Submitted to Bernoulli

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

SHAOJUN GUO[1] and XINGHAO QIAO[2]

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

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

Modelling a large collection of functional time series arises in a broad spectral of real applications. Under such a scenario, not only the number of functional variables can be diverging with, or even larger than the number of temporally dependent functional observations, but each function itself is an infinite-dimensional object, posing a challenging task. In this paper, we propose a three-step procedure to estimate high-dimensional functional time series models. To provide theoretical guarantees for the three-step procedure, we focus on multivariate stationary processes and propose a novel functional stability measure based on their spectral properties. Such stability measure facilitates the development of some useful concentration bounds on sample (auto)covariance functions, which serve as a fundamental tool for further convergence analysis in high-dimensional settings. As functional principal component analysis (FPCA) is one of the key dimension reduction techniques in the first step, we also investigate the non-asymptotic properties of the relevant estimated terms under a FPCA framework. To illustrate with an important application, we consider vector functional autoregressive models and develop a regularization approach to estimate autoregressive coefficient functions under the sparsity constraint. Using our derived non-asymptotic results, we investigate convergence properties of the regularized estimate under high-dimensional scaling. Finally, the finite-sample performance of the proposed method is examined through both simulations and a public financial dataset.

Keywords: functional principal component analysis, functional stability measure, high-dimensional functional time series, non-asymptotics, sparsity, vector functional autoregression.

1. Introduction

In functional data analysis, it is commonly assumed that each measured function, treated as the unit of observation, is independently sampled from some realization of an underlying stochastic process. Functional time series, on the other hand, refers to a collection of curves observed consecutively over time, where the temporal dependence across observations exhibits. The literature has mainly focused on univariate or bivariate functional time series, see, e.g., Hörmann and Kokoszka (2010); Cho et al. (2013); Panaretos

and Tavakoli (2013); Hörmann, Kidziński and Hallin (2015); Jirak (2016); Li, Robinson and Shang (2020) and the reference therein. Recent advances in technology have made multivariate or even high-dimensional functional time series datasets become increasingly common in many applications. Examples include cumulative intraday return trajectories (Horváth, Kokoszka and Rice, 2014) and functional volatility processes (Müller, Sen and Stadtmüller, 2011) for a large number of stocks, daily concentration curves of particulate matter and gaseous pollutants at different locations (Li et al., 2017), and intraday energy consumption curves for thousands of London households (available at https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households). These applications require understanding relationships among a relatively large collection of functional variables based on temporally dependent functional observations.

Throughout the paper, suppose we observe p-dimensional vector of functional time series, $\mathbf{X}_t(\cdot) = \big(X_{t1}(\cdot), \ldots, X_{tp}(\cdot)\big)^{\mathrm{T}}$ for $t = 1, \ldots, n$, defined on a compact interval $\mathcal{U}$. Addressing multivariate or even high-dimensional functional time series problems poses challenges and is largely untouched in the literature. Under such a scenario, not only p is large relative to n, but each $X_{tj}(\cdot)$ is an infinite-dimensional object with temporal dependence across observations. A standard procedure towards the estimation of models involving high-dimensional functional data consists of three steps. In the first step, due to the infinite-dimensional nature of functional data, some form of dimension reduction, e.g. data-driven basis expansion via functional principal component analysis (FPCA) or its dynamic version (Hörmann, Kidziński and Hallin, 2015) and pre-fixed basis expansion (Fan, James and Radchenko, 2015), is needed to approximate each $X_{tj}(\cdot)$ by the q_j-dimensional truncation, which transforms the problem of modelling a p-dimensional vector of functional time series into that of modelling $(\sum_{j=1}^{p} q_j)$-dimensional vector time series. The second step involves the estimation under a high-dimensional and dependent setting, where some lower-dimensional structure is commonly imposed on the model parameter space. One large class assumes sparse function-valued parameters involved in high-dimensional functional time series models. Under functional sparsity constraints, the first step results in the estimation of block sparse vector- or matrix-valued parameters in the second step, where different regularized estimation procedures can be developed in a blockwise fashion, see, e.g., under an independent setting, Fan, James and Radchenko (2015); Kong et al. (2016) and Qiao, Guo and James (2019). Finally, for interpretation and prediction, the third step recovers functional sparse estimates from those block sparse estimates obtained in the second step.

The essential challenge to support such three-step estimation procedure is to provide theoretical guarantees in a range of high-dimensional settings, e.g., $\log p/n \to 0$. Within such high-dimensional statistics framework, one main goal is to obtain some non-asymptotic results, i.e. error bounds on a given performance metric that hold with high probability for a finite sample size n and provide explicit control on the dimension p as well as other structural parameters. Compared with non-functional data, the intrinsic infinite-dimensionality of each process $X_{tj}(\cdot)$ leads to a significant rise in theoretical complexity of the problem, since one needs to develop some operator- and FPCA-based non-asymptotic results for dependent processes within an abstract Hilbert space and to

propose a dependence measure to capture the effect of temporal dependence on non-asymptotic properties. The existing theoretical work mainly for the first step has focused on studying its asymptotic properties by treating p as fixed under a moment-based dependence structure (Hörmann and Kokoszka, 2010) or its non-asymptotic properties under either an independent setting (Koltchinskii and Lounici, 2017; Qiao et al., 2020; Araya Valdivia, 2020) or a special autoregressive structure (Bosq, 2000). These results, however, are not sufficient to evaluate the performance of the three-step procedure in a high-dimensional regime with a general dependence structure. Such a challenging task motivates us to develop some essential non-asymptotic results under the setting we consider, which fills the gap between practical implementation and theoretical justification and forms the core of our paper.

A key innovation in our paper is to propose a functional stability measure for a large class of stationary Gaussian processes, $\{\mathbf{X}_t(\cdot)\}$, based on their spectral density functions. Such stability measure provides new insights into the effect of temporal dependence on theoretical properties of $\widehat{\mathbf{\Sigma}}_h$'s, the estimators for autocovariance functions $\mathbf{\Sigma}_h(u,v) = \mathrm{Cov}\{\mathbf{X}_t(u), \mathbf{X}_{t+h}(v)\}$ with $h = 0, \pm 1, \ldots$ and $(u,v) \in \mathcal{U}^2$, and, in particular, facilitates the development of some novel concentration bounds on $\widehat{\mathbf{\Sigma}}_h$ serving as a fundamental tool for further convergence analysis under high-dimensional scaling. Based on these concentration bounds, we establish non-asymptotic error bounds on relevant estimated terms under a FPCA framework so as to provide theoretical guarantees for our proposed three-step procedure. Such concentration results can also lead to convergence analysis of other possible high-dimensional functional time series models, e.g. those mentioned in Section A of the Supplementary Material (Guo and Qiao, 2022). It is worth noting that the functional stability measure is fundamentally different from the direct extension of the stability measure (Basu and Michailidis, 2015) to the functional domain. This is because, for truly infinite-dimensional functional objects, in contrast to the functional analog of Basu et al.'s stability measure, which just controls the maximum eigenvalue for spectral density functions of $\{\mathbf{X}_t(\cdot)\}$, the functional stability measure utilizes the functional Rayleigh quotients of spectral density functions relative to $\mathbf{\Sigma}_0$ and hence can more precisely capture the effect of small decaying eigenvalues. Such functional stability measure also leads to non-asymptotic results of normalized versions of relevant estimated terms, making the characterization of relevant tail behaviours more accurate for small eigenvalues. To the best of our knowledge, we are the first to propose such a dependence measure for high-dimensional functional time series and rely on it to develop some essential non-asymptotic results.

To illustrate the proposed three-step approach and the usefulness of the derived non-asymptotic results with an important application, we consider vector functional autoregressive (VFAR) models, which characterize the temporal and cross-sectional interrelationships in $\{\mathbf{X}_t(\cdot)\}$. One advantage of a VFAR model is that it accommodates dynamic linear interdependencies in $\{\mathbf{X}_t(\cdot)\}$ into a static framework within a Hilbert space. Moreover, a sparse VFAR model facilitates the extraction of Granger causal networks (Basu, Shojaie and Michailidis, 2015) under the functional domain. The VFAR estimation is intrinsically a very high-dimensional problem, since, in the second step of our

procedure, we need to fit a *vector autoregressive* (VAR) model, whose dimensionality, $(\sum_{j=1}^{p} q_j)^2$, grows quadratically with $\sum_{j=1}^{p} q_j$. For example, estimating a VFAR model of order 1 with $p = 20$ and $q_j = 5$ requires estimating $100^2 = 10,000$ parameters. Under high-dimensional scaling and the sparsity assumption on the functional transition matrices, the second step requires to estimate a block sparse VAR model. We then propose the regularized estimates of block transition matrices, on which the block sparsity constraint is enforced via a standardized group lasso penalty (Simon and Tibshirani, 2012). Using the derived non-asymptotic results, we show that the proposed three-step approach can produce consistent estimates for sparse VFAR models in high dimensions.

Related literature. Our work lies in the intersection of high-dimensional statistics, functional data analysis and time series analysis, each of which corresponds to a vast literature, hence we will only review the most closely related intersectional work to ours. (i) For high-dimensional independent functional data, regularization methods have recently been proposed to estimate different types of functional sparse models, e.g., functional additive regression (Fan, James and Radchenko, 2015; Kong et al., 2016), static functional graphical models (Li and Solea, 2018; Qiao, Guo and James, 2019) and its dynamic version (Qiao et al., 2020). (ii) For high-dimensional time series, some essential concentration bounds were established for Gaussian processes (Basu and Michailidis, 2015), linear processes with more general noise distributions (Sun et al., 2018) and heavy tailed non-Gaussian processes (Wong, Li and Tewari, 2020). (iii) For examples of recent developments in high-dimensional VAR models, Kock and Callot (2015); Basu and Michailidis (2015) and Wong, Li and Tewari (2020) studied the theoretical properties of ℓ_1-type regularized estimates. Basu, Shojaie and Michailidis (2015) and Billio, Casarin and Rossini (2019) considered extracting Granger causal networks from sparse VAR models. See also Han, Lu and Liu (2015), Guo, Wang and Yao (2016) and Ghosh, Khare and Michailidis (2019). (iv) For examples of research on functional autoregressive models, see Bosq (2000); Kokoszka and Reimherr (2013); Aue, Norinho and Hörmann (2015) and the reference therein.

Outline of the paper. In Section 2, we first introduce a functional stability measure and rely on it to establish concentration bounds on the quadratic/bilinear forms of $\widehat{\mathbf{\Sigma}}_h$ leading to elementwise concentration results for $\widehat{\mathbf{\Sigma}}_h$. We then establish theoretical guarantees for the proposed three-step approach by deriving some useful non-asymptotic error bounds under a FPCA framework. In Section 3, we develop a three-step procedure to estimate the sparse VFAR model, connect with casual network modelling, present the convergence analysis of the regularized estimate and finally examine the finite-sample performance through both simulation studies and an analysis of a public financial dataset.

Notation. We summarize here some notation to be used throughout the paper. Let $\mathbb{Z}$ denotes the set of integers. For two positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ if there exists an absolute constant c, such that $a_n \leqslant cb_n$ for all n. We write $a_n \asymp b_n$ if and only if $b_n \lesssim a_n$ and $a_n \lesssim b_n$. We use $x \vee y = \max(x, y)$. For matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{p_1 \times p_2}$, we let $\langle\langle \mathbf{A}, \mathbf{B} \rangle\rangle = \text{trace}(\mathbf{A}^{\mathrm{T}}\mathbf{B})$ and denote the Frobenius norm of $\mathbf{B}$ by $||\mathbf{B}||_{\mathrm{F}} = \left(\sum_{j,k} \mathrm{B}_{jk}^2\right)^{1/2}$. Let $L_2(\mathcal{U})$ denote a Hilbert space of square integrable functions defined on the compact set $\mathcal{U}$ equipped with the inner product $\langle f, g \rangle = \int_{\mathcal{U}} f(u)g(u)du$

for $f, g \in L_2(\mathcal{U})$ and the induced norm $\|\cdot\| = \langle\cdot,\cdot\rangle^{1/2}$. We denote its p-fold Cartesian product by $\mathbb{H} = L_2(\mathcal{U}) \times \cdots \times L_2(\mathcal{U})$ and the tensor product by $\mathbb{S} = L_2(\mathcal{U}) \otimes L_2(\mathcal{U})$. For $\mathbf{f} = (f_1, \ldots, f_p)^{\mathrm{T}}$ and $\mathbf{g} = (g_1, \ldots, g_p)^{\mathrm{T}}$ in $\mathbb{H}$, we denote the inner product by $\langle \mathbf{f}, \mathbf{g} \rangle_{\mathbb{H}} = \sum_{j=1}^{p} \langle f_j, g_j \rangle$ and the induced norm by $\|\cdot\|_{\mathbb{H}} = \langle\cdot,\cdot\rangle_{\mathbb{H}}^{1/2}$. We use $\|\mathbf{f}\|_0 = \sum_{j=1}^{p} I(\|f_j\| \neq 0)$ to denote the functional version of vector ℓ_0 norm. For any $K \in \mathbb{S}$, it can be viewed as the kernel function of a linear operator acting on $L_2(\mathcal{U})$, i.e. for each $f \in L_2(\mathcal{U})$, K maps $f(u)$ to $K(f)(u) = \int_{\mathcal{U}} K(u,v) f(v) dv$. For notational economy, we will use K to denote both the kernel function and the operator. Moreover, we denote the operator and Hilbert-Schmidt norms by $\|K\|_{\mathcal{L}} = \sup_{\|f\| \leqslant 1} \|K(f)\|$ and $\|K\|_{\mathcal{S}} = \left(\int\int K(u,v)^2 du dv \right)^{1/2}$, respectively. For $\mathbf{A} = \left(\mathrm{A}_{jk}\right)_{1 \leqslant j,k \leqslant p}$ with its (j,k)-th component $\mathrm{A}_{jk} \in \mathbb{S}$, we define the functional versions of Frobenius, elementwise ℓ_∞ and matrix ℓ_∞ norms by $\|\mathbf{A}\|_{\mathrm{F}} = \left(\sum_{j,k} \|\mathrm{A}_{jk}\|_{\mathcal{S}}^2 \right)^{1/2}$, $\|\mathbf{A}\|_{\max} = \max_{j,k} \|\mathrm{A}_{jk}\|_{\mathcal{S}}$ and $\|\mathbf{A}\|_{\infty} = \max_j \sum_k \|\mathrm{A}_{jk}\|_{\mathcal{S}}$, respectively.

2. Main results

Suppose that $\{\mathbf{X}_t(\cdot)\}_{t \in \mathbb{Z}}$, defined on $\mathcal{U}$, is a sequence of p-dimensional vector of centered and covariance-stationary Gaussian processes with mean zero and $p \times p$ autocovariance functions, $\mathbf{\Sigma}_h = \left(\Sigma_{jk}^{(h)}\right)_{1 \leqslant j,k \leqslant p}$ with its (j,k)-th component $\Sigma_{jk}^{(h)} \in \mathbb{S}$ for $h \in \mathbb{Z}$. In particular when $h = 0$, one typically refers to $\Sigma_{jj}^{(0)}$ as marginal-covariance functions for $j = k$, and cross-covariance functions for $j \neq k$. To simplify notation, we will also use $\mathbf{\Sigma}_h$ to denote the lag-h autocovariance operator induced from the kernel function $\mathbf{\Sigma}_h$, i.e., for any given $\mathbf{\Phi} \in \mathbb{H}$,

$$\mathbf{\Sigma}_h(\mathbf{\Phi})(u) := \int_{\mathcal{U}} \mathbf{\Sigma}_h(u,v) \mathbf{\Phi}(v) dv = \left(\left\langle \boldsymbol{\sigma}_1^{(h)}(u,\cdot), \mathbf{\Phi}(\cdot) \right\rangle_{\mathbb{H}}, \ldots, \left\langle \boldsymbol{\sigma}_p^{(h)}(u,\cdot), \mathbf{\Phi}(\cdot) \right\rangle_{\mathbb{H}} \right)^{\mathrm{T}} \in \mathbb{H},$$

where $\boldsymbol{\sigma}_j^{(h)}(u,\cdot) = \left(\Sigma_{j1}^{(h)}(u,\cdot), \ldots, \Sigma_{jp}^{(h)}(u,\cdot)\right)^{\mathrm{T}}$ for $j = 1, \ldots, p$. In the special case of $h = 0$, the covariance function $\mathbf{\Sigma}_0$ is symmetric and non-negative definite, i.e. $\mathbf{\Sigma}_0(u,v) = \mathbf{\Sigma}_0(v,u)^{\mathrm{T}}$ for any $(u,v) \in \mathcal{U}^2$ and $\langle \mathbf{\Phi}, \mathbf{\Sigma}_0(\mathbf{\Phi}) \rangle_{\mathbb{H}} \geqslant 0$ for any $\mathbf{\Phi} \in \mathbb{H}$.

2.1. Functional stability measure

To introduce the functional stability measure, we first consider the spectral density operator of $\{\mathbf{X}_t(\cdot)\}_{t \in \mathbb{Z}}$, defined from the Fourier transform of autocovariance operators $\{\mathbf{\Sigma}_h\}_{h \in \mathbb{Z}}$, which encodes the second-order dynamical properties of $\{\mathbf{X}_t(\cdot)\}_{t \in \mathbb{Z}}$.

***Definition* 1.** *We define the spectral density operator of $\{\mathbf{X}_t(\cdot)\}_{t \in \mathbb{Z}}$ at frequency θ by*

$$f_{\mathbf{X},\theta} = \frac{1}{2\pi} \sum_{h \in \mathbb{Z}} \mathbf{\Sigma}_h \exp(-ih\theta), \quad \theta \in [-\pi, \pi]. \tag{1}$$

The spectral density operator (or function) generalizes the notion of the spectral density matrix (Basu and Michailidis, 2015) to the functional domain, and it can also be viewed as a generalization of the spectral density operator (or function) (Panaretos and Tavakoli, 2013) to the multivariate setting. Furthermore, if $\sum_{h=0}^{\infty}\|\mathbf{\Sigma}_h\|_{\mathcal{L}}<\infty$, then $f_{\mathbf{X},\theta}$ is uniformly bounded and continuous in θ with respect to $\|\cdot\|_{\mathcal{L}}$, where we denote by $\|\mathbf{\Sigma}_h\|_{\mathcal{L}}=\sup_{\|\mathbf{\Phi}\|_{\mathbb{H}}\leqslant 1,\mathbf{\Phi}\in\mathbb{H}}\|\mathbf{\Sigma}_h(\mathbf{\Phi})\|_{\mathbb{H}}$ the operator norm of $\mathbf{\Sigma}_h$, and the following inversion formula holds:

$$\mathbf{\Sigma}_h=\int_{-\pi}^{\pi}f_{\mathbf{X},\theta}\exp(ih\theta)d\theta,\qquad\text{for all } h\in\mathbb{Z}. \tag{2}$$

The inversion relationships in (1) and (2) indicate that spectral density operators and autocovariance operators comprise a Fourier transform pair. Hence, to study the second-order dynamics of $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$, we can impose conditions on $\mathbf{\Sigma}_0$ and $\{f_{\mathbf{X},\theta},\theta\in[-\pi,\pi]\}$ in the following Conditions 1 and 2, respectively, which together imply that $\{f_{\mathbf{X},\theta},\theta\in[-\pi,\pi]\}$ are trace-class operators.

***Condition* 1.** *(i) The marginal-covariance functions, $\Sigma_{jj}^{(0)}$'s, are continuous on $\mathcal{U}^2$ and uniformly bounded over $j\in\{1,\ldots,p\}$; (ii) $\lambda_0=\max_{1\leqslant j\leqslant p}\int_{\mathcal{U}}\Sigma_{jj}^{(0)}(u,u)du=O(1)$.*

***Condition* 2.** *(i) The spectral density operators $f_{\mathbf{X},\theta},\theta\in[-\pi,\pi]$ exist; (ii) The functional stability measure of $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$ defined as follows, is bounded, i.e.*

$$\mathcal{M}(f_{\mathbf{X}})=2\pi\cdot\operatorname*{esssup}_{\theta\in[-\pi,\pi],\mathbf{\Phi}\in\mathbb{H}_0^p}\frac{\big\langle\mathbf{\Phi},f_{\mathbf{X},\theta}(\mathbf{\Phi})\big\rangle_{\mathbb{H}}}{\big\langle\mathbf{\Phi},\mathbf{\Sigma}_0(\mathbf{\Phi})\big\rangle_{\mathbb{H}}}<\infty, \tag{3}$$

where $\mathbb{H}_0=\big\{\mathbf{\Phi}\in\mathbb{H}:\langle\mathbf{\Phi},\mathbf{\Sigma}_0(\mathbf{\Phi})\rangle_{\mathbb{H}}\in(0,\infty)\big\}$.

In general, we can relax Condition 1(ii) by allowing λ_0 to grow at some slow rate as p increases. Then our established non-asymptotic bounds, e.g., those in (17) and (18), will depend on λ_0. We next provide several comments for Condition 2. First, the functional stability measure is proportional to the essential supremum of the functional Rayleigh quotient of $f_{\mathbf{X},\boldsymbol{\theta}}$ relative to $\mathbf{\Sigma}_0$ over $\theta\in[-\pi,\pi]$. In particular, under the non-functional setting with $\mathbf{\Phi}\in\mathbb{R}^p$ and $f_{\mathbf{X},\theta},\mathbf{\Sigma}_0\in\mathbb{R}^{p\times p}$, (3) reduces to

$$2\pi\cdot\operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\mathbf{\Phi}\neq\mathbf{0}}\frac{\mathbf{\Phi}^{\mathrm{T}}f_{\mathbf{X},\theta}\mathbf{\Phi}}{\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Sigma}_0\mathbf{\Phi}}<\infty,$$

which is equivalent to the upper bound condition for the stability measure, $\widetilde{\mathcal{M}}(f_{\mathbf{X}})$, introduced by Basu and Michailidis (2015), i.e.

$$\widetilde{\mathcal{M}}(f_{\mathbf{X}})=\operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\mathbf{\Phi}\neq\mathbf{0}}\frac{\mathbf{\Phi}^{\mathrm{T}}f_{\mathbf{X},\theta}\mathbf{\Phi}}{\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}}<\infty.$$

Second, if $X_{t1}(\cdot),\ldots,X_{tp}(\cdot)$ are finite-dimensional objects, the upper bound conditions for $\mathcal{M}(f_{\mathbf{X}})$ and the functional analog of $\widetilde{\mathcal{M}}(f_{\mathbf{X}})$ are equivalent. However, for truly infinite-dimensional functional objects, $\mathcal{M}(f_{\mathbf{X}})$ makes more sense, since it can more precisely

capture the effect of small eigenvalues of $f_{\mathbf{X},\theta}$'s relative to those of $\mathbf{\Sigma}_0$. Moreover, Condition 2 is satisfied by a large class of infinite-dimensional functional data, see examples in Section B of the Supplementary Material (Guo and Qiao, 2022). Third, it is clear that, unlike $\widetilde{\mathcal{M}}(f_{\mathbf{X}})$, $\mathcal{M}(f_{\mathbf{X}})$ is a scale-free stability measure. In the special case of no temporal dependence, $\mathcal{M}(f_{\mathbf{X}}) = 1$. Fourth, since the autocovariance function characterizes a multivariate Gaussian process, it can be used to quantify the temporal and cross-sectional dependence for this class of models. In particular, the spectral density functions provide insights into the stability of the process. In our analysis of high-dimensional functional time series, we will use $\mathcal{M}(f_{\mathbf{X}})$ as a stability measure of the process of $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$. Larger values of $\mathcal{M}(f_{\mathbf{X}})$ would correspond to a less stable process.

We next illustrate the superiority of the functional stability measure to possible competitors using VFAR models as an example. See Section 3 for details on VFAR models. In particular, we consider a VFAR model of order 1, denoted by VFAR(1), as follows

$$\mathbf{X}_t(u) = \int_{\mathcal{U}} \mathbf{A}(u,v)\mathbf{X}_{t-1}(v)dv + \boldsymbol{\varepsilon}_t(u),\ u \in \mathcal{U}. \tag{4}$$

In the special case of a symmetric $\mathbf{A}$, i.e. $\mathbf{A}(u,v) = \mathbf{A}(v,u)^{\mathrm{T}}$, equation (4) has a stationary solution if and only if $||\mathbf{A}||_{\mathcal{L}} < 1$. See Theorem 3.5 of Bosq (2000) for $p = 1$. However, this restrictive condition is violated by many stable VFAR(1) models with non-symmetric $\mathbf{A}$. Moreover, it does not generalize beyond VFAR(1) models.

We consider an illustrative example with

$$\mathbf{A}(u,v) = \begin{pmatrix} a\psi_1(u)\psi_1(v) & b\psi_1(u)\psi_2(v) \\ 0 & a\psi_2(u)\psi_2(v) \end{pmatrix}, \mathbf{X}_t(u) = \begin{pmatrix} x_{t1}\psi_1(u) \\ x_{t2}\psi_2(u) \end{pmatrix}, \boldsymbol{\varepsilon}_t(u) = \begin{pmatrix} e_{t1}\psi_1(u) \\ e_{t2}\psi_2(u) \end{pmatrix}, \tag{5}$$

where $(e_{t1}, e_{t2})^{\mathrm{T}} \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{I}_2)$ and $\|\psi_j\| = 1$ for $j = 1, 2$. Section E of the Supplementary Material (Guo and Qiao, 2022) provides details to calculate $\rho(\mathbf{A})$ (spectral radius of $\mathbf{A}$), $||\mathbf{A}||_{\mathcal{L}}$ and $\mathcal{M}(f_{\mathbf{X}})$ for this example. In particular, $\rho(\mathbf{A}) = |a| < 1$ corresponds to a stationary solution to equation (4). Figure 1 visualizes $||\mathbf{A}||_{\mathcal{L}}$ and $\mathcal{M}(f_{\mathbf{X}})$ for various values of $a \in (0,1)$. We observe a few apparent patterns. First, increasing a results in a value for larger $||\mathbf{A}||_{\mathcal{L}}$. As $|b|$ grows large enough, the condition of $||\mathbf{A}||_{\mathcal{L}} < 1$ will be violated, but equation (4) still have a stationary solution. Second, processes with stronger temporal dependence, i.e. with larger values of a or $|b|$, have larger values of $\mathcal{M}(f_{\mathbf{X}})$ and will be considered less stable. For a high-dimensional VFAR(1) model, it is more sensible to use $\mathcal{M}(f_{\mathbf{X}})$ rather than $\|\mathbf{A}\|_{\mathcal{L}}$ or $\|\mathbf{A}^j\|_{\mathcal{L}}$ for some $j \geqslant 1$ (Theorem 5.1 of (Bosq, 2000)) as a measure of stability of the process.

***Definition* 2.** *For all k-dimensional subprocesses of $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$, i.e. $\{(X_{tj}(\cdot)) : j \in J\}_{t\in\mathbb{Z}}$, for $J \subseteq \{1,\ldots,p\}$ and $|J| \leqslant k$, we define the corresponding functional stability measure by*

$$\mathcal{M}_k(f_{\mathbf{X}}) = 2\pi \cdot \operatorname*{ess\,sup}_{\theta\in[-\pi,\pi],\|\mathbf{\Phi}\|_0\leqslant k,\mathbf{\Phi}\in\mathbb{H}_0} \frac{\langle \mathbf{\Phi}, f_{\mathbf{X},\theta}(\mathbf{\Phi})\rangle_{\mathbb{H}}}{\langle \mathbf{\Phi}, \mathbf{\Sigma}_0(\mathbf{\Phi})\rangle_{\mathbb{H}}},\ k = 1,\ldots,p. \tag{6}$$

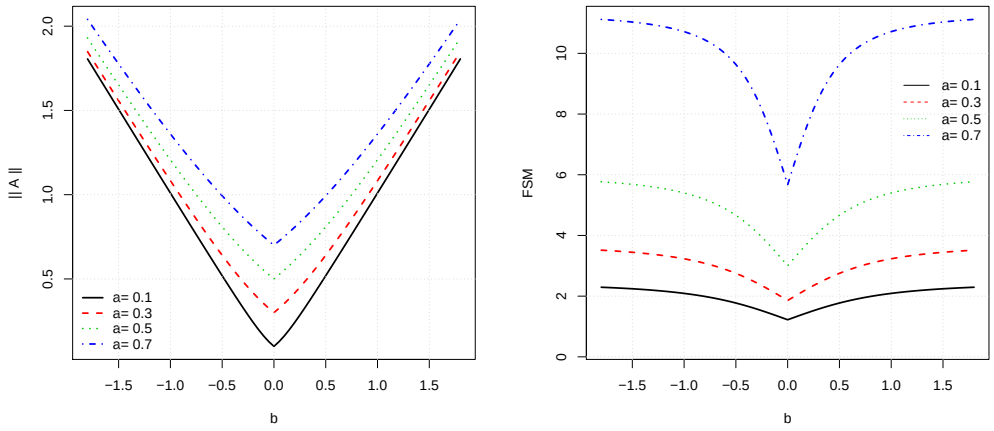

Figure 1. *The illustrative VFAR(1) model. Left:* $||\mathbf{A}||_{\mathcal{L}}$ *as a function of* a *and* b*, plotted against* b *for different* a*. Right:* $\mathcal{M}(f_{\mathbf{X}})$ *as a function of* a *and* b*, plotted against* b *for different* a*.*

It is obvious from definitions in Condition 2 and (6) that $\mathcal{M}_1(f_{\mathbf{X}}) \leqslant \mathcal{M}_2(f_{\mathbf{X}}) \leqslant \cdots \leqslant \mathcal{M}_p(f_{\mathbf{X}}) = \mathcal{M}(f_{\mathbf{X}}) < \infty$ and $\mathcal{M}_1(f_{\mathbf{X}}) = \max_{1\leqslant j\leqslant p} \mathcal{M}(f_{X_j})$, which is allowed to evolve with p satisfying $\mathcal{M}_1(f_{\mathbf{X}}) \lesssim M^{-(2\alpha+1)}\{n/\log(pM)\}^{1/2}$, implied from Theorems 3–4 and Condition 9 below.

2.2. Concentration bounds on $\widehat{\boldsymbol{\Sigma}}_h$

Based on n temporally dependent observations $\mathbf{X}_1(\cdot), \ldots, \mathbf{X}_n(\cdot)$, we construct an empirical estimator of $\boldsymbol{\Sigma}_h$ by

$$\widehat{\boldsymbol{\Sigma}}_h(u,v) = \frac{1}{n-h}\sum_{t=1}^{n-h} \mathbf{X}_t(u)\mathbf{X}_{t+h}(v)^{\mathrm{T}}, \ h = 0, 1, \ldots, \ (u,v) \in \mathcal{U}^2. \tag{7}$$

The following theorem provides concentration bounds on $\widehat{\boldsymbol{\Sigma}}_0$ under the quadratic and bilinear forms. These concentration bounds form the core of our theoretical results, which serve as a starting point to establish further non-asymptotic error bounds presented in Sections 2.2 and 2.3.

Theorem 1. *Suppose that Conditions 1 and 2 hold. Then for any given vectors* $\boldsymbol{\Phi}_1$, $\boldsymbol{\Phi}_2 \in \mathbb{H}_0$ *satisfying* $\|\boldsymbol{\Phi}_1\|_0 \vee \|\boldsymbol{\Phi}_2\|_0 \leqslant k$ *with some integer* k $(1 \leqslant k \leqslant p)$*, there exists some universal constant* $c > 0$ *such that for any* $\eta > 0$*,*

$$P\left\{\left|\frac{\big\langle \boldsymbol{\Phi}_1, (\widehat{\boldsymbol{\Sigma}}_0 - \boldsymbol{\Sigma}_0)(\boldsymbol{\Phi}_1)\big\rangle_{\mathbb{H}}}{\big\langle \boldsymbol{\Phi}_1, \boldsymbol{\Sigma}_0(\boldsymbol{\Phi}_1)\big\rangle_{\mathbb{H}}}\right| > \mathcal{M}_k(f_{\mathbf{X}})\eta\right\} \leqslant 2\exp\Big\{-cn\min(\eta^2, \eta)\Big\}, \tag{8}$$

and

$$P\left\{\left|\frac{\big\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0)(\mathbf{\Phi}_2)\big\rangle_{\mathbb{H}}}{\big\langle \mathbf{\Phi}_1, \mathbf{\Sigma}_0(\mathbf{\Phi}_1)\big\rangle_{\mathbb{H}} + \big\langle \mathbf{\Phi}_2, \mathbf{\Sigma}_0(\mathbf{\Phi}_2)\big\rangle_{\mathbb{H}}}\right| > \mathcal{M}_k(f_{\mathbf{X}})\eta\right\} \leqslant 4\exp\Big\{-cn\min(\eta^2, \eta)\Big\}. \quad (9)$$

The concentration inequalities in (8) and (9) suggest that the temporal dependence may affect the tail behaviors via $\mathcal{M}_k(f_{\mathbf{X}})$ in two different ways, depending on which term in the tail bounds is dominant. With suitable choices of $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$, we can derive non-asymptotic results for entries of $\widehat{\mathbf{\Sigma}}_h$ and relevant estimated terms under a FPCA framework. For example, under the Karhunen-Loève expansion of each $X_{tj}(\cdot)$ (see details in Section 2.3), choosing $\mathbf{\Phi}_1 = (0, \ldots, 0, \phi_{jl}, 0, \ldots, 0)^{\mathrm{T}}$ leads to $\langle \mathbf{\Phi}_1, \mathbf{\Sigma}_0(\mathbf{\Phi}_1)\rangle_{\mathbb{H}} = \lambda_{jl}$ and $\langle \mathbf{\Phi}_1, \widehat{\mathbf{\Sigma}}_0(\mathbf{\Phi}_1)\rangle_{\mathbb{H}} = n^{-1}\sum_{t=1}^{n}\langle X_{tj}, \phi_{jl}\rangle^2$, both of which are useful terms in our further analysis. Moreover, if we choose $\mathbf{\Phi}_1 = (0, \ldots, 0, \lambda_{jl}^{-1/2}\phi_{jl}, 0, \ldots, 0)^{\mathrm{T}}$ and $\mathbf{\Phi}_2 = (0, \ldots, 0, \lambda_{km}^{-1/2}\phi_{km}, 0, \ldots, 0)^{\mathrm{T}}$, an application of (9) and some calculations yield elementwise concentration bounds on $\mathbf{\Sigma}_0$ as stated in the following theorem.

Theorem 2. *Suppose that Conditions 1 and 2 hold. Then there exists some universal constant $\tilde{c} > 0$ such that for any $\eta > 0$ and each $j, k = 1, \ldots, p$,*

$$P\left\{\big\|\widehat{\Sigma}_{jk}^{(0)} - \Sigma_{jk}^{(0)}\big\|_{\mathcal{S}} > 2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\eta\right\} \leqslant 4\exp\Big\{-\tilde{c}n\min(\eta^2, \eta)\Big\}, \quad (10)$$

and

$$P\left\{\big\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\big\|_{\max} > 2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\eta\right\} \leqslant 4p^2\exp\Big\{-\tilde{c}n\min(\eta^2, \eta)\Big\}. \quad (11)$$

In particular, if $n \geqslant \rho^2 \log p$, where ρ is some constant with $\rho > \sqrt{2}\tilde{c}^{-1/2}$, then with probability greater than $1 - 4p^{2-\tilde{c}\rho^2}$, the estimate $\widehat{\mathbf{\Sigma}}_0$ satisfies

$$\big\|\widehat{\mathbf{\Sigma}}_0 - \mathbf{\Sigma}_0\big\|_{\max} \leqslant 2\mathcal{M}_1(f_{\mathbf{X}})\lambda_0\rho\sqrt{\frac{\log p}{n}}. \quad (12)$$

Under an independent setting, the diagonalwise concentration properties of $\widehat{\mathbf{\Sigma}}_0$ were studied in Koltchinskii and Lounici (2017) and Qiao, Guo and James (2019). In particular, Koltchinskii and Lounici (2017) established the concentration bound under operator norm, i.e., with probability greater than $1 - e^{-\delta}$ for all $\delta \geqslant 1$,

$$\|\widehat{\Sigma}_{jj}^{(0)} - \Sigma_{jj}^{(0)}\|_{\mathcal{L}} \lesssim \|\Sigma_{jj}^{(0)}\|_{\mathcal{L}}\left\{\sqrt{\frac{\tilde{r}\big(\Sigma_{jj}^{(0)}\big)}{n}} \vee \frac{\tilde{r}\big(\Sigma_{jj}^{(0)}\big)}{n} \vee \sqrt{\frac{\delta}{n}} \vee \frac{\delta}{n}\right\}, \quad (13)$$

where $\tilde{r}(\Sigma_{jj}^{(0)}) \asymp \int_{\mathcal{U}} \Sigma_{jj}^{(0)}(u, u)du / \|\Sigma_{jj}^{(0)}\|_{\mathcal{L}}$. By contrast, when $\mathbf{X}_1(\cdot), \ldots, \mathbf{X}_n(\cdot)$ are temporally dependent, (10) implies that, with probability greater than $1 - 4e^{-\delta}$,

$$\|\widehat{\Sigma}_{jk}^{(0)} - \Sigma_{jk}^{(0)}\|_{\mathcal{S}} \lesssim \mathcal{M}_1(f_{\mathbf{X}}) \max_j \int_{\mathcal{U}} \Sigma_{jj}^{(0)}(u, u)du \left(\sqrt{\frac{\delta}{n}} \vee \frac{\delta}{n}\right). \quad (14)$$

Note (13) and (14) are both essential concentration bounds of independent interests and share the the common term, $\int_{\mathcal{U}} \Sigma_{jj}^{(0)}(u,u)du$, and the same rate via $(n^{-1}\delta)^{1/2} \vee n^{-1}\delta$ but with different multiplicative terms. However, concentration bounds in (14) under Hilbert–Schmidt norm with free choices of (j,k) play a crutial role in deriving the convergence results for PFCA via Theorems 3 and 4 below in a high-dimensional and dependent setting. In the bounds established in Theorem 2–4, the effects of temporal dependence are commonly captured by $\mathcal{M}_1(f_{\mathbf{X}})$ with larger values yielding a slower convergence rate.

We next present similar concentration bounds on $\widehat{\mathbf{\Sigma}}_h$'s for $h > 0$.

Proposition 1. *Suppose that Conditions 1–2 hold and h is fixed. Then for any given vectors $\mathbf{\Phi}_1$, $\mathbf{\Phi}_2 \in \mathbb{H}_0$ satisfying $\|\mathbf{\Phi}_1\|_0 \vee \|\mathbf{\Phi}_2\|_0 \leqslant k$ with some integer k $(1 \leqslant k \leqslant p)$, there exists some universal constant $c > 0$ such that for any $\eta > 0$,*

$$P\left\{\left|\frac{\big\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}_h)(\mathbf{\Phi}_1)\big\rangle_{\mathbb{H}}}{\big\langle \mathbf{\Phi}_1, \mathbf{\Sigma}_0(\mathbf{\Phi}_1)\big\rangle_{\mathbb{H}}}\right| > 2\mathcal{M}_k(f_{\mathbf{X}})\eta\right\} \leqslant 4\exp\Big\{-cn\min(\eta^2,\eta)\Big\}, \tag{15}$$

and

$$P\left\{\left|\frac{\big\langle \mathbf{\Phi}_1, (\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}_h)(\mathbf{\Phi}_2)\big\rangle_{\mathbb{H}}}{\big\langle \mathbf{\Phi}_1, \mathbf{\Sigma}_0(\mathbf{\Phi}_1)\big\rangle_{\mathbb{H}} + \big\langle \mathbf{\Phi}_2, \mathbf{\Sigma}_0(\mathbf{\Phi}_2)\big\rangle_{\mathbb{H}}}\right| > 2\mathcal{M}_k(f_{\mathbf{X}})\eta\right\} \leqslant 8\exp\Big\{-cn\min(\eta^2,\eta)\Big\}. \tag{16}$$

With the same choices of $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ as those used in applying Theorem 1 to prove Theorem 2, the concentration bounds in (15) and (16) can lead to $\big\|\widehat{\mathbf{\Sigma}}_h - \mathbf{\Sigma}_h\big\|_{\max} = O_P\big\{\mathcal{M}_1(f_{\mathbf{X}})(\log p/n)^{1/2}\big\}$. Moreover, these concentration results are useful to address other important high-dimensional functional time series problems, e.g., high-dimensional functional factor models and non-asymptotic analysis of dynamic FPCA, as discussed in Section A of the Supplementary Material (Guo and Qiao, 2022).

2.3. Rates in elementwise ℓ_∞-norm under a FPCA framework

For each $j = 1,\ldots,p$, we assume that $X_{tj}(\cdot)$ admits the Karhunen-Loève expansion, i.e. $X_{tj}(\cdot) = \sum_{l=1}^{\infty} \xi_{tjl}\phi_{jl}(\cdot)$, which forms the foundation of FPCA. The coefficients $\xi_{tjl} = \langle X_{tj}, \phi_{jl}\rangle$, $l \geqslant 1$, namely *functional principal component* (FPC) scores, correspond to a sequence of random variables with $E(\xi_{tjl}) = 0$, $\mathrm{Var}(\xi_{tjl}) = \lambda_{jl}$ and $\mathrm{Cov}(\xi_{tjl}, \xi_{tjl'}) = 0$ if $l \neq l'$. The eigenpairs $\{(\lambda_{jl}, \phi_{jl})\}_{l\geqslant 1}$ satisfy the eigen-decomposition $\langle \Sigma_{jj}^{(0)}(u,\cdot), \phi_{jl}(\cdot)\rangle = \lambda_{jl}\phi_{jl}(u)$ with $\lambda_{j1} \geqslant \lambda_{j2} \geqslant \cdots$. We say that $X_{tj}(\cdot)$ is d_j-dimensional if $\lambda_{jd_j} \neq 0$ and $\lambda_{j(d_j+1)} = 0$ for some positive integer d_j. If $d_j = \infty$, all the eigenvalues are nonzero and $X_{tj}(\cdot)$ is a truly infinite-dimensional functional object.

To implement FPCA based on realizations $\{X_{1j}(\cdot),\ldots,X_{nj}(\cdot)\}$, we first compute the sample estimator of $\Sigma_{jj}^{(0)}$ by $\widehat{\Sigma}_{jj}^{(0)}(u,v) = n^{-1}\sum_{t=1}^{n} X_{tj}(u)X_{tj}(v)$. Performing an eigen-analysis on $\widehat{\Sigma}_{jj}^{(0)}$, i.e. $\langle \widehat{\Sigma}_{jj}^{(0)}(u,\cdot), \widehat{\phi}_{jl}(\cdot)\rangle = \widehat{\lambda}_{jl}\widehat{\phi}_{jl}(u)$ for $l \geqslant 1$, leads to estimated eigenpairs $(\widehat{\lambda}_{jl}, \widehat{\phi}_{jl})$ and estimated FPC scores $\widehat{\xi}_{tjl} = \langle X_{tj}, \widehat{\phi}_{jl}\rangle$. In the following, we will provide the non-asymptotic analysis of estimated eigenpairs.

2.3.1. Eigenvalues and eigenfunctions

We first impose the following regularity condition.

***Condition* 3.** *For each $j = 1, \ldots, p$, all the nonzero eigenvalues of $\Sigma_{jj}^{(0)}$ are different, i.e. $\lambda_{j1} > \lambda_{j2} > \cdots > 0$, and there exist some positive constants c_0 and $\alpha > 1$ such that $\lambda_{jl} - \lambda_{j(l+1)} \geqslant c_0 l^{-\alpha-1}$ for $l = 1, \ldots, \infty$.*

Condition 3 is standard in functional data analysis literature, see, e.g., Hall and Horowitz (2007) and Kong et al. (2016). The parameter α controls lower bounds for spacings between adjacent eigenvalues with larger values of α allowing tighter eigengaps. This condition also implies that $\lambda_{jl} \geqslant c_0 \alpha^{-1} l^{-\alpha}$ as $\lambda_{jl} = \sum_{k=l}^{\infty} \{\lambda_{jk} - \lambda_{j(k+1)}\} \geqslant c_0 \sum_{k=l}^{\infty} k^{-\alpha-1}$.

In the following theorem, we present relative error bounds on $\{\widehat{\lambda}_{jl}\}$ and $\|\widehat{\phi}_{jl} - \phi_{jl}\|$ in elementwise ℓ_∞ norm, which plays an crucial rule for the further consistency analysis under high-dimensional scaling.

Theorem 3. *Suppose that Conditions 1–3 hold. Let M be a positive integer possibly depending on (n, p). If $n \gtrsim M^{4\alpha+2} \mathcal{M}_1^2(f_{\mathbf{X}}) \log(pM)$, then there exist some positive constants c_1 and c_2 independent of (n, p, M) such that, with probability greater than $1 - c_1(pM)^{-c_2}$, the estimates $\{\widehat{\lambda}_{jl}\}$ and $\{\widehat{\phi}_{jl}(\cdot)\}$ satisfy*

$$\max_{1 \leqslant j \leqslant p, 1 \leqslant l \leqslant M} \left\{ \left| \frac{\widehat{\lambda}_{jl} - \lambda_{jl}}{\lambda_{jl}} \right| + \left\| \frac{\widehat{\phi}_{jl} - \phi_{jl}}{l^{\alpha+1}} \right\| \right\} \lesssim \mathcal{M}_1(f_{\mathbf{X}}) \sqrt{\frac{\log(pM)}{n}}. \tag{17}$$

We provide three remarks for the relative errors of $\{\widehat{\lambda}_{jl}\}$ and $\|\widehat{\phi}_{jl} - \phi_{jl}\|$. First, compared with the non-asymptotic results for the absolute errors of $\{\widehat{\lambda}_{jl}\}$ under an independent setting (Qiao, Guo and James, 2019), Theorem 3 does not require the upper bound condition for eigenvalues. Moreover, when $\mathcal{M}_1(f_{\mathbf{X}})$ remains constant with respect to p, provided that λ_{jl} converges to zero as l grows to infinity, (17) leads to a faster rate of convergence for small eigenvalues, i.e. $|\widehat{\lambda}_{jl} - \lambda_{jl}| = O_P(\lambda_{jl} n^{-1/2})$. Such rate is also sharp in the sense of Jirak (2016). See also Araya Valdivia (2020), which, under an independent setting, established relative concentration bounds on $\{\widehat{\lambda}_{jl}\}$ similar to Lemma 1 in the Supplementary Material (Guo and Qiao, 2022). Second, when each $X_{tj}(\cdot)$ is finite-dimensional with $\lambda_{jl} = 0$ for $l > d_j$, the estimators of zero-eigenvalues enjoy the faster rates due to the property of first order degeneracy (Bathia, Yao and Ziegelmann, 2010). Third, error bounds on $\widehat{\phi}_{jl}$ under ℓ_2 norm are derived using a well-known pathway bound in Lemma 4.3 of Bosq (2000). For finite-dimensional functional objects, such error bounds lead to the optimal $\sqrt{n}$-rate. Under an infinite-dimensional setting, our derived error bounds on $\widehat{\phi}_{jl}(\cdot)$'s can possibly be improved from a probabilistic perspective as long as l diverges. Interestingly, Jirak (2016) established a sharper rate by $\Lambda_{jl}^{-1/2} \|\widehat{\phi}_{jl} - \phi_{jl}\| = O_P(n^{-1/2})$ with $\Lambda_{jl} = \sum_{k \neq l}^{\infty} \frac{\lambda_{jl} \lambda_{jk}}{(\lambda_{jl} - \lambda_{jk})^2}$. In particular, when $\lambda_{jl} \asymp l^{-\alpha}$

with $\alpha > 1$, $\Lambda_{jl} \lesssim l^2$ as $l \to \infty$. We believe that, with the help of techniques used in Jirak (2016), the relevant bounds in Theorems 3 and 4 can be further sharpened. The extra complication, however, will make the theoretical justification of the VFAR estimate in Section 3 more challenging. We leave the development of optimal results as a topic for future research.

Finally, we give two remarks on the parameter M. First, M can be viewed as the truncated dimension of $X_{tj}(\cdot) \approx \sum_{l=1}^{M} \xi_{tjl}\phi_{jl}(\cdot)$. In general, M can depend on j, say M_j, then the right-side of (17) becomes $\mathcal{M}_1(f_{\mathbf{X}})\{\log(\sum_{j=1}^{p} M_j)/n\}^{1/2}$. Second, the sample size lower bound in Theorem 3 implies that $M \lesssim \left[n/\{\mathcal{M}_1^2(f_{\mathbf{X}})\log p\}\right]^{1/(4\alpha+2)}$, which controls the rate that M can grow at most as a function of $n, p, \mathcal{M}_1(f_{\mathbf{X}})$ and α, with larger values of n or smaller values of p or α allowing a larger M.

2.3.2. Covariance between FPC scores

For each $j, k = 1, \ldots, p$, $l, m = 1, 2, \ldots$, and $h = 0, 1, \ldots$, let $\sigma_{jklm}^{(h)} = E(\xi_{tjl}\xi_{(t+h)km})$ and its sample estimator be $\widehat{\sigma}_{jklm}^{(h)} = (n-h)^{-1}\sum_{t=1}^{n-h}\widehat{\xi}_{tjl}\widehat{\xi}_{(t+h)km}$. In the second step of the three-step procedure, our main target is to fit a sparse model based on temporally dependent estimated FPC scores, $\{\widehat{\xi}_{tjl}\}$. To provide the theoretical guarantee for this step, we present the convergence analysis of $\{\widehat{\sigma}_{jklm}^{(h)}\}$ in elementwise ℓ_∞ norm as follows.

Theorem 4. *Suppose that Conditions 1–3 hold and h is fixed. Let M be a positive integer possibly depending on (n, p). If $n \gtrsim M^{4\alpha+2}\mathcal{M}_1^2(f_{\mathbf{X}})\log(pM)$, then there exist some positive constants c_3 and c_4 independent of (n, p, M) such that, with probability greater than $1 - c_3(pM)^{-c_4}$, the estimates $\{\widehat{\sigma}_{jklm}^{(h)}\}$ satisfy*

$$\max_{\substack{1\leqslant j,k\leqslant p\\ 1\leqslant l,m\leqslant M}} \frac{\left|\widehat{\sigma}_{jklm}^{(h)} - \sigma_{jklm}^{(h)}\right|}{(l\vee m)^{\alpha+1}\lambda_{jl}^{1/2}\lambda_{km}^{1/2}} \lesssim \mathcal{M}_1(f_{\mathbf{X}})\sqrt{\frac{\log(pM)}{n}}. \tag{18}$$

We provide three comments here. First, compared with the convergence rate of absolute errors of $\{\widehat{\sigma}_{jklm}^{(0)}\}$ under an independent setting, i.e. $|\widehat{\sigma}_{jklm}^{(0)} - \sigma_{jklm}^{(0)}| = O_P\{(l+m)^{\alpha+1}n^{-1/2}\}$ (Qiao, Guo and James, 2019), we obtain that of scaled errors of $\{\widehat{\sigma}_{jklm}^{(h)}\}$ as $\lambda_{jl}^{-1/2}\lambda_{km}^{-1/2}|\widehat{\sigma}_{jklm}^{(h)} - \sigma_{jklm}^{(h)}| = O_P\{\mathcal{M}_1(f_{\mathbf{X}})(l\vee m)^{\alpha+1}n^{-1/2}\}$, which can more precisely characterize the effect of small eigenvalues on the convergence. Second, in the special case of $j = k$ and $l = m$ with $\sigma_{jjll}^{(0)} = \lambda_{jl}$ and $\widehat{\sigma}_{jjll}^{(0)} = \widehat{\lambda}_{jl}$, the scaled errors of $\{\widehat{\sigma}_{jjll}^{(0)}\}$, i.e. $\lambda_{jl}^{-1}|\widehat{\lambda}_{jl} - \lambda_{jl}|$, would correspond to a faster convergence rate due to (17). Third, if we relax Condition 3 by allowing the parameter α to depend on j, the resulting scaled error rate becomes $\lambda_{jl}^{-1/2}\lambda_{km}^{-1/2}|\widehat{\sigma}_{jklm}^{(h)} - \sigma_{jklm}^{(h)}| = O_P\{\mathcal{M}_1(f_{\mathbf{X}})(l^{\alpha_j+1}\vee m^{\alpha_k+1})n^{-1/2}\}$.

3. Vector functional autoregressive models

Inspired from the standard VAR formulation, we propose a VFAR model of lag L, namely VFAR(L), which is able to characterize linear inter-dependencies in $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$ as follows

$$\mathbf{X}_t(u) = \sum_{h=1}^{L} \int_{\mathcal{U}} \mathbf{A}_h(u,v)\mathbf{X}_{t-h}(v)dv + \boldsymbol{\varepsilon}_t(u), \quad t = L+1,\ldots,n, \tag{19}$$

where $\boldsymbol{\varepsilon}_t(\cdot) = \big(\varepsilon_{t1}(\cdot),\ldots,\varepsilon_{tp}(\cdot)\big)^{\mathrm{T}}$ are independently sampled from a p-dimensional vector of mean zero Gaussian processes, independent of $\mathbf{X}_{t-1}(\cdot), \mathbf{X}_{t-2}(\cdot),\ldots$, and $\mathbf{A}_h = \big(A_{jk}^{(h)}\big)_{1\leqslant j,k\leqslant p}$ is the transition function at lag h with $A_{jk}^{(h)} \in \mathbb{S}$. The structure of transition functions provides insights into the temporal and cross-sectional inter-relationship amongst p functional time series. To make a feasible fit to (19) in a high-dimensional regime, we assume the functional sparsity in $\mathbf{A}_1,\ldots,\mathbf{A}_L$, i.e. most of the components in $\big\{X_{(t-h)k}(\cdot) : h = 1,\ldots,L, k = 1,\ldots,p\big\}$ are unrelated to $X_{tj}(\cdot)$ for $j = 1,\ldots,p$.

Due to the infinite-dimensional nature of functional data, for each j, we take a standard dimension reduction approach through FPCA to approximate $X_{tj}(\cdot)$ using the leading q_j principal components, i.e. $X_{tj}(\cdot) \approx \sum_{l=1}^{q_j} \xi_{tjl}\phi_{jl}(\cdot) = \boldsymbol{\xi}_{tj}^{\mathrm{T}}\boldsymbol{\phi}_j(\cdot)$, where $\boldsymbol{\xi}_{tj} = (\xi_{tj1},\ldots,\xi_{tjq_j})^{\mathrm{T}}$, $\boldsymbol{\phi}_j(\cdot) = \big(\phi_{j1}(\cdot),\ldots,\phi_{jq_j}(\cdot)\big)^{\mathrm{T}}$ and q_j is chosen data-adaptively to provide a reasonable approximation to the trajectory $X_{tj}(\cdot)$.

Once the FPCA has been performed for each $X_{tj}(\cdot)$, we let $\mathbf{V}_j^{(h)} \in \mathbb{R}^{(n-L)\times q_j}$ with its row vectors given by $\boldsymbol{\xi}_{(L+1-h)j},\ldots,\boldsymbol{\xi}_{(n-h)j}$ and $\boldsymbol{\Psi}_{jk}^{(h)} = \int_{\mathcal{U}}\int_{\mathcal{U}} \boldsymbol{\phi}_k(v)A_{jk}^{(h)}(u,v)\boldsymbol{\phi}_j(u)^{\mathrm{T}}dudv \in \mathbb{R}^{q_k\times q_j}$. Then further derivations in Section F.1 of the Supplementary Material (Guo and Qiao, 2022) lead to the matrix representation of (19) as

$$\mathbf{V}_j^{(0)} = \sum_{h=1}^{L}\sum_{k=1}^{p} \mathbf{V}_k^{(h)}\boldsymbol{\Psi}_{jk}^{(h)} + \mathbf{R}_j + \mathbf{E}_j, \quad j = 1,\ldots,p, \tag{20}$$

where $\mathbf{R}_j$ and $\mathbf{E}_j$ are $(n-L)\times q_j$ error matrices whose row vectors are formed by the truncation and random errors, respectively. Hence, we can rely on the block sparsity pattern in $\{\boldsymbol{\Psi}_{jk}^{(h)} : h = 1,\ldots,L, j,k = 1,\ldots,p\}$ to recover the functional sparsity structure in $\{A_{jk}^{(h)} : h = 1,\ldots,L, j,k = 1,\ldots,p\}$. It is also worth noting that (20) can be viewed as a $\big(\sum_{j=1}^{p} q_j\big)$-dimensional VAR(L) model with the error vector consisting of both the truncation and random errors.

3.1. Estimation procedure

The estimation procedure proceeds in the following three steps.

Step 1. We perform FPCA based on observed curves, $X_{1j}(\cdot),\ldots,X_{nj}(\cdot)$ and thus obtain estimated eigenfunctions $\widehat{\boldsymbol{\phi}}_j(\cdot) = \big(\widehat{\phi}_{jl}(\cdot),\ldots,\widehat{\phi}_{jq_j}(\cdot)\big)^{\mathrm{T}}$ and FPC scores $\widehat{\boldsymbol{\xi}}_{tj} = \big(\widehat{\xi}_{tj1},\ldots,\widehat{\xi}_{tjq_j}\big)^{\mathrm{T}}$ for each j. See

Step 2. Motivated from the matrix representation of a VFAR(L) model in (20), we propose a penalized *least squares* (LS) approach, which minimizes the following optimization criterion over $\{\mathbf{\Psi}_{jk}^{(h)} : h = 1, \ldots, L, k = 1, \ldots, p\}$:

$$\frac{1}{2}\left\|\widehat{\mathbf{V}}_j^{(0)} - \sum_{h=1}^{L}\sum_{k=1}^{p}\widehat{\mathbf{V}}_k^{(h)}\mathbf{\Psi}_{jk}^{(h)}\right\|_F^2 + \gamma_{nj}\sum_{h=1}^{L}\sum_{k=1}^{p}\left\|\widehat{\mathbf{V}}_k^{(h)}\mathbf{\Psi}_{jk}^{(h)}\right\|_F, \tag{21}$$

where $\widehat{\mathbf{V}}_j^{(h)}$, the estimate of $\mathbf{V}_j^{(h)}$, is a $(n-L)\times q_j$ matrix with its i-th row vector given by $\widehat{\boldsymbol{\xi}}_{(L+i-h)j}$ for $i = 1, \ldots, (n-L)$, and $\gamma_{nj} \geqslant 0$ is a regularization parameter. The ℓ_1/ℓ_2 type of standardized group lasso penalty (Simon and Tibshirani, 2012) in (21) forces the elements of $\mathbf{\Psi}_{jk}^{(h)}$ to either all be zero or non-zero. Potentially, one could modify (21) by adding an unstandardized group lasso penalty (Yuan and Lin, 2006) in the form of $\gamma_{nj}\sum_{h=1}^{L}\sum_{k=1}^{p}\left\|\mathbf{\Psi}_{jk}^{(h)}\right\|_F$ to produce the block sparsity in $\{\mathbf{\Psi}_{jk}^{(h)}\}$. However, orthonormalization within each group would correspond to the uniformly most powerful invariant test for inclusion of a group, hence we use a standardized group lasso penalty here. In Section G.3 of the Supplementary Material (Guo and Qiao, 2022), we develop a block version of fast iterative shrinkage-thresholding algorithm (FISTA), which mirrors recent gradient-based techniques (Beck and Teboulle, 2009; O'Donoghue and Candès, 2015), to solve the optimization problem in (21) with the solution given by $\{\widehat{\mathbf{\Psi}}_{jk}^{(h)}\}$. The proposed block FISTA algorithm is easy to implement and converges very fast, thus is suitable for solving large-scale optimization problems.

Step 3. Finally, we recover the functional sparse estimates of elements in $\{A_{jk}^{(h)}\}$ by the block sparse estimates in $\{\widehat{\mathbf{\Psi}}_{jk}^{(h)}\}$ via

$$\widehat{A}_{jk}^{(h)}(u,v) = \widehat{\boldsymbol{\phi}}_k(v)^{\mathrm{T}}\widehat{\mathbf{\Psi}}_{jk}^{(h)}\widehat{\boldsymbol{\phi}}_j(u), \quad h = 1, \ldots, L,\ j,k = 1, \ldots, p. \tag{22}$$

3.2. Functional network Granger causality

In this section, we extend the definition of network Granger causality (NGC) under a VAR framework (Lütkepohl, 2005) to the functional domain and then use the extended definition under our proposed VFAR framework to understand the causal relationship in $\{\mathbf{X}_t(\cdot)\}_{t\in\mathbb{Z}}$

In an analogy to the NGC formulation, a functional NGC (FNGC) model consists of p nodes, one for each functional variable, and a number of edges with directions connecting a subset of nodes. Specifically, functional times series of $\{X_{tk}(\cdot)\}_{t\in\mathbb{Z}}$ is defined to be Granger causal for that of $\{X_{tj}(\cdot)\}_{t\in\mathbb{Z}}$ or equivalently there is an edge from node k to node j, if $A_{jk}^{(h)}(u,v) \neq 0$ for some $(u,v) \in \mathcal{U}^2$ or $h \in \{1, \ldots, L\}$. Then our proposed FNGC model can be represented by a directed graph $G = (V, E)$ with vertex set $V = \{1, \ldots, p\}$ and edge set

$$E = \left\{(k,j) : A_{jk}^{(h)}(u,v) \neq 0 \text{ for some } (u,v) \in \mathcal{U}^2 \text{ or } h \in \{1, \ldots, L\}, (j,k) \in V^2\right\}.$$

It is worth noting that, at lag h, $\|A_{jk}^{(h)}\|_{\mathcal{S}}$ can be viewed as explaining the global Granger-type casual impact of $X_k(\cdot)$ on $X_j(\cdot)$, while $A_{jk}^{(h)}(u,v)$ itself accounts for the local Granger-type casual impact of $X_k(v)$ on $X_j(u)$. To explore the FNGC structure and the direction of influence from one node to the other, we need to develop an approach to estimate E, i.e. identifying the locations of non-zero entries in $\widehat{\mathbf{A}}_1, \ldots, \widehat{\mathbf{A}}_L$ under the Hilbert-Schmidt norm, the details of which are presented in Section 3.1.

3.3. Theoretical properties

According to Section F.2 of the Supplementary Material (Guo and Qiao, 2022), all VFAR(L) models in (19) can be reformulated as a VFAR(1) model. Without loss of generality, we consider a VFAR(1) model in the form of

$$\mathbf{X}_t(u) = \int_{\mathcal{U}} \mathbf{A}(u,v)\mathbf{X}_{t-1}(v)dv + \boldsymbol{\varepsilon}_t(u), \quad t = 2, \ldots, n, \ u \in \mathcal{U}.$$

To simplify our notation in this section, we focus on the setting where q_j's are the same across $j = 1, \ldots, p$. However, our theoretical results extend naturally to the more general setting. In our empirical studies, we select different q_j's, see Section G.1 of the Supplementary Material (Guo and Qiao, 2022) for details. Let $\widehat{\mathbf{Z}} = (\widehat{\mathbf{V}}_1^{(1)}, \ldots, \widehat{\mathbf{V}}_p^{(1)})^{\mathrm{T}} \in \mathbb{R}^{(n-1)\times pq}$, $\boldsymbol{\Psi}_j = \big((\boldsymbol{\psi}_{j1}^{(1)})^{\mathrm{T}}, \ldots, (\boldsymbol{\psi}_{jp}^{(1)})^{\mathrm{T}}\big)^{\mathrm{T}} \in \mathbb{R}^{pq\times q}$. and $\widehat{\mathbf{D}} = \mathrm{diag}\big(\widehat{\mathbf{D}}_1, \ldots, \widehat{\mathbf{D}}_p\big) \in \mathbb{R}^{pq\times pq}$, where $\widehat{\mathbf{D}}_k = \big\{(n-1)^{-1}(\widehat{\mathbf{V}}_k^{(1)})^{\mathrm{T}}\widehat{\mathbf{V}}_k^{(1)}\big\}^{1/2} \in \mathbb{R}^{q\times q}$ for $k = 1, \ldots, p$. Then minimizing (21) over $\boldsymbol{\Psi}_j \in \mathbb{R}^{pq\times q}$ is equivalent to minimizing the following criterion over $\mathbf{B}_j \in \mathbb{R}^{pq\times q}$,

$$-\langle\langle \widehat{\mathbf{Y}}_j, \mathbf{B}_j \rangle\rangle + \frac{1}{2}\langle\langle \mathbf{B}_j, \widehat{\boldsymbol{\Gamma}}\mathbf{B}_j \rangle\rangle + \gamma_{nj}\|\mathbf{B}_j\|_1^{(q)}, \tag{23}$$

where $\widehat{\mathbf{Y}}_j = (n-1)^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{V}}_j^{(0)}$, $\widehat{\boldsymbol{\Gamma}} = (n-1)^{-1}\widehat{\mathbf{D}}^{-1}\widehat{\mathbf{Z}}^{\mathrm{T}}\widehat{\mathbf{Z}}\widehat{\mathbf{D}}^{-1}$. Let $\widehat{\mathbf{B}}_j$ be the minimizer of (23), then $\widehat{\boldsymbol{\Psi}}_j = \widehat{\mathbf{D}}^{-1}\widehat{\mathbf{B}}_j$ with its k-th row block given by $\widehat{\boldsymbol{\Psi}}_{jk}$ and $\widehat{\mathbf{A}} = (\widehat{A}_{jk})$ with its (j,k)-th entry, $\widehat{A}_{jk}(u,v) = \widehat{\boldsymbol{\phi}}_k(v)^{\mathrm{T}}\widehat{\boldsymbol{\Psi}}_{jk}\widehat{\boldsymbol{\phi}}_j(u)$ for $j,k = 1, \ldots, p$ and $(u,v) \in \mathcal{U}^2$.

Before imposing the condition on the entries of $\mathbf{A} = (A_{jk})$, we begin with some notation. For the j-th row of $\mathbf{A}$, we denote the set of non-zero functions by $S_j = \big\{k \in \{1, \ldots, p\} : \|A_{jk}\|_{\mathcal{S}} \neq 0\big\}$ and its cardinality by $s_j = |S_j|$ for $j = 1, \ldots, p$. We also denote the maximum degree or row-wise cardinality by $s = \max_j s_j$ (possibly depends on n and p), corresponding to the maximum number of non-zero functions in any row of $\mathbf{A}$. For a sparse VFAR(L) model in (19) with $s_{jh} = \sum_{k=1}^{p} I(\|A_{jk}^{(h)}\|_{\mathcal{S}} \neq 0)$ for $j = 1, \ldots, p$ and $h = 1, \ldots, L$, according to Section F.2 of the Supplementary Material (Guo and Qiao, 2022), the equivalent sparse VFAR(1) model in (F.47) has the maximum degree $s = \max_j(\sum_h s_{jh}) \lesssim \max_{j,h} s_{jh}$ assuming a fixed lag order L.

***Condition* 4.** *For each $j = 1, \ldots, p$ and $k \in S_j$, $A_{jk}(u,v) = \sum_{l,m=1}^{\infty} a_{jklm}\phi_{jl}(u)\phi_{km}(v)$ and there exist some positive constants $\beta > \alpha/2 + 1$ and μ_{jk} such that $|a_{jklm}| \leqslant \mu_{jk}(l+m)^{-\beta-1/2}$ for $l, m \geqslant 1$.*

For each (j,k), the basis with respect to which coefficients $\{a_{jklm}\}_{l,m\geqslant 1}$ are defined is determined by $\{\phi_{jl}(\cdot)\}_{l\geqslant 1}$ and $\{\phi_{km}(\cdot)\}_{m\geqslant 1}$ The parameter β in Condition 4 determines the decay rate of the upper bounds for coefficients $\{a_{jklm}\}_{l,m\geqslant 1}$ and hence characterizes the degree of smoothness in $\{A_{jk}\}$, with larger values of β yielding smoother functions. See also Hall and Horowitz (2007) and Kong et al. (2016) for similar smoothness conditions in functional linear models.

We next establish the consistency of the VFAR estimate based on the sufficient conditions in Conditions 5–7 below. To be specific, we first establish an upper bound on $\|\widehat{\mathbf{A}}-\mathbf{A}\|_\infty$ in Theorem 5 below. Using the convergence results in Section 2.3, we then show that all VFAR models in (19) satisfy Conditions 5–7 with high probability through Propositions 2–4 below. As a consequence, the error bound in Theorem 5 holds with high probability.

Before stating these conditions, we give some notation. For a block matrix $\mathbf{B}=(\mathbf{B}_{jk})\in\mathbb{R}^{p_1q\times p_2q}$ with its (j,k)-th block $\mathbf{B}_{jk}\in\mathbb{R}^{q\times q}$, we define its q-block versions of Frobenius norm, elementwise ℓ_∞ norm and matrix ℓ_1 norm by $\|\mathbf{B}\|_{\mathrm{F}}=\big(\sum_{j,k}\|\mathbf{B}_{jk}\|_{\mathrm{F}}^2\big)^{1/2}$, $\|\mathbf{B}\|_{\max}^{(q)}=\max_{j,k}\|\mathbf{B}_{jk}\|_{\mathrm{F}}$ and $\|\mathbf{B}\|_1^{(q)}=\max_k\sum_j\|\mathbf{B}_{jk}\|_{\mathrm{F}}$, respectively.

***Condition* 5.** *The symmetric matrix $\widehat{\mathbf{\Gamma}}\in\mathbb{R}^{pq\times pq}$ satisfies the restricted eigenvalue condition with tolerance $\tau_1>0$ and curvature $\tau_2>0$ if*

$$\boldsymbol{\theta}^{\mathrm{T}}\widehat{\mathbf{\Gamma}}\boldsymbol{\theta}\geqslant\tau_2\|\boldsymbol{\theta}\|^2-\tau_1\|\boldsymbol{\theta}\|_1^2\quad\forall\boldsymbol{\theta}\in\mathbb{R}^{pq}. \tag{24}$$

Condition 5 comes from a class of conditions commonly referred to as restricted eigenvalue (RE) conditions in the lasso literature (Bickel, Ritov and Tsybakov, 2009; Loh and Wainwright, 2012). Intuitively speaking, Condition 5 implies that $\boldsymbol{\theta}^{\mathrm{T}}\widehat{\mathbf{\Gamma}}\boldsymbol{\theta}/\|\boldsymbol{\theta}\|^2$ is strictly positive as long as $\|\boldsymbol{\theta}\|_1$ is not large relative to $\|\boldsymbol{\theta}\|$. Denote the estimation error by $\mathbf{\Delta}_j=\widehat{\mathbf{B}}_j-\mathbf{B}_j$, this condition ensures that $\langle\langle\mathbf{\Delta}_j,\widehat{\mathbf{\Gamma}}\mathbf{\Delta}_j\rangle\rangle\geqslant\tau_2\|\mathbf{\Delta}_j\|_{\mathrm{F}}^2/2$ if $\tau_2\geqslant 32\tau_1q^2s$, as stated in Theorem 5 below.

***Condition* 6.** *There exist some positive constants C_λ and C_ϕ independent of (n,p,q) such that*

$$\begin{aligned}&\max_{1\leqslant j\leqslant p,1\leqslant l\leqslant q}\left|\frac{\widehat{\lambda}_{jl}^{-1/2}-\lambda_{jl}^{-1/2}}{\lambda_{jl}^{-1/2}}\right|\leqslant C_\lambda\mathcal{M}_1(f_{\mathbf{X}})\sqrt{\frac{\log(pq)}{n}},\\&\max_{1\leqslant j\leqslant p,1\leqslant l\leqslant q}\|\widehat{\phi}_{jl}-\phi_{jl}\|\leqslant C_\phi\mathcal{M}_1(f_{\mathbf{X}})q^{\alpha+1}\sqrt{\frac{\log(pq)}{n}}.\end{aligned} \tag{25}$$

***Condition* 7.** *There exits some positive constant C_E independent of (n,p,q) such that*

$$\left\|\widehat{\mathbf{Y}}_j-\widehat{\mathbf{\Gamma}}\mathbf{B}_j\right\|_{\max}^{(q)}\leqslant C_E\mathcal{M}_1(f_{\mathbf{X}})s_j\left\{q^{\alpha+2}\sqrt{\frac{\log(pq)}{n}}+q^{-\beta+1}\right\},\ j=1,\ldots,p. \tag{26}$$

Condition 6 and 7 are two deviation conditions, which guarantee the good behaviours of relevant estimated terms by controlling their deviation bounds. Specifically, Condition 7 ensures that $\widehat{\mathbf{Y}}_j$ and $\widehat{\mathbf{\Gamma}}$ are nicely concentrated around their population versions. See also similar deviation conditions in the lasso literature (Loh and Wainwright, 2012; Basu and Michailidis, 2015).

We are now ready to present the theorem on the convergence rate of the VFAR estimate.

Theorem 5. *Suppose that Conditions 1–7 hold with* $\tau_2 \geqslant 32\tau_1 q^2 s$. *Then, for any regularization parameter,* $\gamma_{nj} \geqslant 2C_E \mathcal{M}_1(f_{\mathbf{X}}) s_j \{q^{\alpha+2}(\log(pq)/n)^{1/2} + q^{-\beta+1}\}$, $\gamma_n = \max_j \gamma_{nj}$ *and* $q^{\alpha/2} s \gamma_n \to 0$ *as* $n, p, q \to \infty$, *any minimizer* $\widehat{\mathbf{B}}_j$ *of (23) satisfies*

$$\|\widehat{\mathbf{B}}_j - \mathbf{B}_j\|_F \leqslant \frac{24 s_j^{1/2} \gamma_{nj}}{\tau_2}, \quad \|\widehat{\mathbf{B}}_j - \mathbf{B}_j\|_1^{(q)} \leqslant \frac{96 s_j \gamma_{nj}}{\tau_2} \quad \text{for } j = 1, \ldots, p,$$

and the estimated transition function, $\widehat{\mathbf{A}}$, *satisfies*

$$\|\widehat{\mathbf{A}} - \mathbf{A}\|_\infty \leqslant \frac{96\alpha^{1/2} q^{\alpha/2} s \gamma_n}{c_0^{1/2} \tau_2} \Big\{1 + o(1)\Big\}. \tag{27}$$

The convergence rate of $\widehat{\mathbf{A}}$ under functional matrix ℓ_∞ norm is governed by dimensionality parameters (n, p, s) and internal parameters $(\mathcal{M}_1(f_{\mathbf{X}}), q, \tau_1, \tau_2, \alpha, \beta)$. We provide three remarks for the error bound in (27). First, it is easy to see that larger values of α (tighter eigengaps) or $\mathcal{M}_1(f_{\mathbf{X}})$ (less stable process of $\{\mathbf{X}_t(\cdot)\}$) or s (denser structure in $\mathbf{A}$) yield a slower convergence rate, while enlarging β or τ_2 will increase the entrywise smoothness in $\mathbf{A}$ or the curvature of the RE condition, respectively, thus resulting in a faster rate. Second, the convergence rate consists of two terms corresponding to the variance-bias tradeoff as commonly considered in nonparametric statistics. Specifically, the variance is of the order $O_P\big[\mathcal{M}_1(f_{\mathbf{X}}) s^2 q^{(3\alpha+4)/2} \{\log(pq)/n\}^{1/2}\big]$ and the bias term is bounded by $O\big\{\mathcal{M}_1(f_{\mathbf{X}}) s^2 q^{(\alpha-2\beta+2)/2}\big\}$. To balance both terms, we can choose an optimal q satisfying $\log(pq) q^{2\alpha+2\beta+2} \asymp n$, which leads to $q \asymp \{n/\log(p \vee n)\}^{1/(2\alpha+2\beta+2)}$. Third, when each $X_{tj}(\cdot)$ is finite-dimensional, although the truncation step is no longer required, the FPC scores still need to be estimated. The resulting convergence rate becomes $O_P\{\mathcal{M}_1(f_{\mathbf{X}}) s^2 (\log p/n)^{1/2}\}$, which is slightly different from that of the high-dimensional VAR estimate in Basu and Michailidis (2015).

Finally, we turn to verify that, when $\mathbf{X}_1(\cdot), \ldots, \mathbf{X}_n(\cdot)$ are drawn from model (19), Conditions 5–7 are satisfied with high probability as stated in the following Propositions 2–4. Before presenting these propositions, we list two regularity conditions.

***Condition* 8.** *For* $\mathbf{\Sigma}_0 = \big(\Sigma_{jk}^{(0)}\big)_{1 \leqslant j,k \leqslant p}$, *we denote by* $\mathbf{D}_0 = diag(\Sigma_{11}^{(0)}, \ldots, \Sigma_{pp}^{(0)})$ *the diagonal function. The infimum* $\underline{\mu}$ *of the functional Rayleigh quotient of* $\mathbf{\Sigma}_0$ *relative to* $\mathbf{D}_0$, *defined as follows, is bounded below by zero, i.e.*

$$\underline{\mu} = \inf_{\mathbf{\Phi} \in \bar{\mathbb{H}}_0} \frac{\big\langle \mathbf{\Phi}, \mathbf{\Sigma}_0(\mathbf{\Phi}) \big\rangle_{\mathbb{H}}}{\big\langle \mathbf{\Phi}, \mathbf{D}_0(\mathbf{\Phi}) \big\rangle_{\mathbb{H}}} > 0, \tag{28}$$

where $\bar{\mathbb{H}}_0 = \{\mathbf{\Phi} \in \mathbb{H} : \langle \mathbf{\Phi}, \mathbf{D}_0(\mathbf{\Phi}) \rangle_{\mathbb{H}} \in (0, \infty)\}$.

***Condition* 9.** *The sample size n satisfies the bound $n \gtrsim q^{4\alpha+2}\mathcal{M}_1^2(f_{\mathbf{X}})\log(pq)$.*

In Condition 8, the lower bound on $\underline{\mu}$, chosen as the curvature τ_2 in the proof of Proposition 2, can be understood as requiring the minimum eigenvalue of the correlation function for $\mathbf{X}_t(\cdot)$ to be bounded below by zero. Specially, if $X_{tj}(\cdot)$ is d_j-dimensional for $j = 1, \ldots, p$ it is easy to show that $\underline{\mu}$ reduces to the minimum eigenvalue of the correlation matrix for the $(\sum_j d_j)$-dimensional vector, $\boldsymbol{\xi}_t = (\xi_{t11}, \ldots, \xi_{t1d_1}, \ldots, \xi_{tp1}, \ldots, \xi_{tpd_p})^{\mathrm{T}}$. Condition 9 is required here due to its presence in Theorems 3 and 4.

Proposition 2. *(Verify Condition 5) Suppose that Conditions 2–3 and 8–9 hold. Then there exist three positive constants C_Γ, c_5 and c_6 independent of (n, p, q) such that*

$$\boldsymbol{\theta}^T \hat{\mathbf{\Gamma}} \boldsymbol{\theta} \geqslant \underline{\mu}\|\boldsymbol{\theta}\|_2^2 - C_\Gamma \mathcal{M}_1(f_{\mathbf{X}}) q^{\alpha+1}\sqrt{\frac{\log(pq)}{n}}\|\boldsymbol{\theta}\|_1^2$$

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

Proposition 3. *(Verify Condition 6) Suppose that Conditions 2–3 and 9 hold. Then there exist four positive constants C_ϕ, C_λ c_5 and c_6 independent of (n, p, q) such that (25) holds with probability greater than $1 - c_5(pq)^{-c_6}$.*

Proposition 4. *(Verify Condition 7) Suppose that Conditions 2–4 and 9 hold. Then there exist three positive constants C_E, c_5 and c_6 independent of (n, p, q) such that (26) holds with probability greater than $1 - c_5(pq)^{-c_6}$.*

Propositions 2–4 can be proved by applying the convergence results in Theorems 3 and 4. With suitable choices of common constants c_5 and c_6 in Propositions 2–4, we can show that the joint probability for the three events corresponding to the non-asymptotic upper bounds in (17) and (18) is greater than $1 - c_5(pq)^{-c_6}$. Consequently, with probability greater than $1 - c_5(pq)^{-c_6}$, the estimate $\widehat{\mathbf{A}}$ satisfies the error bound in (27).

3.4. Simulation studies

In this section, we conduct a number of simulations to compare the finite-sample performance of our proposed method to potential competitors.

In each simulated scenario, we generate functional variables by $X_{tj}(u) = \mathbf{s}(u)^{\mathrm{T}}\boldsymbol{\theta}_{tj}$ for $j = 1, \ldots, p$ and $u \in \mathcal{U} = [0, 1]$, where $\mathbf{s}(\cdot)$ is a 5-dimensional Fourier basis function and each $\boldsymbol{\theta}_t = (\boldsymbol{\theta}_{t1}^{\mathrm{T}}, \ldots, \boldsymbol{\theta}_{tp}^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{5p}$ is generated from a stationary VAR(1) process, $\boldsymbol{\theta}_t = \mathbf{B}\boldsymbol{\theta}_{t-1} + \boldsymbol{\eta}_t$, with block transition matrix $\mathbf{B} \in \mathbb{R}^{5p\times 5p}$, whose (j, k)-th block is given by $\mathbf{B}_{jk}$ for $j, k = 1, \ldots, p$ and innovations $\boldsymbol{\eta}_t$'s being independently sampled from $N(\mathbf{0}, \mathbf{I}_{5p})$. To mimic a real data setting, we generate observed values, W_{tjs}, with measurement errors,

$W_{tjs} = X_{tj}(u_s) + e_{tjs}$, from $T = 50$ equally spaced time points, $0 = u_1, \ldots, u_T = 1$ with errors e_{tjs}'s being randomly sampled from $N(0, 0.5^2)$. In our simulations, we generate $n = 100$ or $n = 200$ observations of $p = 40$ or $p = 80$ functional variables, and we aim to show that, although our method is developed for fully observed functional time series, it still works well even for the dense design with measurement error. It is worth noting that, as discussed in Section 1, the VFAR estimation is naturally a very high-dimensional problem. For example, to fit a VFAR(1) model under our most "low-dimensional" setting with $p = 40$ and $n = 200$, we need to estimate $40^2 \times 5^2 = 40,000$ parameters based on only 200 observations.

According to Section F.3 of the Supplementary Material (Guo and Qiao, 2022), $\mathbf{X}_t(\cdot)$ follows from a VFAR(1) model in (4), where $\varepsilon_{tj}(u) = \mathbf{s}(u)^{\mathrm{T}}\boldsymbol{\eta}_{tj}$ and autocoefficient functions satisfy $A_{jk}(u, v) = \mathbf{s}(u)^{\mathrm{T}}\mathbf{B}_{jk}\mathbf{s}(v)$ for $j, k = 1, \ldots, p, (u, v) \in \mathcal{U}^2$. Hence, the functional sparsity structure in $\mathbf{A}$ can be characterized by the block sparsity pattern in $\mathbf{B}$. In the following, we consider two different scenarios to generate $\mathbf{B}$.

(i) **Block sparse**. We generate a block sparse $\mathbf{B}$ without any special structure. Specifically, we generate $\mathbf{B}_{jk} = w_{jk}\mathbf{C}_{jk}$ for $j, k = 1 \ldots, p$, where entries in $\mathbf{C}_{jk}$ are randomly sampled from $N(0, 1)$ and w_{jk}'s are generated from $\{0, 1\}$ under the constraint of $\sum_{k=1}^{p} w_{jk} = 5$ for each j, such that the same row-wise cardinality for $\mathbf{B}$ can be produced in a blockwise fashion. To guarantee the stationarity of $\{\mathbf{X}_t(\cdot)\}$, we rescale $\mathbf{B}$ by $\iota\mathbf{B}/\rho(\mathbf{B})$, where ι is generated from Unif[0.5,1].

(ii) **Block banded**. We generate a block banded $\mathbf{B}$, with entries in $\mathbf{B}_{jk}$ being randomly sampled from $N(0, 1)$ if $|j - k| \leqslant 2$, and being zero at other locations. $\mathbf{B}$ is then rescaled as described in (i).

For each $j = 1, \ldots, p$, we perform regularized FPCA (Ramsay and Silverman, 2005) on observations $\{W_{tjs}\}$ to obtain smoothed estimates of $\phi_{jl}(\cdot)$'s, and use 5-fold cross-validation to choose q_j and the smoothing parameter, the details of which are presented in Sections G.1 and G.2 of the Supplementary Material (Guo and Qiao, 2022). Typically $q_j = 4$, 5 or 6 are selected in our simulations. To choose the regularization parameters γ_{nj}'s, there exist a number possible methods such as AIC/BIC and cross-validation. While the third one is computationally intensive and remains largely unexplored in the literature for serially dependent functional observations, we take an approach motivated by the information criterion for sparse additive models (Voorman, Shojaie and Witten, 2014). For each j, our proposed information criterion is

$$\mathrm{IC}_j(\gamma_{nj}) = n\log\Big\{\big\|\widehat{\mathbf{V}}_j^{(0)} - \sum_{h=1}^{L}\sum_{k=1}^{p}\widehat{\mathbf{V}}_k^{(h)}\widehat{\boldsymbol{\Phi}}_{jk}^{(h)}(\gamma_{nj})\big\|_F^2\Big\} + \kappa_n \mathrm{df}_j(\gamma_{nj}), \tag{29}$$

where $\kappa_n = 2$ and $\log n$ correspond to AIC and BIC, respectively, and $\mathrm{df}_j(\gamma_{nj})$ is the effective degrees of freedom used in fitting (21) with

$$\mathrm{df}_j(\gamma_{nj}) = \sum_{h=1}^{L}\sum_{k=1}^{p}\left\{I\Big((h,k) : ||\widehat{\boldsymbol{\Phi}}_{jk}^{(h)}(\gamma_{nj})||_F \neq 0\Big) + (q_j q_k - 1)\frac{||\widehat{\mathbf{V}}_k^{(h)}\widehat{\boldsymbol{\Phi}}_{jk}^{(h)}(\gamma_{nj})||_F^2}{||\widehat{\mathbf{V}}_k^{(h)}\widehat{\boldsymbol{\Phi}}_{jk}^{(h)}(\gamma_{nj})||_F^2 + \gamma_{nj}}\right\}. \tag{30}$$

Table 1. The mean and standard error (in parentheses) of AUROCs over 100 simulation runs. The best values are in bold font.

		Model (i)			Model (ii)		
n	p	ℓ_1/ℓ_2-LS$_a$	ℓ_1/ℓ_2-LS$_2$	ℓ_1-LS$_1$	ℓ_1/ℓ_2-LS$_a$	ℓ_1/ℓ_2-LS$_2$	ℓ_1-LS$_1$
100	40	**0.840(0.018)**	0.690(0.019)	0.591(0.023)	**0.872**(0.016)	0.719(0.022)	0.609(0.024)
	80	**0.829(0.015)**	0.682(0.017)	0.585(0.015)	**0.869**(0.014)	0.714(0.017)	0.600(0.017)
200	40	**0.951(0.011)**	0.764(0.020)	0.616(0.021)	**0.971**(0.006)	0.795(0.018)	0.639(0.023)
	80	**0.948(0.010)**	0.770(0.017)	0.626(0.015)	**0.969**(0.005)	0.799(0.014)	0.644(0.015)

We compare our proposed ℓ_1/ℓ_2-penalized LS estimate using all selected principal components, namely ℓ_1/ℓ_2-LS$_\mathrm{a}$, to its two competitors. One method, ℓ_1/ℓ_2-LS$_2$, relies on minimizing ℓ_1/ℓ_2-penalized LS based on the first two estimated principal components, which capture partial curve information. The other approach, ℓ_1-LS$_1$, projects the functional data into a standard format by computing the first estimated FPC score and then implements an ℓ_1 regularization approach (Basu and Michailidis, 2015) for the VAR estimation on this data. We examine the sample performance of three approaches, ℓ_1/ℓ_2-LS$_\mathrm{a}$, ℓ_1/ℓ_2-LS$_2$, and ℓ_1-LS$_1$ in terms of model selection consistency and estimation accuracy.

- **Model selection**. We plot the true positive rates against false positive rates, defined as $\frac{\#\{(j,k):||\widehat{A}_{jk}^{(\boldsymbol{\gamma}_n)}||_{\mathcal{S}}\neq 0 \text{ and } ||A_{jk}||_{\mathcal{S}}\neq 0\}}{\#\{(j,k):||A_{jk}||_{\mathcal{S}}\neq 0\}}$ and $\frac{\#\{(j,k):||\widehat{A}_{jk}^{(\boldsymbol{\gamma}_n)}||_{\mathcal{S}}\neq 0 \text{ and } ||A_{jk}||_{\mathcal{S}}= 0\}}{\#\{(j,k):||A_{jk}||_{\mathcal{S}}= 0\}}$, respectively, over a sequence of $\boldsymbol{\gamma}_n = (\gamma_{n1}, \ldots, \gamma_{np})$ values to produce a ROC curve. We compute the area under the ROC curve (AUROC) with values closer to one indicating better performance in recovering the functional sparsity structure in $\mathbf{A}$.
- **Estimation error**. We calculate the relative estimation accuracy for $\widehat{\mathbf{A}}$ by $\|\widehat{\mathbf{A}} - \mathbf{A}\|_\mathrm{F}/\|\mathbf{A}\|_\mathrm{F}$, where $\widehat{\mathbf{A}}$ is the regularized estimate based on the optimal regularization parameters selected by minimizing AICs or BICs in (29).

To investigate the support recovery consistency, we report the average AUROCs of three comparison methods under both model settings in Table 1. In all simulations, we observe that ℓ_1/ℓ_2-LS$_\mathrm{a}$ with most of curve information being captured, provides highly significant improvements over its two competitors and ℓ_1-LS$_1$ gives the worst results. To evaluate the estimation accuracy, Table 2 presents numerical results of relative errors of different regularized estimates. We also report the performance of the LS estimate in the oracle case, where we know locations of non-zero entries of $\mathbf{A}$ in advance. Several conclusions can be drawn from Table 2. First, in all scenarios, the proposed BIC-based ℓ_1/ℓ_2-LS$_\mathrm{a}$ method provides the highest estimation accuracy among all the comparison methods. Second, the performance of AIC-based methods severally deteriorate in comparison with their BIC-based counterparts. Given the high-dimensional, dependent and functional natural of the model structure, computing the effective degrees of freedom in (30) leads to a very challenging task and requires further investigation. In practice, with some prior knowledge about the targeted graph density under a FNGC framework, one can choose the values of γ_{nj} that result in the graph with a desired sparsity level. Third, LS$_\mathrm{oracle}$ estimates give much worse results than BIC-based regularized estimates. This is not surprising, since even in the "large n, small p" scenario, e.g. $n = 100, p = 40$ for

Table 2. The mean and standard error (in parentheses) of relative estimation errors of $\widehat{\mathbf{A}}$ over 100 simulation runs. The best values are in bold font.

Model	(n,p)	γ_n	ℓ_1/ℓ_2-LS$_{\mathrm{a}}$	ℓ_1/ℓ_2-LS$_2$	ℓ_1-LS$_1$	LS$_{\mathrm{oracle}}$
(i)	(100,40)	AIC	1.783(0.044)	1.455(0.032)	1.047(0.008)	1.483(0.044)
		BIC	**0.971(0.007)**	0.997(0.002)	1.002(0.002)	
	(100,80)	AIC	1.546(0.037)	1.576(0.030)	1.124(0.013)	1.530(0.050)
		BIC	**0.991(0.003)**	0.999(0.001)	1.002(0.001)	
	(200,40)	AIC	1.419(0.035)	1.159(0.016)	1.020(0.004)	0.990(0.029)
		BIC	**0.850(0.012)**	0.989(0.003)	0.999(0.002)	
	(200,80)	AIC	1.544(0.045)	1.350(0.022)	1.032(0.004)	1.016(0.033)
		BIC	**0.915(0.013)**	0.994(0.002)	1.000(0.001)	
(ii)	(100,40)	AIC	1.679(0.040)	1.400(0.026)	1.039(0.008)	1.363(0.039)
		BIC	**0.957(0.009)**	0.995(0.002)	1.000(0.002)	
	(100,80)	AIC	1.435(0.028)	1.489(0.020)	1.105(0.012)	1.383(0.036)
		BIC	**0.983(0.004)**	0.998(0.001)	1.001(0.001)	
	(200,40)	AIC	1.329(0.028)	1.128(0.013)	1.014(0.004)	0.909(0.024)
		BIC	**0.824(0.009)**	0.985(0.003)	0.998(0.002)	
	(200,80)	AIC	1.428(0.027)	1.296(0.013)	1.024(0.004)	0.926(0.019)
		BIC	**0.881(0.010)**	0.990(0.002)	0.998(0.001)	

Model (i), implementing LS requires estimating $5 \times 5^2 = 125$ parameters based on only 100 observations, which intrinsically results in a high-dimensional estimation problem.

3.5. Real data analysis

In this section, we apply our proposed method to a public financial dataset, which was downloaded from `https://wrds-web.wharton.upenn.edu/wrds`. The dataset consists of high-frequency observations of prices for a collection of S&P100 stocks from $n = 251$ trading days in year 2017. We removed several stocks for which the data are not available during the observational period. See Table 3 in the Supplementary Material (Guo and Qiao, 2022) for tickers, company names and classified sectors of the inclusive $p = 98$ stocks. We then obtained data at a sampling frequency one-minute per datum such that the impact of microstructure noise is reduced (Zhang, Mykland and Aït-Sahalia, 2005). The daily trading period (9:30-16:00) is thus converted to $\mathcal{U} = [0, T]$ with $T = 390$ minutes. Let $P_{tj}(u_k)$ $(t = 1, \ldots, n, j = 1, \ldots, p, k = 1, \ldots, T)$ be the price of the j-th stock at intraday time u_k on the t-th trading day. We denote the cumulative intraday return (CIDR) trajectory, in percentage, by $r_{tj}(u_k) = 100[\log\{P_{tj}(u_k)\} - \log\{P_{tj}(u_1)\}]$ (Horváth, Kokoszka and Rice, 2014). This transformation not only guarantees the shape of CIDR curves nearly the same as original daily price curves, but also makes the assumption of stationarity for curves more plausible. According to the definition, CIDR curves always start from zero enforcing level stationary and the logarithm helps reduce potential scale inflation. See Horváth, Kokoszka and Rice (2014) for an empirical study on testing the stationary of CIDR curves.

Our main target is to construct a directed graph under a FNGC framework and to

display the Granger-type casual relationships among intraday returns of different stocks. To achieve this goal, we first center all the series of $\{r_{tj}(\cdot)\}$ about their empirical means and then apply the three-step approach to fit a sparse VFAR(1) model on the demeaned curves. In the first step, we implement regularized FPCA to smooth these curves and therefore the effects of microstructure noise are further reduced to be almost negligible. See Sections G.1-G.2 and G.3 of the Supplementary Material (Guo and Qiao, 2022) for details on regularized FPCA with the selection of relevant tuning parameters and block FISTA algorithm, respectively.

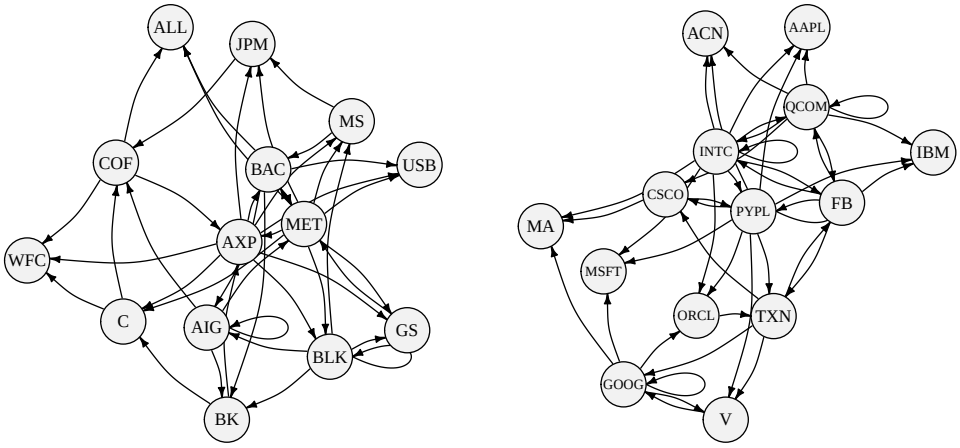

Figure 2. *Left and right graphs plot the directed networks with indegree=3 for $p = 14$ stocks in the financial and IT sectors, respectively.*

To better visualize and interpret the network, we focus on $p = 14$ stocks in the financial and information technology (IT) sections, respectively and set the row-wise sparsity to 3/14, i.e. each node receives connections from 3 (indegree) out of 14 nodes. A more systematic method for determining the network sparsity level, e.g. via a significance testing, needs to be developed. Figure 2 displays two directed networks based on the identified sparsity structures in $\widehat{\mathbf{A}}$ (estimated transition function) for stocks in two sectors. It suggests that "MET" (Metlife) and "INTC" (Intel) together with "PYPL" (PayPal), placed in the center of each network, provide the lowest levels of column-wise sparsity in the financial and IT sectors, respectively, thus resulting in highest Granger-type causal impacts on all the stocks in terms of their CIDR curves. Moreover, we consider $p = 28$ stocks in both financial and IT sectors. Setting the row-wise sparsity to 1/7, we plot a larger directed graph in Figure 3. We observe that more IT companies, e.g. PayPal, Qualcomm and Intel, have relatively higher causal impacts. Interestingly, PayPal, as a leading financial technology (FinTech) company belonging to both financial and IT sectors leads to the the highest causal influence on others. See also Section H.2 of the Supplementary

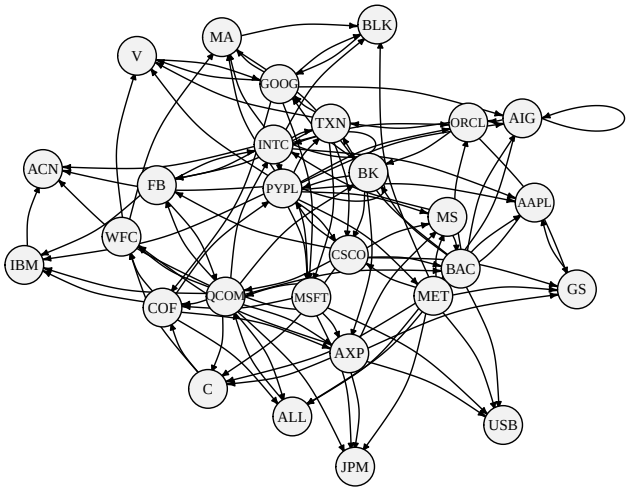

Figure 3. *The directed graph with indegree=4 for $p = 28$ stocks in the financial and IT sectors.*

Material (Guo and Qiao, 2022) for additional empirical analysis.

Acknowledgements

We are grateful to the Editor, the Associate Editor and two referees for their insightful comments, which have led to significant improvement of our paper. Shaojun Guo was partially supported by the National Natural Science Foundation of China (No. 11771447).

Supplementary Material

This supplementary material contains the discussion of several potential extensions, examples for infinite-dimensional functional data satisfying Condition 2, all technical proofs, derivations of functional stability measure for the illustrative VFAR(1) example, some derivations for VFAR models, details of the algorithms to fit sparse VFAR models and additional empirical results.

References

Araya Valdivia, E. (2020). Relative concentration bounds for the spectrum of kernel matrices. *Preprint. Available at arXiv:1812.02108v8.*

Aue, A., Norinho, D. D. and Hörmann, S. (2015). On the prediction of stationary functional time series. *J. Amer. Statist. Assoc.* **110** 378–392. MR3338510

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

Basu, S., Shojaie, A. and Michailidis, G. (2015). Network Granger causality with inherent grouping structure. *J. Mach. Learn. Res.* **16** 417–453. MR3335801

Bathia, N., Yao, Q. and Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series. *Ann. Statist.* **38** 3352–3386. MR2766855

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

Bickel, P. J., Ritov, Y. and Tsybakov, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469

Billio, M., Casarin, R. and Rossini, L. (2019). Bayesian nonparametric sparse VAR models. *J. Econometrics* **212** 97–115. MR3994009

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

Cho, H., Goude, Y., Brossat, X. and Yao, Q. (2013). Modeling and forecasting daily electricity load curves: a hybrid approach. *J. Amer. Statist. Assoc.* **108** 7–21. MR3174599

Fan, Y., James, G. M. and Radchenko, P. (2015). Functional additive regression. *Ann. Statist.* **43** 2296–2325. MR3396986

Ghosh, S., Khare, K. and Michailidis, G. (2019). High-dimensional posterior consistency in Bayesian vector autoregressive models. *J. Amer. Statist. Assoc.* **114** 735–748. MR3963176

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

Guo, S., Wang, Y. and Yao, Q. (2016). High-dimensional and banded vector autoregressions. *Biometrika* **103** 889–903. MR3620446

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

Han, F., Lu, H. and Liu, H. (2015). A direct estimation of high dimensional stationary vector autoregressions. *J. Mach. Learn. Res.* **16** 3115–3150. MR3450535

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

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Ann. Statist.* **38** 1845–1884. MR2662361

Horváth, L., Kokoszka, P. and Rice, G. (2014). Testing stationarity of functional time series. *J. Econometrics* **179** 66–82. MR3153649

Jirak, M. (2016). Optimal eigen expansions and uniform bounds. *Probab. Theory Related Fields* **166** 753–799. MR3568039

Kock, A. B. and Callot, L. (2015). Oracle inequalities for high dimensional vector autoregressions. *J. Econometrics* **186** 325–344. MR3343790

Kokoszka, P. and Reimherr, M. (2013). Determining the order of the functional autoregressive model. *J. Time Series Anal.* **34** 116–129. MR3008019

Koltchinskii, V. and Lounici, K. (2017). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli* **23** 110 – 133. MR3556768

Kong, D., Xue, K., Yao, F. and Zhang, H. H. (2016). Partially functional linear regression in high dimensions. *Biometrika* **103** 147–159. MR3465827

Li, D., Robinson, P. M. and Shang, H. L. (2020). Long-range dependent curve time series. *J. Amer. Statist. Assoc.* **115** 957–971. MR4107692

Li, B. and Solea, E. (2018). A nonparametric graphical model for functional data with application to brain networks based on fMRI. *J. Amer. Statist. Assoc.* **113** 1637–1655. MR3902235

Li, R., Cui, L., Li, J., Zhao, A., Fu, H., Wu, Y., Zhang, L., L., K. and Chen, J. (2017). Spatial and temporal variation of particulate matter and gaseous pollutants in China during 2014-2016. *Atmospheric environment* **161** 235-246.

Loh, P.-L. and Wainwright, M. J. (2012). High-dimensional regression with noisy and missing data: provable guarantees with nonconvexity. *Ann. Statist.* **40** 1637–1664. MR3015038

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer-Verlag, Berlin. MR2172368

Müller, H.-G., Sen, R. and Stadtmüller, U. (2011). Functional data analysis for volatility. *J. Econometrics* **165** 233–245. MR2846647

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

Panaretos, V. M. and Tavakoli, S. (2013). Fourier analysis of stationary time series in function space. *Ann. Statist.* **41** 568–603. MR3099114

Qiao, X., Guo, S. and James, G. M. (2019). Functional graphical models. *J. Amer. Statist. Assoc.* **114** 211–222. MR3941249

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

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

Simon, N. and Tibshirani, R. (2012). Standardization and the group Lasso penalty. *Statist. Sinica* **22** 983–1001. MR2987480

Sun, Y., Li, Y., Kuceyeski, A. and Basu, S. (2018). Large spectral density matrix estimation by thresholding. *arXiv:1812.00532*.

Voorman, A., Shojaie, A. and Witten, D. (2014). Graph estimation with joint additive models. *Biometrika* **101** 85–101. MR3180659

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

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 49–67. MR2212574

Zhang, L., Mykland, P. A. and Aït-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100** 1394–1411. MR2236450