \[ \text{(i) and (ii)} \] hold, $\{Z_t\}_{t \in [n]}$ is sub-Gaussian linear process and $h$ is fixed. Let $\tilde{d}$ be a positive integer possibly depending on $(n,p,\tilde{p})$ and $M_{W,Z} = M_{W,Z}^{W_1} + M_{W,Z}^{W_1,1} + M_{W,Z}^{W_2}$. For $n \geq (M_{W,Z})^{2} \log(p\tilde{p})$, there exist some positive constants $c_9$ and $c_{10}$ independent of $(n,p,\tilde{p},d)$ such that

$$\max_{j \in [\tilde{p}], k \in [\tilde{p}], l \in [d]} \left| \frac{\tilde{\theta}_{h,jkl} - \tilde{\theta}_{h,jkl}}{l^{\alpha+1} + m^{\tilde{d}+1}} \right| \leq M_{W,Z} \sqrt{\frac{\log(p\tilde{p})}{n}},$$

(A.4)

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

B Technical proofs

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

B.1 Auxiliary lemmas

Lemma 1. Suppose that Condition \[ \text{(i)} \] holds. Then $\| \tilde{\theta}_{S} \|_{1}^{(d,d)} \leq \| \tilde{\theta}_{S} \|_{1}^{(d,d)}$ with probability at least $1 - \delta_{n2}$.

Proof. It follows from Condition \[ \text{(i)} \] and $\theta_{0,S} = 0$ by definition that with probability at least $1 - \delta_{n2}$, $\| \theta_{0,S} \|_{1}^{(d,d)} \leq \| \tilde{\theta}_{S} \|_{1}^{(d,d)}$, which implies that $\| \theta_{0,S} \|_{1}^{(d,d)} \geq \| \tilde{\theta}_{S} \|_{1}^{(d,d)}$, $\tilde{\theta}_{S} - \theta_{0,S} \|_{1}^{(d,d)}$. By cancelling $\| \theta_{0,S} \|_{1}^{(d,d)}$ on both sides above, we obtain $\| \tilde{\theta}_{S} - \theta_{0,S} \|_{1}^{(d,d)} \leq \| \tilde{\theta}_{S} - \theta_{0,S} \|_{1}^{(d,d)}$.

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

Proof. Let $v_j$ denotes the $j$-th row of $V^Tx$ for $j \in [r]$. Write $\sigma_r^2 \| x \|_{F}^2 \leq \| Ax \|_{F}^2 = \text{tr}(x^T A^T A x) = \text{tr}(x^T V \Lambda^2 V^T x) = (\sum_{j=1}^{r} \sigma_j^2 v_j^T v_j)^{1/2} \leq \sigma_1 \| x \|_{F}^2$, where, in the inequalities above, we have used $\| V^T x \|_{F} \leq \| x \|_{F}$ due to the orthonormality of $V$. 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 $\sigma_{\min}(m, G)$ and $\sigma_{\max}(m, G)$, respectively.

**Lemma 3.** It holds that

$$
\kappa(\theta_0) \geq \max_{m \geq s} \left\{ \frac{\sigma_{\min}(m)}{\sqrt{m}} - \frac{2\sigma_{\max}(m)}{\sqrt{m}} \sqrt{\frac{s}{m}} \right\} \frac{s^{-1/2}}{2(1 + 2\sqrt{s/m})}.
$$

**Proof.** Let $T \subset [p]$ and $\|\delta_T\|_1^{(d, \tilde{d})} \leq \|\delta_T\|_1^{(d, d)}$ by [21]. Let $T_1$ denote the largest $m$ components of $\{\|\delta_i\|_F\}_{i \in [p]}$, and $T_2$ be the subsequent $m$-largest, etc. Let $V_\mu = \text{diag}(\mu \otimes 1_d)$ where $\mu \in \mathbb{R}^q$ with $\sum_{i=1}^q I(|\mu_i|) \leq m$ and $1_2 = (1, \ldots, 1)^T \in \mathbb{R}^d$. We let $\|\mu\| = (\sum_{i=1}^q \mu_i^2)^{1/2}$ and $\|\mu\|_\infty = \max_{i \in [q]} |\mu_i|$. Then, we have

$$
\|G\delta\|_{\max}^{(d, \tilde{d})} = \max_{\mu} \left\| \frac{1}{\sqrt{m}} V_\mu G \delta \right\|_F \geq \left\| \frac{1}{\sqrt{m}} \|\mu\|_\infty V_\mu G \delta \right\|_F \\
\geq \left\| \frac{1}{\sqrt{m}} \|\mu\|_\infty V_{\mu G \cdot T_1} \delta_{T_1} \right\|_F - \sum_{j \geq 2} \left\| \frac{1}{\sqrt{m}} \|\mu\|_\infty V_\mu G \cdot T_j \delta_{T_j} \right\|_F,
$$

(B.1)

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

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

$$
\left\| \frac{1}{\sqrt{m}} \|\mu\|_\infty V_{\mu G \cdot T_1} \delta_{T_1} \right\|_F = \left\| \frac{1}{\sqrt{m}} G_{\tilde{J}_1, T_1} \delta_{T_1} \right\|_F \\
\geq \frac{\sigma_{\min}(G_{\tilde{J}_1, T_1})}{\sqrt{m}} \|\delta_{T_1}\|_F = \frac{1}{\sqrt{m}} \max_{|J| \leq m} \sigma_{\min}(G_{J, T_1}) \|\delta_{T_1}\|_F
$$

(B.2)

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

$$
\sum_{j \geq 2} \left\| \frac{1}{\sqrt{m}} \|\mu\|_\infty V_{\mu G \cdot T_j} \delta_{T_j} \right\|_F = \frac{1}{\sqrt{m}} \sum_{j \geq 2} \|G_{\tilde{J}_j, T_j} \delta_{T_j}\|_F \\
\leq \frac{1}{\sqrt{m}} \sum_{j \geq 2} \sigma_{\max}(G_{\tilde{J}_j, T_j}) \|\delta_{T_j}\|_F = \frac{1}{\sqrt{m}} \sum_{j \geq 2} \max_{|J| \leq m} \sigma_{\max}(G_{J, T_j}) \|\delta_{T_j}\|_F
$$

(B.3)

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

$$
\sum_{j \geq 2} \|\delta_{T_j}\|_F \leq \frac{1}{\sqrt{m}} \sum_{j \geq 1} \|\delta_{T_j}\|_1^{(d, \tilde{d})} \leq \|\delta_1\|_1^{(d, \tilde{d})} \sqrt{\frac{m}{\sqrt{m}}}.
$$

(B.4)
Combining (B.2), (B.3) and (B.4) yields

\[
\|G\delta\|_{\text{max}}^{(d,\tilde{d})} \geq \frac{\sigma_{\min}(m)}{\sqrt{m}}\|\delta_T\|_F - \frac{\sigma_{\max}(m)}{\sqrt{m}}\|\delta_1^{(d,\tilde{d})}\|_1/\sqrt{m} \\
\geq \frac{\sigma_{\min}(m)}{\sqrt{m}}\|\delta_T\|_F - \frac{\sigma_{\max}(m)}{\sqrt{m}}2\sqrt{\frac{s}{m}}\|\delta_T\|_F \\
= \left\{ \frac{\sigma_{\min}(m)}{\sqrt{m}} - 2\frac{\sigma_{\max}(m)}{\sqrt{m}}\sqrt{\frac{s}{m}}\|\delta_T\|_F \right\}\|\delta_T\|_F
\]

(B.5)

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

\[
\|\delta\|_F \leq \|\delta_T\|_F + \sum_{j \geq 2}\|\delta_{T_j}\|_F \leq \|\delta_T\|_F + 2\sqrt{s/m}\|\delta_T\|_F \leq (1 + 2\sqrt{s/m})\|\delta_T\|_F.
\]

(B.6)

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

\[
\|G\delta\|_{\text{max}}^{(d,\tilde{d})} \geq \left\{ \frac{\sigma_{\min}(m)}{\sqrt{m}} - 2\frac{\sigma_{\max}(m)}{\sqrt{m}}\sqrt{\frac{s}{m}}\|\delta_T\|_F \right\}\|\delta_T\|_F
\]

(B.7)

where the second inequality comes from \(\|\delta\|_F \geq \|\delta_T\|_F \geq \|\delta_T\|_1^{(d,\tilde{d})}/\sqrt{s} \geq \|\delta_1^{(d,\tilde{d})}/\sqrt{s}\). We complete our proof by (21) and dividing \(\|\delta_1^{(d,\tilde{d})}\) on both sides of (B.7).

\[\square\]

Lemma 4. Suppose that Condition [5] holds. Then there exists some positive constant \(c\) such that \(\kappa(\theta_0) \geq c\mu^2/(24s)\).

Proof. Applying Lemma [3] and choosing \(m = 16s/\mu^2\) yields that

\[
\kappa(\theta_0) \geq \max_{m \geq s} \frac{\sigma_{\max}(m, G)}{\sqrt{m}} \left\{ \frac{\sigma_{\min}(m, G)}{\sigma_{\max}(m, G)} - \frac{\mu}{2} \right\} \frac{s^{-1/2}}{2(1 + \mu/2)} \\
\geq \frac{c\mu}{4\sqrt{s}} \left( \mu - \frac{\mu}{2} \right) \left\{ 2\left( 1 + \frac{\mu}{2} \right) \right\}^{-1} s^{-1/2} \geq \frac{c\mu^2}{24s},
\]

which completes our proof.

\[\square\]

For each \(j \in [p]\), let \(\omega_{j1} \geq \omega_{j2} \geq \cdots \geq 0\) be the eigenvalues of \(\Sigma_{0,jj}^X\) with the corresponding eigenfunctions \(\nu_{j1}(\cdot), \nu_{j2}(\cdot), \ldots\). Similarly, let \(\{(\omega_{j1}^W, \nu_{j1}^W(\cdot))\}_{i=1}^{\infty}\) be the eigenpairs of \(\Sigma_{0,jj}^W\).

Lemma 5. Suppose that Condition [2] holds. Then we have \(\omega_0^W = \max_j \sum_{l=1}^{\infty} \omega_{jl}^W = O(1)\).

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

\[\square\]

Lemma 6. For \(p \times p\) lag-h autocovariance function of \(\{W_t(\cdot)\}, \{\Sigma_{h,jk}(\cdot, \cdot)\}_{j,k \in [p]}\), we have \(\|\Sigma_{h,jk}\|_S \leq \omega_0^W\)

\(\|\Sigma_{h,jk}(\psi_{km})\|_S \leq \omega_{km}^W(\omega_0^W)^{1/2}\) for \(m \geq 1\).

Proof. This lemma follows directly from Lemma 8 of [Guo and Qiao (2022)] and hence the proof is omitted here.

\[\square\]
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, 2022], we can obtain that for \( h \geq 1 \)

\[
\mathbb{P}\left\{ \left| \frac{\langle \Phi_1, \Sigma_h^W - \Sigma_0^W \rangle (\Phi_2) \rangle}{\langle \Phi_1, \Sigma_0^W (\Phi_1) \rangle + \langle \Phi_2, \Sigma_0^W (\Phi_2) \rangle} \right| > 2M_2^W \delta \right\} \leq 8 \exp\{-cn \min(\delta^2, \delta)\}.
\]  
(B.8)

For each \( j \in [p] \), consider the spectral decomposition \( \Sigma_{h,jj}(u,v) = \sum_{l=1}^L \omega_{jl}^W \nu_{jl}^W (u) \nu_{jl}^W (v) \) and \( \omega_0 = \max_j \sum_{l=1}^L \omega_{jl}^W = O(1) \), implied from Lemma 3. For each \( (j,k,l,m) \), choosing \( \Phi_1 = (0, \ldots, 0, (\omega_{jl}^W)^{-1/2} \nu_{jl}^W, 0, \ldots, 0)^T \) and \( \Phi_2 = (0, \ldots, 0, (\omega_{km}^{W})^{-1/2} \nu_{km}^W, 0, \ldots, 0)^T \) on (B.8) and following the same procedure to prove Theorem 2 of [Guo and Qiao, 2022] with the choice of suitable constant \( \tilde{c} \), we can obtain that

\[
\mathbb{P}\{ |\hat{\Sigma}_{h,jk}^W - \Sigma_{h,jk}^W|_S > M_1^W \delta \} \leq 8 \exp\{-\tilde{c}n \min(\delta^2, \delta)\}.
\]  
(B.9)

By (B.7), (B.8) and Cauchy–Schwarz inequality, we have \( \|\tilde{K}_j - K_j\|_S^2 \leq 2L \sum_{h=1}^L \|\hat{\Sigma}_{h,jj}^W - \Sigma_{h,jj}^W\|_S^2 \|\hat{\Sigma}_{h,jj}^W\|_S^2 + L \sum_{h=1}^L \|\hat{\Sigma}_{h,jj}^W - \Sigma_{h,jj}^W\|_S^4 \). Let \( \Omega_{\omega,jj}^{(h)} = \{ \|\hat{\Sigma}_{h,jk}^W - \Sigma_{h,jk}^W\|_S \leq \omega_0 \} \) and \( \Omega_{\omega,jj}^{(h)} = \{ \|\hat{\Sigma}_{h,jk}^W - \Sigma_{h,jk}^W\|_S \leq M_1^W \delta \} \). On the event \( \Lambda_j = \Omega_{\omega,jj}^{(1)} \cap \cdots \cap \Omega_{\omega,jj}^{(L)} \cap \Omega_{\omega,jj}^{(L)} \cap \cdots \cap \Omega_{\omega,jj}^{(L)} \), it follows from the above results and Lemma 6 that

\[
\|\hat{K}_j - K_j\|_S \leq \sqrt{3}L \omega_0 M_1^W \delta.
\]  
(B.10)

Applying (B.9) and choosing \( \delta = (\Lambda_1^{-1} - \omega_0^2 \Omega_{\omega,jj}^{(1)} \cdots \Omega_{\omega,jj}^{(L)} \Omega_{\omega,jj}^{(L)} \cdots \Omega_{\omega,jj}^{(L)} ) \), we obtain

\[
\mathbb{P}\{ |\hat{K}_j - K_j|_S > M_1^W \delta \} \leq \tilde{c} \exp\{-cn \min(\delta^2, \delta)\} + \tilde{c} \exp\{-cn\}.
\]
(B.11)

For each \( j \in [p] \), it follows from Lemma 4.3 of [Bose, 2000] and Condition 3 with \( \min_{k=1}^t \{ \lambda_{jk} - \lambda_{j(k+1)} \} \geq \alpha_0 l^{-\alpha - 1} \) that

\[
\max_{l \in [d]} |\lambda_{jl} - \lambda_{jl}| \leq \|\hat{K}_j - K_j\|_S \quad \text{and} \quad \max_{l \in [d]} (\|\hat{\psi}_{jl} - \psi_{jl}\|)^{1/\alpha + 1} \leq 2\sqrt{2} \alpha_0^{-1} \|\hat{K}_j - K_j\|_S.
\]
(B.12)

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

\[
\mathbb{P}\left( \max_{j \in [p]} |\lambda_{jl} - \lambda_{jl}| > M_1^W \delta \right) \cup \mathbb{P}\left\{ \max_{j \in [p], l \in [d]} (\|\hat{\psi}_{jl} - \psi_{jl}\|)^{1/\alpha + 1} > 2\sqrt{2} \alpha_0^{-1} M_1^W \delta \right\}
\]

\[
\leq \tilde{c} \rho \exp\{-cn \min(\delta^2, \delta)\} + \tilde{c} \rho \exp\{-cn\}.
\]

Let \( \delta = \rho \sqrt{n^{-1} \log p} \leq 1 \). Choosing suitable positive constants \( \tilde{c} \) and \( \tilde{c} = \rho \alpha_0^{-4} - 1 \), we obtain that holds with probability greater than \( 1 - \tilde{c} \rho \alpha_0^{-4} \), 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

\[
\hat{\sigma}_{jkm}^{(h)} - \sigma_{jkm}^{(h)} = \langle \hat{\psi}_{jl}, \hat{\Sigma}_{h,jk}^W (\hat{\psi}_{km}) \rangle - \langle \psi_{jl}, \Sigma_{h,jk}^W (\psi_{km}) \rangle
\]

\[
= \langle (\hat{\psi}_{jl} - \psi_{jl}), \Sigma_{h,jk}^W (\psi_{km} - \psi_{km}) \rangle + \langle \psi_{jl}, (\hat{\Sigma}_{h,jk}^W - \Sigma_{h,jk}^W) (\psi_{km}) \rangle.
\]
On the event \( \tilde{\Omega}_{jk} = \Omega_{\omega,jk}^{(h)} \cap \Omega_{\omega,jk}^{(h)} \cap \Lambda_j \cap \Lambda_k \), it follows from Lemma \( \text{(B.10)} \) and \( \text{(B.12)} \), the orthonormality of \( \{ \hat{\psi}_l \}, \{ \hat{\psi}_m \} \) that \( \max_{l,m \in [d]} \{ |J_l|/(l \vee m)^{\alpha+1} \} \lesssim (M_1^W)^2 \delta^2, \) \( \max_{l,m \in [d]} \{ |J_2| \} \lesssim M_1^W \delta \), \( \max_{l,m \in [d]} \{ |J_3|/(l \vee m)^{\alpha+1} \} \lesssim M_1^W \delta \). Then \( \max_{l,m \in [d]} \{ \sum_{i=1}^4 |J_i|/(l \vee m)^{\alpha+1} \} \lesssim cM_1^W \delta + c\delta^{\alpha+1} (M_1^W)^2 \delta^2 \)\). Applying \( \text{(B.9)} \) and choosing \( \delta = (M_1^W)^{-1} \omega_0 \) for \( \Omega_{\omega,jk}^{(h)}, \Omega_{\omega,jj}^{(h)}, \ldots, \Omega_{\omega,j,j}^{(h)} \) yields that \( \mathbb{P}(\tilde{\Omega}_{jk}^c) \leq (16L + 8) \exp\{-cn \min(\delta^2, \delta)\} + (16L + 8) \exp\{-cn \min((M_1^W)^{-2} \delta^2, (M_1^W)^{-1} \omega_0)\} \).

Combining the above results, choosing suitable positive constants \( \tilde{c}, \tilde{c}, \tilde{c} \), and applying the union bound of probability yields that

\[
\mathbb{P} \left\{ \max_{j,k \in [p], l,m \in [d]} \left| \frac{\hat{\sigma}_{\delta,jkm}^{(h)} - \sigma_{\delta,jkm}^{(h)}}{(l \vee m)^{\alpha+1}} \right| > M_1^W \delta + c\delta^{\alpha+1} (M_1^W)^2 \delta^2 \right\} \leq \tilde{c}p^2 \left[ \exp\{-\tilde{c}n \min(\delta^2, \delta)\} + \exp(-\tilde{c}n) \right].
\]

(B.13)

Choosing \( \delta = \rho_1 \sqrt{n^{-1}} \log p \leq 1 \) and \( 1 + c\delta^{\alpha+1} M_1^W \delta \leq \rho_2 \) for some positive constants \( \rho_1, \rho_2 \), which can be achieved for sufficiently large \( n \gtrsim d^{2(\alpha+2)(M_1^W)^2} \log p \), it follows from \( \text{(B.13)} \) that there exist positive constants \( \tilde{c}, \tilde{c}, \tilde{c} \) such that, with probability greater than \( 1 - \tilde{c}p^{-\tilde{c}} \),

\[
\max_{j,k \in [p], l,m \in [d]} \left| \frac{\hat{\sigma}_{\delta,jkm}^{(h)} - \sigma_{\delta,jkm}^{(h)}}{(l \vee m)^{\alpha+1}} \right| \leq \rho_1 \rho_2 M_1^W \sqrt{\log p \over n},
\]

which completes the proof of Theorem \( \text{2} \). \( \square \)

### B.4 Proof of Proposition \( \text{2} \)

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

\[
\hat{\sigma}_{h,jkm}^{W,Y} - \sigma_{h,jkm}^{W,Y} = \left\{ [\hat{\psi}_l - \psi_l, (\hat{\psi}_m - \psi_m)] \right\} + \left\{ [\hat{\psi}_l - \psi_l, (\hat{\psi}_m - \psi_m)] \right\} + \left\{ [\hat{\psi}_l - \psi_l, (\hat{\psi}_m - \psi_m)] \right\} + \left\{ [\hat{\psi}_l - \psi_l, (\hat{\psi}_m - \psi_m)] \right\} =: I_1 + I_2 + I_3 + I_4.
\]

Let \( \Omega_{0,kk}^{\delta} = \{ \| \hat{\Sigma}_{0,kk} - \Sigma_{0,kk}^{W,Y} \| \leq M_1^{W,Y} \delta \} \) and \( \Omega_{h,jk}^{W,Y} = \{ \| \hat{\Sigma}_{h,jk}^{W,Y} - \Sigma_{h,jk}^{W,Y} \| \leq M_{W,Y} \delta \} \). On the event \( \Lambda_j \cap \Omega_{0,kk}^{\delta} \cap \Lambda_k \), it follows from \( \| \hat{\Sigma}_{h,jk}^{W,Y}, \phi_{km} \| \lesssim \omega_0^{1/2} \rho_{km}^{1/2} \) and \( \| \hat{\Sigma}_{h,jk}^{W,Y}, \phi_{km} \| \lesssim \omega_0^{1/2} \rho_{km}^{1/2} \), derived by the similar techniques to prove Lemma \( \text{(B.10)} \) and \( \text{(B.12)} \), the orthonormality of \( \{ \hat{\psi}_l \}, \{ \hat{\psi}_m \} \) and Condition \( \text{(B.10)} \) that \( \max_{l \in [d], m \in [d]} \{ |I_1|/(l^{2(\alpha+1)} + m^{2(\alpha+1)}) \} \lesssim (M_1^W)^2 \delta^2 + (M_1^Y)^2 \delta^2 \), \( \max_{l \in [d], m \in [d]} \{ |I_2| \} \lesssim M_{W,Y} \delta + M_{W,Y} \delta \) \( \lesssim M_{W,Y} \delta + M_{W,Y} \delta \). Combining the above results and \( M_{W,Y} = M_1^W + M_1^Y + M_{W,Y}^{W,Y} \) yields that \( \max_{l \in [d], m \in [d]} \{ \sum_{i=1}^4 |I_i|/(l^{2(\alpha+1)} + m^{2(\alpha+1)}) \} \lesssim cM_{W,Y} \delta + \tilde{c}(d^{\alpha+1} + d^{\alpha+1})(M_{W,Y})^2 \delta^2 \). Following the same developments to prove \( \text{(B.13)} \), we apply \( \text{(B.11)} \), Theorem \( \text{2} \), Lemma 24 of \( \text{Fang et al. (2022)} \) and the
union bound of probability, choose suitable positive constants \(\tilde{c}, \tilde{c}, \tilde{c}\) and hence obtain that

\[
\Pr\left\{ \max_{j \in [p], k \in [\tilde{p}], l \in [d], m \in [d]} \frac{|\hat{\theta}_{hljkm} - \theta_{hljkm}|}{\ell_{\alpha + 1} \sqrt{m^{\alpha + 1}}} > \mathcal{M}_{W,Y} \delta + \tilde{c}(d^{\alpha + 1} \cdot d^{\alpha + 1}) \mathcal{M}_{W,Y}^2 \delta^2 \right\} 
\leq \tilde{c} p\bar{p} \left[ \exp\{-\tilde{c}n \min(\delta^2, \delta)\} + \exp(-\tilde{c}n) \right].
\]

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

\[
\max_{j \in [p], k \in [\tilde{p}], l \in [d], m \in [d]} \frac{|\hat{\theta}_{hljkm} - \theta_{hljkm}|}{\ell_{\alpha + 1} \sqrt{m^{\alpha + 1}}} \leq \rho_3 \rho_4 \mathcal{M}_{W,Y} \sqrt{\frac{\log(p\bar{p})}{n}},
\]

which completes the proof of Proposition 2.

### B.5 Proof of Proposition 3

For each \((h, j, k, l)\), we write \(\hat{\theta}_{hljkl} = \hat{\theta}_{hjkl} = (\hat{\psi}_{jkl} - \psi_{jkl}), (\hat{\Sigma}_{hj}, \Sigma_{hj}), \hat{\psi}_{jkl} - \psi_{jkl}) + (\psi_{jkl} - (\hat{\Sigma}_{hj}, \Sigma_{hj})), \hat{\Sigma}_{hj} = \hat{\Sigma}_{hj}, \Sigma_{hj}) = T_1 + T_2 + T_3.\) Let \(\Omega_{hj}^{W,Z} = \{\|\hat{\Sigma}_{hj} - \Sigma_{hj}\|_F \leq \mathcal{M}_{W,Z} \delta\}.\) On the event \(\Lambda_{hj} \cap \Omega_{hj}^{W,Z}\), it follows from (B.10), (B.12), the orthonormality of \(\{\hat{\psi}_{jkl}\}\) and \(\|\hat{\Sigma}_{hj}\| \leq \omega_1/2 \|\omega_{hj}\|\) that \(\max_{l \in [d]} |T_1|/l^{\alpha + 1} \leq \mathcal{M}_{W,Z} \delta \mathcal{M}_{W,Z} \delta, \max_{l \in [d]} |T_2| \leq \mathcal{M}_{W,Z} \delta\) and \(\max_{l \in [d]} |T_3|/l^{\alpha + 1} \leq \mathcal{M}_{W,Z} \delta.\) Combining the above results and \(\mathcal{M}_{W,Z} = \mathcal{M}_{W}^2 + \mathcal{M}_{Z}^2 + \mathcal{M}_{1,1}^2 \) implies that \(\max_{l \in [d]} \sum_{k=1}^3 |T_i|/l^{\alpha + 1} \leq c \mathcal{M}_{W,Z} \delta + \tilde{c}(\mathcal{M}_{W,Z}^2)^2 \delta^2.\) 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 \(\tilde{c}, \tilde{c}, \tilde{c}\) and hence obtain that

\[
\Pr\left\{ \max_{j \in [p], k \in [\tilde{p}], l \in [d], m \in [d]} \frac{|\hat{\theta}_{hljkm} - \theta_{hljkm}|}{\ell_{\alpha + 1} \sqrt{m^{\alpha + 1}}} > \mathcal{M}_{W,Z} \delta + \tilde{c}(\mathcal{M}_{W,Z})^2 \delta^2 \right\} \leq \tilde{c} p\bar{p} \left[ \exp\{-\tilde{c}n \min(\delta^2, \delta)\} + \exp(-\tilde{c}n) \right].
\]

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

\[
\max_{j \in [p], k \in [\tilde{p}], l \in [d], m \in [d]} \frac{|\hat{\theta}_{hljkm} - \theta_{hljkm}|}{\ell_{\alpha + 1} \sqrt{m^{\alpha + 1}}} 
\leq \rho_5 \rho_6 \mathcal{M}_{W,Z} \sqrt{\frac{\log(p\bar{p})}{n}},
\]

which completes the proof of Proposition 3.

### B.6 Proof of Theorem 3

By \(g(\theta) = G\theta + g(0)\) and (19), we have \(g(\hat{\theta}) = \hat{G}\hat{\theta} + g(0), G\theta_0 + g(0) + R = 0\) and \(\hat{g}(\hat{\theta}) = \hat{G}\hat{\theta} + \hat{g}(0)\). Consider event \(A = \{\|\hat{G} - G\|_{(d,d)} \max \|\hat{g}(0) - g(0)\|_{(d,d)} \leq \epsilon_n\} \cap \{\|\hat{g}(\theta_0)\|_{(d,d)} \leq \gamma_n\}.\) By the union bound of probability and Conditions (i) and (iii), this event occurs with probability at least \(1 - \delta_{n1} - \delta_{n2}\). On
event $A$, we have

$$
\|G(\hat{\theta} - \theta_0)\|_{\max}^{(d,d)} \leq \|g(\hat{\theta})\|_{\max}^{(d,d)} + \|R\|_{\max}^{(d,d)}
\leq \|G - G\|_{\max}^{(d,d)} + \|\hat{g}(\hat{\theta}) - g(\theta)\|_{\max}^{(d,d)} + \|R\|_{\max}^{(d,d)}
\leq \max_i \{\|\hat{G} - G\|_{\max}^{(d,d)} + \|\hat{g}(0) - g(0)\|_{\max}^{(d,d)} + \|R\|_{\max}^{(d,d)}\}
\leq K\epsilon_n + \epsilon_n + 1 + \gamma_n + \epsilon_2,
$$

where, in the last two inequalities, we have used facts that $\max_i \{\|\hat{G} - G\|_{\max}^{(d,d)} = \max_{i\in[\theta]} \sum_{j=1}^p \|\hat{G} - G\|_{\max}^{(d,d)}\}$ by Condition $4(ii)$. On event $A$, choosing the set $T = S$ in (21) and applying Lemma 1 under Condition $4(iii)$ yields $\max_i \{\|\hat{G} - G\|_{\max}^{(d,d)}\}$ and hence $\hat{\theta} \in C_S$. Then by (21), (B.16) and Lemma 1 under Condition $5$, we have $\|\hat{G} - G\|_{\max}^{(d,d)} = \kappa(\theta_0)^{-1} \cdot \max_i \{\|\hat{G} - G\|_{\max}^{(d,d)}\} \leq \gamma_n \kappa(\theta_0)^{-1} \leq \max_i \{\|\hat{G} - G\|_{\max}^{(d,d)}\} \leq \gamma_n \kappa(\theta_0)^{-1} \leq \gamma_n \kappa(\theta_0)^{-1}$, which completes the proof. $
$
of \{\psi_{jl}\}, Cauchy–Schwartz inequality and Condition 6(i) that

\[
\{||R||_{\text{max}}^{(d,1)}\}^2 = \max_{k,h} \{E(\eta_{(t-h)km}^2)\} = \max_{k,h} \left\{ \sum_{m=1}^{d} \sum_{j=1}^{\infty} \eta_{(t-h)km}^2 \right\}^2 \\
\leq \max_{k,h} \left( \sum_{j=1}^{\infty} \sum_{l=d+1}^{\infty} \sqrt{E(\eta_{(t-h)km}^2)E(\eta_{(t-h)km}^2)} \right)^2 \\
\leq s^2 \max_{k,j} \sum_{m=1}^{d} \left( \sum_{l=d+1}^{\infty} \lambda_{km} \lambda_{kj} a_{jl} \right)^2 \\
\leq s^2 \max_{k,j} \sum_{m=1}^{d} \lambda_{km} \max_j \left\{ \sum_{l=d+1}^{\infty} \lambda_{kj} a_{jl} \right\} \leq \lambda_0^2 s^2 \sum_{j=1}^{\infty} l^{-2\tau} = O(s^2 d^{-2\tau+1}).
\]

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

\[
||R||_{\text{max}}^{(d,1)} \leq csd^{-\tau+1/2} = \epsilon_2.
\]  

(B.21)

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

\[
||b_0||_{1,1}^{(d,1)} = \sum_{j=1}^{d} \left( \sum_{l=d+1}^{\infty} a_{jl}^2 \right)^{1/2} \leq \max_{j=1}^{d} \left( \sum_{l=d+1}^{\infty} l^{-2\tau} \right)^{1/2} = O(s).
\]  

(B.22)

Finally, we verify Condition 4(iii) for SFLR. On event \(I_1 \cap I_2\), combining (B.20) (B.21) and (B.22) yields that

\[
||\tilde{g}(b_0) - g(b_0)||_{\text{max}}^{(d,1)} \leq ||\tilde{g}(b_0) - g(b_0)||_{\text{max}}^{(d,1)} + ||R||_{\text{max}}^{(d,1)} \\
\leq ||\tilde{G} - G||_{\text{max}}^{(d,1)} + ||\tilde{g}(0) - g(0)||_{\text{max}}^{(d,1)} + ||R||_{\text{max}}^{(d,1)} \\
\leq ||\tilde{G} - G||_{\text{max}}^{(d,1)} + ||\tilde{g}(0) - g(0)||_{\text{max}}^{(d,1)} + ||R||_{\text{max}}^{(d,1)} \\
\leq cs \left( d^{\alpha+2} M_{W,Y} \sqrt{\frac{\log p}{n}} + d^{-\tau+1/2} \right) = \gamma_n.
\]

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

\[
||\tilde{b} - b_0||_{1,1}^{(d,1)} = O_p \left\{ \mu^{-2} s^2 d^\alpha \left( d^{\alpha+2} M_{W,Y} \sqrt{\frac{\log p}{n}} + d^{-\tau+1/2} \right) \right\}.
\]  

(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 \(||R_j||^2 = \sum_{l=d+1}^{\infty} a_{jl}^2 \psi_{jl}^2 = \sum_{l=d+1}^{\infty} a_{jl}^2 \leq d^{-2\tau+1} \) for \(j \in \{\psi_{jl}\}\) under Condition 6(i), we have

\[
||\tilde{b}_j - b_{0j}|| = ||\tilde{\psi}_{jl}^T \tilde{b}_j - \psi_{jl}^T b_{0j} - R_j|| \leq \left( ||\tilde{\psi}_{jl} - \psi_{jl}|\tilde{b}_j - \psi_{jl}^T b_{0j}|| \right) + ||R_j|| \\
\leq d^{1/2} \max_{l=1}^{d} ||\tilde{\psi}_{jl} - \psi_{jl}|| \|\tilde{b}_j - b_{0j}\| + ||R_j|| + O(d^{-\tau+1/2}),
\]

which implies that \(||\tilde{b} - b_0|| \leq d^{1/2} \max_{j \in [p]} \|\tilde{\psi}_{jl} - \psi_{jl}|| \|\tilde{b}_j - b_{0j}\| + O(sd^{-\tau+1/2})\), where
the third term above is of a smaller order of the second term due to (B.23). By \(|\|\mathbf{b}_1^{(d,1)}\|_{1} \leq \|\mathbf{b} - \mathbf{b}_0\|_{1}^{(d,1)} + \|\mathbf{b}_0\|_{1}^{(d,1)} + \) (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

\[
I_3 = \left\{ \max_{k \in \{p, \ldots, r_p\}, k \in \{L, \ldots, h\}} \left| \frac{1}{n-h} \sum_{t=h+1}^{n} \hat{\eta}_{(t-h)km} \zeta_{tt} - \mathbb{E}\{\eta_{(t-h)km} \zeta_{tt}\} \right| \leq \tilde{c} d^{a \wedge \alpha+1} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}} \right\}
\]

for some sufficiently large \(\tilde{c}\). On event \(I_1 \cap I_3\), we have

\[
\|\mathbf{g}(0) - \mathbf{g}(0)\|_{\max}^{(d,\hat{d})} = \max_{k \in \{p, \ldots, r_p\}, k \in \{L, \ldots, h\}} \left| \frac{1}{n-h} \sum_{t=h+1}^{n} \hat{\eta}_{(t-h)km} \zeta_{tt} - \mathbb{E}\{\eta_{(t-h)km} \zeta_{tt}\} \right|_{F} \leq \tilde{c} d^{a \wedge \alpha+2} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}} .
\] (B.24)

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

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

We next verify Condition 4(ii) for FFLR. If follows from \(r_t = (r_{t1}, \ldots, r_{td})^T\) with each \(r_{tm'} = \sum_{j=1}^{p} \sum_{l=d+1}^{\infty} \eta_{lj} (\psi_{jl}, \beta_{0j}, \phi_{m'})\), orthonormality of \(\{\psi_{jl}\}, \{\phi_{m'}\}\), Cauchy–Schwartz inequality and Condition 7 that

\[
\left\{ \mathbb{E}\|\mathbf{R}\|_{\max}^{(d,\hat{d})} \right\}^{2} = \max_{k \in \{p, \ldots, r_p\}, k \in \{L, \ldots, h\}} \mathbb{E}\{\eta_{(t-h)km} \zeta_{tt}^T\} \|_{F}^{2} = \max_{k,h} \sum_{m=1}^{d} \sum_{m'=1}^{\hat{d}} \left\{ \mathbb{E}\{\eta_{(t-h)km} \zeta_{tt}^T\} \mathbb{E}\{\eta_{tl}^2 a_{jm'}\} \right\}^{2} \leq \max_{k,h} \sum_{m=1}^{d} \sum_{m'=1}^{\hat{d}} \left( \sum_{j \in S} \sum_{l=d+1}^{\infty} \mathbb{E}\{\eta_{(t-h)km} \zeta_{tt}^T\} \mathbb{E}\{\eta_{tl}^2 a_{jm'}\} \right)^{2} \leq s^2 \max_{k} \sum_{m=1}^{d} \sum_{m'=1}^{\hat{d}} \left( \sum_{l=d+1}^{\infty} \lambda_{km} \lambda_{jl}^{1/2} \lambda_{jm'}^{1/2} \right)^{2} \leq s^2 \max_{k} \sum_{m=1}^{d} \sum_{m'=1}^{\hat{d}} \left( \sum_{l=d+1}^{\infty} \lambda_{jl} \sum_{m'=1}^{\hat{d}} \sum_{l=d+1}^{\infty} \lambda_{jm'} \right)^{2} \leq \lambda_{0}^2 s^2 \sum_{m'=1}^{\hat{d}} \sum_{l=d+1}^{\infty} (l + m')^{-2\tau - 1} = O(s^2 d^{-2\tau + 1}) ,
\]

which implies that

\[
\|\mathbf{R}\|_{\max}^{(d,\hat{d})} \leq \tilde{c} s d^{-\tau + 1/2} = \epsilon_2 .
\] (B.26)

By the similar technique above and Condition 7

\[
\|\mathbf{B}_0\|_{1}^{(d,\hat{d})} = \sum_{j \in S} \left( \sum_{l=1}^{d} \sum_{m=1}^{\hat{d}} \eta_{jm'}^2 \right)^{1/2} \leq s \max_{j \in S} \left( \sum_{l=1}^{d} \sum_{m=1}^{\hat{d}} (l + m')^{-2\tau - 1} \right)^{1/2} = O(s) .
\] (B.27)
Finally, we verify Condition \( \Psi \) (iii) for FFLR. On event \( I_1 \cap I_3 \), combining (B.25), (B.26) and (B.27), and applying the similar techniques for SFLR, we have
\[
\| \hat{g}(B_0) \|_{\text{max}}^{(d,d)} \leq \| \hat{G} - G \|_{\text{max}}^{(d,d)} \| B_0 \|_{\text{max}}^{(d,d)} + \| \hat{g}(0) - g(0) \|_{\text{max}}^{(d,d)} + \| R \|_{\text{max}}^{(d,d)}
\leq cs \left( d^{\alpha + \tilde{\alpha} + 2} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}} + d^{-\tau + 1/2} \right) = \gamma_n.
\]
By Condition \( \Theta \) and Proposition \( \Phi \) under Condition \( \Psi \) (ii), we have
\[
\| \hat{B} - B_0 \|_1^{(d,d)} = O_p \left\{ \mu^{-2} s^2 d^\alpha \left( d^{\alpha + \tilde{\alpha} + 2} \mathcal{M}_{W,Y} \sqrt{\frac{\log p}{n}} + d^{-\tau + 1/2} \right) \right\}.
\]
(B.28)

For each \( j \in [p] \), let \( R_j(u,v) = (\sum_{l=1}^d \sum_{m=1}^d - \sum_{l,m=1}^\infty) a_{jlm} \nu^j(u) \nu^j(v) \) and write
\[
\hat{\beta}_j(u,v) - \beta_0(u,v) = \hat{\psi}_j(u)^T \hat{B}_j \phi(v) - \psi_j(u)^T B_0 \phi(v) + R_j(u,v)
= \hat{\psi}_j(u)^T \hat{B}_j \{ \phi(v) - \phi(v) \} + (\hat{\psi}_j(u) - \psi_j(u))^T \hat{B}_j \phi(v)
+ \psi_j(u)^T \{ \hat{B}_j - B_0 \} \phi(v) + R_j(u,v).
\]

By Lemma 9 of Guo and Qiao (2022), we bound the first three terms by
\[
\| \hat{\psi}_j^T \hat{B}_j (\phi - \phi) \|_S \leq d^{1/2} \max_{m \in [d]} \| \hat{\nu}_m - \phi_m \|_S \| \hat{B}_j \|_F,
\| (\hat{\psi}_j - \psi_j)^T \hat{B}_j \phi \|_S \leq d^{1/2} \max_{l \in [d]} \| \hat{\psi}_j - \psi_j \|_S \| \hat{B}_j \|_F,
\| \psi_j^T (\hat{B}_j - B_0) \phi \|_S = \| \hat{B}_j - B_0 \|_F.
\]
(B.29)

We next bound the fourth term. For \( j \in S \), by the orthonormality of \{\psi_j\} and \{\phi_m\},
\[
\| R_j \|_S^2 = \left\| \left( \sum_{l=1}^d \sum_{m=1}^d - \sum_{l,m=1}^\infty \right) a_{jlm} \nu^j \nu^m \right\|_S^2
= O(1) \cdot \left\{ \sum_{l=1}^d \sum_{m=d+1}^\infty a_{jlm}^2 + \sum_{l=1}^\infty \sum_{m=1}^d a_{jlm}^2 \right\}
= O(1) \cdot \left\{ \sum_{l=1}^d \sum_{m=d+1}^\infty (l + m)^{-2\tau - 1} + \sum_{l=1}^\infty \sum_{m=1}^d (l + m)^{-2\tau - 1} \right\} = O(d^{-2\tau + 1}).
\]
(B.30)

Combining (B.29) and (B.30), we obtain \( \| \hat{\beta} - \beta_0 \|_1 \leq \| \hat{B}_j \|_1^{(d,d)} \{ d^{1/2} \max_{m \in [d]} \| \hat{\nu}_m - \phi_m \| + d^{1/2} \max_{j \in [p], l \in [d]} \| \hat{\psi}_j - \psi_j \| \} + \| \hat{\psi}_j - \psi_j \| + O(s d^{-\tau + 1/2}) \), where the third term above is of a smaller order of the second term due to (B.28). By \( \| \hat{B}_j \|_1^{(d,d)} \leq \| \hat{B} - B_0 \|_1^{(d,d)} + \| B_0 \|_1^{(d,d)} \), (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. \( \square \)
B.10 Proof of Theorem \textit{6}

For each \( j \in [p] \), we first verify Condition \textit{i}) for VFAR. On event \( I_1 \) in (B.17),
\[
\|\hat{G}_j - G_{j}\|^2_{\max} = \max_{j',k \in [p], h \in [L], h' \in [H]} \left\| \frac{1}{n - H - h} \sum_{t = H + h + 1}^{n} \eta_{(t - H - h)k} \hat{H}_{j'}^{T} - \mathbb{E}\{\eta_{(t - H - h)k}H_{j'}^{T}\} \right\|_{F}^2 \\
\leq c d^{2 + 2M_1 W} \sqrt{\frac{\log p}{n}},
\]  
\[\tag{B.31}\]

It follows from Theorem \textit{2} that \( \mathbb{P}(I_1) \geq 1 - \bar{c} p^{-c} \) for some positive constants \( \bar{c}, \bar{c} \). By (B.31) and (B.32), Condition \textit{i}) can be verified by choosing \( \epsilon_{n_1} = c d^{2 + 2M_1 W} \sqrt{\frac{\log p}{n}}. \) \[\tag{B.33}\]

We next verify Condition \textit{ii}) for VFAR. It follows from \( r_{ij} = (r_{ij1}, \ldots, r_{ijd})^{T} \) with each \( r_{ijm'} = \sum_{h' = 1}^{H} \sum_{j'}^{p} \sum_{l = d + 1}^{\infty} \eta_{(t - h')j'} \langle \psi_{j'}^T, A_{0,j'}^{(h')} \rangle, \) orthonormality of \( \{\psi_{j'}\} \), Cauchy–Schwarz inequality and Condition \textit{i}) that
\[
\{ \| R_{j} \|^2_{\max} \}^2 = \max_{k \in [p], h \in [L]} \left\| \mathbb{E}\{\eta_{(t - H - h)k}A_{j}^{T}\} \right\|_{F}^2 \\
= \max_{k, h} \sum_{m = 1}^{d} \sum_{m' = 1}^{d} \left\{ \sum_{(j', h') \in S_{j}} \sum_{l = d + 1}^{\infty} \sqrt{\mathbb{E}\{\eta_{(t - H - h)k}^{2}\} \mathbb{E}\{\eta_{(t - h')j'}^{2}\}} \right\}^2 \\
\leq s_j^2 \max_{k, j', h'} \sum_{m = 1}^{d} \sum_{m' = 1}^{d} \left\{ \sum_{l = d + 1}^{\infty} \lambda_{km}^{1/2} \lambda_{j'T}^{1/2} \right\}^2 \\
\leq s_j^2 \max_{k} \sum_{m = 1}^{d} \lambda_{km} \max_{j', h'} \left[ \sum_{l = d + 1}^{\infty} \lambda_{j'T} \sum_{m' = 1}^{d} \sum_{l = d + 1}^{\infty} \{a_{j'l'm'}^{(h')}\}^2 \right] \\
\leq \lambda_{0}^2 s_j^2 \sum_{m' = 1}^{d} \sum_{l = d + 1}^{\infty} (l + m')^{-2\tau - 1} = O(s_j^2 d^{-2\tau + 1}),
\]
which implies that
\[
\| R_{j} \|^2_{\max} \leq \epsilon s_j d^{-\tau + 1/2} = \epsilon_2.
\]  
\[\tag{B.34}\]
By the similar technique above and Condition 8(i), we have

$$\|\Omega_{j l}^{(d, d)}\|_{1} \leq \sum_{(j', h') \in S_j} \left( \sum_{l=1}^{d} \sum_{m=1}^{d} \left\{ a_{j j' lm}^{(h')} \right\}^2 \right)^{1/2} s_j \max_{(j', h') \in S_j} \left\{ \sum_{l=1}^{d} \sum_{m=1}^{d} (l + m)^{-2\tau - 1} \right\}^{1/2} = O(s_j). \quad (B.35)$$

Finally, we verify Condition 8(iii) for VFAR. On event $I_1$, combining (B.33), (B.34), (B.35) and applying the similar techniques, we have

$$\| g_j(\Omega_{j l}) \|_{\max}^{(d, d)} \leq \| \hat{G}_j - G_j \|_{\max}^{(d, d)} \| \Omega_{j l} \|_{1}^{(d, d)} + \| g_j(0) - g_j(0) \|_{\max}^{(d, d)} \| R_j \|_{\max}^{(d, d)} \leq c s_j \left( d^{a+2} M_1^W \sqrt{\frac{\log p}{n}} + d^{-\tau + 1/2} \right) = \gamma_{n, j}. \quad (B.36)$$

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

$$\hat{A}_{j j'}^{(h')} (u, v) - A_{0,jj'}^{(h')} (u, v) = \hat{\psi}_{j'}(u) \hat{\Omega}_{j j'}^{(h')} \hat{\psi}_{j}(v) - \psi_{j'}(u) \psi_{j}(v) + R_{jj'}^{(h')} (u, v)\quad (B.37)$$

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

$$\| \hat{\psi}_{j'}^{\top} \hat{\Omega}_{j j'}^{(h')} (\hat{\psi}_{j} - \psi_{j}) \|_{S} \leq d^{1/2} \max_{l \in [d]} \| \hat{\psi}_{jl} - \psi_{jl} \| \left\| \hat{\Omega}_{j j'}^{(h')} \right\|_{F},$$

$$\| (\hat{\psi}_{j'} - \psi_{j'})^{\top} \hat{\Omega}_{j j'}^{(h')} \psi_{j} \|_{S} \leq d^{1/2} \max_{m \in [d]} \| \hat{\psi}_{j' m} - \psi_{j' m} \| \left\| \hat{\psi}_{j} \right\| \left\| \hat{\Omega}_{j j'}^{(h')} \right\|_{F}, \quad (B.37)$$

$$\| \hat{\psi}_{j'} \left( \hat{\Omega}_{j j'}^{(h')} - \Omega_{0,jj'}^{(h')} \right) \psi_{j} \|_{S} = \| \hat{\Omega}_{j j'}^{(h')} - \Omega_{0,jj'}^{(h')} \|_{F}. \quad (B.37)$$

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

$$\| R_{jj'}^{(h')} \|_{S}^{2} = \left\| \left( \sum_{l=1}^{d} \sum_{m=1}^{d} - \sum_{l, m=1}^{d} a_{j j' lm}^{(h')} \right) \psi_{j l} \psi_{j' m} \right\|_{S}^{2}$$

$$= O(1) \left\{ \sum_{l=1}^{d} \sum_{m=1}^{d} \sum_{l, m=1}^{d} \sum_{l, m=1}^{d} (l + m)^{-2\tau - 1} \right\} = O(d^{-2\tau + 1}). \quad (B.38)$$

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

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

33
where the third term above is of a smaller order of the second term due to (B.36). By \( \max_j \| \hat{\Omega}_j^{(d,d)} \|_1 \leq \max_j \| \hat{\Omega}_j - \Omega_{0j}^{(d,d)} \|_1 + \max_j \| \Omega_{0j}^{(d,d)} \|_1 \) \( (B.35) \) 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. \( \square \)

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

References

Aue, A., Norinho, D. and Hörmann, S. (2015). On the prediction of stationary functional time series, Journal of the American Statistical Association 110(509): 378–392.

Basu, S. and Michailidis, G. (2015). Regularized estimation in sparse high-dimensional time series models, The Annals of Statistics 43(4): 1535–1567.

Bathia, N., Yao, Q. and Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series, The Annals of Statistics 38(6): 3352–3386.

Belloni, A., Chernozhukov, V., Chetverikov, D., Hansen, C. and 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. and Qiao, X. (2022). Functional linear regression: dependence and error contamination, Journal of Business & Economic Statistics 40(1): 444–457.

Cho, H., Goude, Y., Brossat, X. and Yao, Q. (2013). Modeling and forecasting daily electricity load curves: a hybrid approach, Journal of the American Statistical Association 108(501): 7–21.

Descary, M.-H. and Panaretos, V. M. (2019). Functional data analysis by matrix completion, The Annals of Statistics 47(1): 1–38.

Fan, Y., Foutz, N., James, G. M. and Jank, W. (2014). Functional response additive model estimation with online virtual stock markets, The Annals of Applied Statistics 8(4): 2435–2460.

Fan, Y., James, G. M. and Radchenko, P. (2015). Functional additive regression, The Annals of Statistics 43(5): 2296–2325.

Fang, Q., Guo, S. and Qiao, X. (2022). Finite sample theory for high-dimensional functional/scalar time series with applications, Electronic Journal of Statistics 16(1): 527–591.

Fu, A., Narasimhan, B. and Boyd, S. (2020). CVXR: An R package for disciplined convex optimization, Journal of Statistical Software 94: 1–34.

Gautier, E. and Rose, C. (2019). High-dimensional instrumental variables regression and confidence sets, arXiv:1105.2454.

Guo, S. and Qiao, X. (2022). On consistency and sparsity for high-dimensional functional time series with application to autoregressions, Bernoulli, in press.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression, The Annals of Statistics 35(1): 70–91.

Hall, P. and Vial, C. (2006). Assessing the finite dimensionality of functional data, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68(4): 689–705.

Hamilton, J. D. (1994). Time series analysis, Vol. 2, Princeton New Jersey.

Hörmann, S., Kidziński, L. and Hallin, M. (2015). Dynamic functional principal components, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77(2): 319–348.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data, The Annals of Statistics 38(3): 1845–1884.

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series, Journal of Econometrics 179(1): 66–82.

Kong, D., Xue, K., Yao, F. and Zhang, H. H. (2016). Partially functional linear regression in high dimensions, Biometrika 103(1): 147–159.

Li, D., Robinson, P. M. and Shang, H. L. (2020). Long-range dependent curve time series, Journal of the American
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                        |
| BIB    | 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 COR                      |
| 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                 |
| INJ    | JOHNSON & JOHNSON         | XOM    | EXXON MOBIL CORP                   |

35
Statistical Association 115(530): 957–971.
Luo, R. and Qi, X. (2017). Function-on-function linear regression by signal compression, Journal of the American Statistical Association 112(518): 690–705.
Müller, H.-G., Sen, R. and Stadtmüller, U. (2011). Functional data analysis for volatility, Journal of Econometrics 165(2): 233–245.
Panaretos, V. M. and Tavakoli, S. (2013). Fourier analysis of stationary time series in function space, The Annals of Statistics 41(2): 568–603.
Rudelson, M. and Vershynin, R. (2013). Hanson-wright inequality and sub-gaussian concentration, Electronic Communications in Probability 18: 1–9.
Shang, H. L. (2013). ftsa: An R package for analyzing functional time series., R Journal 5(1): 64–72.
Xue, K. and Yao, F. (2021). Hypothesis testing in large-scale functional linear regression, Statistica Sinica 31: 1101–1123.
Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data, Journal of the American statistical association 100(470): 577–590.
Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68(1): 49–67.