

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of
Business

Lee Kong Chian School of Business

7-2020

A review study of functional autoregressive models with application to energy forecasting

Ying CHEN

Thorsten KOCH

Kian Guan LIM

Xiaofei XU

Nazgul ZAKIYEVA


Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research



Part of the [Management Sciences and Quantitative Methods Commons](#), and the [Statistics and Probability Commons](#)

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

A review study of functional autoregressive models with application to energy forecasting

Ying Chen^{1,2} | Thorsten Koch^{3,4} | Kian Guan Lim⁵ | Xiaofei Xu¹ |
Nazgul Zakiyeva⁶ 

¹Department of Mathematics, National University of Singapore, Singapore

²Risk Management Institute, National University of Singapore, Singapore

³Chair of Software and Algorithms for Discrete Optimization, Technische Universität Berlin, Berlin, Germany

⁴Mathematical Optimization Department, Zuse Institute Berlin, Berlin, Germany

⁵Lee Kong Chian School of Business, Singapore Management University, Singapore

⁶Department of Statistics and Applied Probability, National University of Singapore, Singapore

Correspondence

Nazgul Zakiyeva, Department of Statistics and Applied Probability, National University of Singapore, Singapore.
Email: nazgul.zakiyeva@u.nus.edu

Abstract

In this data-rich era, it is essential to develop advanced techniques to analyze and understand large amounts of data and extract the underlying information in a flexible way. We provide a review study on the state-of-the-art statistical time series models for univariate and multivariate functional data with serial dependence. In particular, we review functional autoregressive (FAR) models and their variations under different scenarios. The models include the classic FAR model under stationarity; the FARX and pFAR model dealing with multiple exogenous functional variables and large-scale mixed-type exogenous variables; the vector FAR model and common functional principal component technique to handle multiple dimensional functional time series; and the warping FAR, varying coefficient-FAR and adaptive FAR models to handle seasonal variations, slow varying effects and the more challenging cases of structural changes or breaks respectively. We present the models' setup and detail the estimation procedure. We discuss the models' applicability and illustrate the numerical performance using real-world data of high-resolution natural gas flows in the high-pressure gas pipeline network of Germany. We conduct 1-day and 14-days-ahead out-of-sample forecasts of the daily gas flow curves. We observe that the functional time series models generally produce stable out-of-sample forecast accuracy.

This article is categorized under:

Statistical Models > Semiparametric Models

Data: Types and Structure > Time Series, Stochastic Processes, and Functional Data

KEYWORDS

energy forecast, functional autoregressive modeling, functional time series, sieve estimation

1 | INTRODUCTION

Analyzing big data provides both advantages and challenges in various research fields, including energy, economics, medicine, biology, and finance. Big data has played an increasingly important role in the investigation of scientific questions and provided rich information for statistical inference. On the other hand, complex dependence structures,

large sample sizes and high dimensions in big data bring another level of statistical and computational difficulties that cannot be solved with traditional statistical methods. It is essential to build advanced and flexible methods to analyze complex structures in big data and understand the underlying dynamics.

This study aims to provide a review on recently developed models and methodologies for handling big data sets that can be naturally represented as a series of dependent functions or curves. In contrast to other reviews on functional data which have focused more on theory or computations of independent functional data (see e.g., Ullah & Finch, 2013; Wang, Chiou, & Müller, 2016), this review focuses on modeling and prediction for serially dependent functional data. As illustration, we demonstrate the implementation of the reviewed models using real data on natural gas demand and supply in the German gas transmission network.

In the high-dimensional area, curves arise naturally in many cases such as natural gas flow curves (Chen, Chua, & Koch, 2018), electricity price curves (Chen & Li, 2017; Chen, Marron, & Zhang, 2019), growth data in Berkeley Growth Study (Tuddenham, 1954), and Canadian weather data (Programm, 1982; Ramsay & Dalzell, 1991). A good review of functional data examples is given in Ramsay and Silverman (2002). Recent advances in functional data analysis (FDA) have enabled efficient methods for analyzing big data with certain features. In FDA, the observed multiple time series are considered as discrete observations of a continuous curve or function. With a natural and parsimonious functional representation, the high-dimensional data is converted to a series of curves and can be analyzed with improved efficiency and accuracy. We refer to Müller and Stadtmüller (2005), Ferraty and Vieu (2006), Ramsay and Silverman (2002), Ramsay and Silverman (2005) and Wang et al. (2016) for details on analyzing independent functional data.

The functional data collected over time naturally exhibit serial dependence, which is inconsistent to the IID assumption widely adopted in for example, functional regressions. This motivates the study of functional time series analysis, which focuses on understanding the serial dependence among the curves, modeling their dynamics over time, and conducting statistical prediction. Functional time series analysis is still a fast developing field. The autoregressive Hilbertian (ARH) process pioneered by Bosq (1991), also called functional autoregressive (FAR) model under Hilbert space, is likely the most popular pioneering work and plays an important role in this context. As a natural extension of the scalar and vector-valued AR process (Brockwell & Davis, 1991) to the infinite-dimensional space, FAR is mathematically and statistically flexible to be used in practice for modeling and prediction of continuous time series (Hörmann, 2013). The FAR model considers serial correlation between the functional observations and models it as a linear operator on the functional space. Bosq (2000) provided a comprehensive theory of the general linear functional time series including the limit theorem of linear process in both Hilbert and Banach spaces. Mas (2007) completed the theoretical study of the FAR model by addressing the issue of weak convergence for estimates from the model, which proved that traditional facts about weak convergence in nonparametric models appear. Dedecker and Merlevède (2002, 2003) proved a conditional central limit theorem for functional linear processes under mild assumptions. Hörmann, Kokoszka, et al. (2010) set up the theoretical framework to investigate the serial dependence in functional time series modeling. See also Mas and Menneteau (2003), Menneteau (2005), Bosq (2007) and Farindon (2011) for more theoretical studies of functional time series.

In the FAR modeling framework, several estimation approaches have been proposed. The most popular estimation methods include the functional Yule–Walker (FYW) estimation and the sieve maximum likelihood (SML) estimation. Later, based on FYW estimation of Bosq (1991), a nonparametric kernel estimator was introduced by Besse, Cardot, and Stephenson (2000). Didericksen, Kokoszka, and Zhang (2012) evaluated several FYW estimation methods and showed that the Bosq (2000) method performed the best overall. Moreover, Antoniadis and Sapatinas (2003) and Antoniadis, Paparoditis, and Sapatinas (2006) applied the wavelet-kernel approach and addressed the inverse problem of FYW, which can not satisfy the positive semidefiniteness condition in the covariance operator, while Kargin and Onatski (2008) adopted the method of predictive factor decomposition. Alternatively, Mourid and Bensmain (2006) proposed using sieves methods initially introduced by Grenander (1981) to reduce the number of dimensions and obtain a closed-form maximum likelihood estimator (MLE). Liu, Xiao, and Chen (2016) derived the estimation procedure of the convolutional FAR model using splines and sieve methods.

In addition to the advancement of estimation approaches, there has been a vivid development of FAR modeling. Damon and Guillas (2002) proposed the FAR model with exogenous variables (FARX) to incorporate the effect of exogenous functional covariates on the serially dependent curves; see also Chen et al. (2018). To investigate the effect of large-dimensional functional and scalar covariates on the serially dependent functional time series, Chen, Koch, and Xu (2019) proposed a partial FAR (pFAR) model. Beyond the univariate FAR models, Chen, Chua, and Härdle (2019) proposed the vector FAR (VFAR) model describing the joint dynamics of the multiple functional time series with serial cross-dependence in a unified framework. When the dimension becomes too large, the VFAR model faces an

overparametrization problem. Factor modeling can be employed to extract low-dimensional factors retaining the dependency information. The relatively low dimensional factors can represent the original high-dimensional curves data well. The factors also inherit the serial dependence that can be further modeled by simple time series models for example, in an autoregressive framework. See Zhang, Chen, Klotz, and Lim (2017) for the model on multi-country yield curves analysis.

The aforementioned works assume constant serial dependence over time, that is, time homogeneity. However, the linear stationarity assumption is not practical for unstable real-world time series data (Xu, Li, & Chen, 2017). According to the literature, considering nonstationary time series methods can be a possible way to address this issue. Horváth, Kokoszka, and Rice (2014) developed a general methodology for stationarity testing, and Hörmann, Kokoszka, and Nisol (2018) derived several methods for periodicity testing in a functional time series. In addition, Kosiorowski, Mielczarek, Rydlewski, and Snarska (2014) adopted a moving functional median method, and Xu et al. (2017) proposed a varying coefficient FAR (VC-FAR) model. In the latter case, serial dependencies among curves are assumed to change over time gradually, and time-dependent coefficients are obtained using a local regression technique. Chen and Li (2017) presented an adaptive FAR (AFAR) model that allows both smooth structural changes and abrupt breaks. The time-dependent parameters of the model are calculated in a data-driven way.

This paper provides a review study of different FAR models and statistical approaches in the functional domain. This review focuses on modeling and methodology for the serially dependent functional data, and demonstrate the implementation of the reviewed models using real data, while the other reviews in Ullah and Finch (2013) and Wang et al. (2016) focused more on theory or computations of independent functional data without real data applications. Specifically, we start with a stationary FAR model that describes the serial dependence of curves varying with time, which was initially introduced by Bosq (2000). We introduce smoothing and estimation techniques and provide useful literature references. Then we extend to the FARX model with exogenous variables by Chen et al. (2018) and the partial FAR (pFAR) model with large-scale exogenous variables of mixed data type by Chen, Koch, and Xu (2019). While the classical FAR models only consider the dependence of curves on their own lagged values, the FARX and pFAR models by incorporating exogenous information can facilitate richer dynamics and also causal interpretation. Next, we will discuss the vector FAR (VFAR) model for multiple-dimensional functional time series that exhibit lead-lag cross dependence by Chen, Chua, and Härdle (2019). When the dimension of functional time series increases, we will show how to build up factors via the common functional principal component (CFPC) analysis, see Zhang et al. (2017). Lastly, we will discuss the scenarios with seasonal variations, slow varying effects and more challenging, unforeseeable structural changes or breaks, in the framework of the warping FAR (WFAR) by Chen, Marron, and Zhang (2019), varying-coefficient FAR (VC-FAR) by Xu et al. (2017) and adaptive FAR (AFAR) respectively Chen and Li (2017). Table 1 compares the features of the various models to be reviewed in the paper. We will detail the model setup and estimation procedure as follows.

The rest of the paper is outlined as follows. In Section 2 we present the classical FAR based models under various scenarios. We detail the model setup and elaborate on the estimation procedures. Section 3 describes the functional

TABLE 1 Features of the reviewed models

Features	FAR	FPC	FARX	pFAR	VFAR	CFPC	WFAR	VC-FAR	AFAR
Serial dependence	✓	✓	✓	✓	✓	✓	✓	✓	✓
Multivariate functional series					✓	✓			
Seasonality							✓		
Causal relation to exogenous variables			✓	✓					
High-dimensionality (in # of series or predictors)				✓		✓			
Nonstationarity								✓	✓

Abbreviations: AFAR, adaptive FAR model; CFPC, common functional principal component; FAR, functional autoregressive; FARX, FAR model with exogenous variables; FPC, functional principle component; pFAR, partial FAR; VFAR, vector FAR; VC-FAR, varying coefficient-FAR; WFAR, warping FAR.

data—the natural gas flow curves and the empirical features. Section 4 contains the numerical results of various models, and we compare and discuss these results. Section 5 contains the conclusions.

2 | MODELS

In this section, we review the recently developed FAR models that can be used for different scenarios.

2.1 | Functional time series and expansion

The classical estimation techniques such as the ML are not privileged in the FAR models possibly due to the infinite dimensional parameter space. It is therefore necessary to conduct regularization before estimation. One popular idea is to decompose the infinite-dimensional functional series to finite parameter space via orthonormal basis expansion with the information loss controlled. In fact, functional series could be roughly classified into two categories: periodic and nonperiodic. Fourier basis (Chen & Li, 2017; Xu et al., 2017) are usually applied to fit periodic functions, and the spline or B-splines (Liu et al., 2016) are more suitable for nonperiodic smooth functions. Other popular basis functions include functional eigenbasis (Kong, Xue, Yao, & Zhang, 2016) and Gaussian basis functions (Matsui, Kawano, & Konishi, 2009; Matsui & Konishi, 2011). In this paper, we consider three popular functional expansion methods, that is, Fourier expansion, spline expansion and functional eigenbasis expansion. Readers can refer to Ramsay and Silverman (2005) for a good review of functional basis expansion.

Let \mathcal{H} be a real and separable Hilbert space with a countable and orthonormal basis $\{e_k, k \in \mathbb{Z}\}$ and the inner product $\langle \cdot, \cdot \rangle$ induces the norm $\| \cdot \|$. Define a sequence of random curves $\{Y_t(\tau), t \in \mathbb{Z}\}$ in \mathcal{H} over a time domain $\tau \in \mathcal{T}$. Here $Y_t(\tau) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathcal{H}, \mathcal{B}_{\mathcal{H}})$ is a measurable function, where $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space and $\mathcal{B}_{\mathcal{H}}$ is a Borel σ -field defined on the functional Hilbert space \mathcal{H} . Without loss of generality, we set $\mathcal{T} = [0, 1]$, and consider L^2 separable Hilbert space, that is, we consider $Y_t(\tau) \in L^2([0, 1])$. Function $Y_t(\tau)$ can be expanded as a linear combination of a set of functional building blocks Φ_k , called basis functions:

$$Y_t(\tau) = \sum_{k=1}^{\infty} c_k \Phi_k(\tau)$$

The parameters c_1, c_2, \dots, c_K are the coefficients of the expansion.

Fourier expansion: The Fourier series is:

$$\Phi_0 = I_{[0,1]}, \Phi_{2k}(\tau) = \sqrt{2} \cos 2\pi k \tau, \Phi_{2k-1}(\tau) = \sqrt{2} \sin 2\pi k \tau,$$

for $k = 1, 2, \dots, \infty$, and we obtain Fourier basis expansion for $Y_t(\tau)$:

$$Y_t(\tau) = a_{t,0} + \sum_{k=1}^{\infty} [b_{t,k} \Phi_{2k-1}(\tau) + a_{t,k} \Phi_{2k}(\tau)], \quad (1)$$

where $a_{t,0}, b_{t,k}$ and $a_{t,k}$ denote the constant, cosine and sine Fourier basis coefficients.

Spline expansion: Alternatively, we can use the B-spline basis functions:

$$B_{j,m}(\tau) = \frac{\tau - w_j}{w_{j+m-1} - w_j} B_{j,m-1}(\tau) + \frac{w_{j+m} - \tau}{w_{j+m} - w_{j+1}} B_{j+1,m-1}(\tau), m \geq 2, \quad (2)$$

where m denotes the order of a spline, $w_1 \leq \dots \leq w_{J+m}$ denote the sequence of knots, and

$$B_{j,1}(\tau) = \begin{cases} 1, & \text{if } w_j \leq \tau < w_{j+1} \\ 0, & \text{otherwise.} \end{cases}$$

In the spline expansion, we obtain:

$$Y_t(\tau) = \sum_{k=1}^{\infty} a_{t,k} B_{k,m}, \quad (3)$$

where $a_{t,k}$ is the B-spline coefficient for the function $Y_t(\tau)$.

Functional eigenbasis expansion: Functional eigenbasis expansion is data-driven without a predetermined basis and is widely used in functional principal component analysis. It is popular when there is no clue or prior knowledge on important features and unsupervised learning is preferred.

Suppose $Y_t(\tau)$ has mean function $\mathbb{E}(Y_t(\tau)) = \mu(\tau)$ and covariance function $\nu(\pi, \tau): L^2([0, 1]) \rightarrow \mathbb{R}$, $\pi, \tau \in [0, 1]$, given by

$$\nu(\pi, \tau) = \text{Cov}(Y_t(\pi), Y_t(\tau)) = \mathbb{E}\{(Y_t(\pi) - \mu(\pi))(Y_t(\tau) - \mu(\tau))\}.$$

The eigenfunctions of $\nu(\pi, \tau)$ are called functional principle components (FPCs), obtained by decomposing the covariance operator of the curves under independence. We define the covariance operator $\Upsilon: L^2(\mathcal{T}) \rightarrow L^2(\mathcal{T})$ with the covariance function ν and obtain the FPCs denoted as $\phi_j, j \geq 1$ by

$$\left(\Upsilon\phi_j\right)(\tau) = \lambda_j\phi_j(\tau), \quad (4)$$

where the corresponding eigenvalue of the covariance operator Υ is denoted as λ_j . The eigenfunctions, ϕ_j , are assumed to be orthonormal with $\int \phi_j(\tau)\phi_m(\tau)d\tau = 0$, for all $m \neq j$ and normalized to unit norm. The statistical approach is to obtain the orthonormal functions ϕ_1, ϕ_2, \dots with the maximal variances of the principal scores $\text{Var}(\xi_j)$, where the functional principal component scores at time t defined as

$$\xi_{t,j} = \int_{\mathcal{T}} [Y_t(\pi) - \mu(\pi)]\phi_j(\pi)d\pi. \quad (5)$$

Given the orthogonality of ϕ_j and ϕ_l for $j \neq l$, it follows that $\mathbb{E}(\xi_j\xi_l) = 0$ for $j \neq l$. We have $\mathbb{E}(\xi_j^2) = \lambda_j$ and $\mathbb{E}(\xi_j) = 0$. Then the function $Y_t(\tau)$ has the Karhunen–Loève expansion as:

$$Y_t(\tau) = \mu(\tau) + \sum_{j=1}^{\infty} \xi_{tj}\phi_j(\tau), \quad (6)$$

where the coefficients ξ_{tj} are the functional principal component scores at time t inheriting the serial dependence in $Y_t(\tau)$.

2.2 | FAR model under stationarity

In this section, we present the FAR model and its derivation in a stationary framework, where the first two moments of the dynamic dependence are constant over time.

2.2.1 | The FAR model

Recall a sequence of stationary random curves $\{Y_t(\tau), t \in \mathbb{Z}\}$ in $L^2([0, 1])$ defined in Section 2.1. Suppose that there exists a Hilbert–Schmidt operator $\rho(\cdot)$ from \mathcal{H} to \mathcal{H} which is bounded linear and admits a representation

$\rho(\cdot) = \sum_{j=1}^{\infty} e_j \langle \cdot, v_j \rangle f_j$, where $\{v_j\}$ and $\{f_j\}$ are two orthonormal bases of \mathcal{H} . The parameter e_j is a real sequence satisfying $\sum_{j=1}^{\infty} e_j^2 < \infty$ and converging to zero. The FAR model of order p , i.e., FAR(p), is defined as:

$$Y_t(\tau) - \mu(\tau) = \sum_{j=1}^p \rho_j (Y_{t-j}(\tau) - \mu(\tau)) + \varepsilon_t(\tau),$$

where $\mu(\tau)$ is the mean function of $Y_t(\tau)$, Y_{t-j} denotes j th lag of curve Y_t and $\varepsilon_t(\tau)$ is a strong \mathcal{H} -white noise with zero mean and finite second moment $E \|\varepsilon_t(\tau)\|^2 < \infty$. It is common to have order 1 in functional time series analysis. Meanwhile, we refer to Kokoszka and Reimherr (2013) for a multiple testing procedure to determine order p if FAR(1) is not adequate.

An implementable form of the Hilbert–Schmidt operator is to use a convolution kernel operator to represent ρ , see Pumo (1998), Mokhtari and Mourid (2003) and Mas and Pumo (2018). We obtain:

$$Y_t(\tau) - \mu(\tau) = \sum_{j=1}^p \int_0^1 \phi_j(\tau - \pi) [Y_{t-j}(\pi) - \mu(\pi)] d\pi + \varepsilon_t(\tau), \quad (7)$$

where $\phi_j(\tau) \in L^2([0, 1])$ is the kernel function of the operator specifying the serial dependence of the curve on its own past value. The kernel ϕ is usually taken to be an even function with $\|\phi\|_2 < 1$ and $\|\cdot\|_2$ denotes the standard L^2 norm. Xu et al. (2017) and Chen, Koch, and Xu (2019) relaxed the assumption on kernel for more flexibility. Yet both studies support the choice in real data analysis as the contribution coming from the odd function part is trivial. Model (7) is only one kind of implementable form of the ARH process of order p (ARH(p)) defined in Bosq (2000) which is associated with more general forms of operators in \mathcal{H} . In this review study, we consider this form (7) with kernel of form $\phi(\tau)$ due to its popularity in the literature. We refer to Bosq (2000) for detailed generalizations.

The functional terms of observations, kernels and innovations in (7) can be expanded with the functional basis introduced in Section 2.1. As an illustration, we detail the expansion of model (7) using Fourier basis and spline basis below. A similar estimation procedure can be applied to other FAR based models covered in this review study.

Fourier basis expansion for the FAR: Due to the orthonormal properties of the basis, it is easy to derive a closed form solution for the FAR models with the Fourier basis expansions. We apply the trigonometric basis functions in $L^2([0, 1])$ defined in (2.1) to expand the functional terms in (7):

$$\begin{aligned} Y_t(\tau) &= a_{t,0} + \sum_{k=1}^{\infty} [b_{t,k} \Phi_{2k-1}(\tau) + a_{t,k} \Phi_{2k}(\tau)], \delta(\tau) = \omega_0 + \sum_{k=1}^{\infty} [\eta_k \Phi_{2k-1}(\tau) + \omega_k \Phi_{2k}(\tau)], \\ \phi_j(\tau) &= c_{j,0} + \sum_{k=1}^{\infty} [d_{j,k} \Phi_{2k-1}(\tau) + c_{j,k} \Phi_{2k}(\tau)], \varepsilon_t(\tau) = \varepsilon_{t,0} + \sum_{k=1}^{\infty} [e_{t,k} \Phi_{2k-1}(\tau) + \varepsilon_{t,k} \Phi_{2k}(\tau)], \end{aligned}$$

where $\delta(\tau)$ is the intercept function defined as $\delta(\tau) = \mu(\tau) - \sum_{j=1}^p \int_0^1 \phi_j(\tau - \pi) \mu(\pi) d\pi$. $a_{t,0}$, $a_{t,k}$, $b_{t,k}$ denote the constant, cosine, and sine Fourier basis coefficients corresponding to the observed curves $Y_t(\tau)$, and thus have known values; $c_{j,0}$, $c_{j,k}$, $d_{j,k}$ are associated with the unknown serial dependence function $\phi_j(\tau)$; ω_0 , ω_k , η_k are related to the intercept function $\delta(\tau)$, and $\varepsilon_{t,0}$, $\varepsilon_{t,k}$, $e_{t,k}$ are that corresponding to the innovation $\varepsilon_t(\tau)$, see Chen et al. (2018) and Chen, Koch, and Xu (2019).

Plugging Fourier expansions into (7) and rearranging the equations, we obtain the recursive relationship of the Fourier coefficients for all $k = 1, \dots, \infty$,

$$\begin{aligned} a_{t,0} &= \omega_0 + \sum_{j=1}^p c_{j,0} a_{t-j,0} + \varepsilon_{t,0}, \\ a_{t,k} &= \omega_k + \sum_{j=1}^p \frac{1}{\sqrt{2}} c_{j,k} a_{t-j,k} - \sum_{j=1}^p \frac{1}{\sqrt{2}} d_{j,k} b_{t-j,k} + \varepsilon_{t,k}, \\ b_{t,k} &= \eta_k + \sum_{j=1}^p \frac{1}{\sqrt{2}} c_{j,k} b_{t-j,k} + \sum_{j=1}^p \frac{1}{\sqrt{2}} d_{j,k} a_{t-j,k} + e_{t,k}. \end{aligned} \quad (8)$$

The functional observation and functional parameters are now characterized by their discrete Fourier coefficients. The estimation problem of the FAR model has now been converted to a problem of estimating the Fourier coefficients.

Spline basis expansion for the FAR: Splines are generally applicable for smooth curves, but it is relatively difficult to obtain an explicit solution. With the B-spline basis functions in $L^2([0, 1])$ defined in (2) we expand the functional terms in (7) as follows:

$$Y_t(\tau) = \sum_{k=1}^{\infty} a_{t,k} B_{k,m}, \quad \delta(\tau) = \sum_{k=1}^{\infty} \eta_k B_{k,m}, \quad \phi_j(\tau) = \sum_{k=1}^{\infty} c_{j,k} B_{k,m}, \quad \varepsilon_t(\tau) = \sum_{k=1}^{\infty} \epsilon_{t,k} B_{k,m},$$

where $a_{t,k}$, η_k , $c_{j,k}$ and $\epsilon_{t,k}$ are, respectively, the B-spline coefficients for the observed functional data Y_t , the intercept function $\delta(\tau)$, the unknown kernel functions $\phi_j(\tau)$, and the innovations ε_t . Using the B-spline expansions, we can represent the FAR model of order 1 in terms of the B-spline coefficient relations:

$$a_{t,k} = \eta_k + \epsilon_{t,k} + \sum_{i=1}^{\infty} \times \left\{ \sum_{j=1}^{\infty} \left(\frac{w_{j+m} - w_{j+1}}{w_{j+m} - w_j} - \frac{w_{j+m+1} - w_{j+2}}{w_{j+m+1} - w_{j+1}} \right) c_{j-k} \right\} \times \frac{w_{i+m} - w_i}{m} a_{t-1,i}, \quad (9)$$

for all $k = 1, \dots, \infty$. The FAR model (7) can be solved by estimating the B-spline coefficients, see Chen, Chua, and Härdle (2019).

2.2.2 | Sieve method

There are, again, an infinite number of functional basis coefficients after expansion since $k = 1, \dots, \infty$ in Equations (8) and (9). Obviously, it is computationally infeasible to estimate the coefficients in infinite dimensional parameter spaces with a finite number of sample. To enable estimation with a finite sample, the method of sieves developed by Grenander (1981) has been extensively used in FDA and has good consistency properties (Chen, 2007). In particular, the sieve method constructs a sequence of subspaces $\{\Theta_{m_n}\}$, called sieves, of the original infinite-dimensional space Θ . The sieves need to be compact and nondecreasing with $\Theta_{m_n} \subseteq \Theta_{m_{n+1}} \subseteq \dots \subseteq \Theta$, and the union of the subspaces $\cup \Theta_{m_n}$ must be dense in Θ .

Specifically, let $\{\Theta_{m_n}\}$ denote the finite-dimensional linear space of certain polynomials, say trigonometric polynomials, on $[0, 1]$ of degree m_n or less, that is

$$\Theta_{m_n} = \{K(\tau) \in L^2 \mid K(\tau) = \theta_0 1_{[0,1]} + \sum_{k=1}^{m_n} \theta_k \Phi_{2k}(\tau) + \sum_{k=1}^{m_n} \vartheta_k \Phi_{2k-1}(\tau), \\ \sum_{k=1}^{m_n} k^2 \theta_k^2 + k^2 \vartheta_k^2 \leq \zeta m_n, \theta_0, \theta_k, \vartheta_k \in \mathbb{R}, \tau \in [0, 1]\},$$

where $\{\theta_0, \theta_k, \vartheta_k\}$ are the coefficients for functional terms in the expansion. Here ζ is a positive constant that ensures the constraint is satisfied without sacrifice of the growth rate of m_n , and $m_n \rightarrow \infty$ as $n \rightarrow \infty$, that is, the number of parameters increases with the sample size. Sieve method provides a tool to represent the relationship of an infinite number of parameters in a finite subspace: the m_n th sieve space spanned by the first m_n functional basis. The hyperparameter m_n controls the smoothing degree and balances the bias and variance of the approximation projected on the sieve space. Some common approaches can be used to select the optimal m_n , such as pseudo-versions of Akaike information criterion (AIC) and Bayesian information criterion (BIC) (Yao, Müller, & Wang, 2005), leave-one-out cross validation (Rice & Silverman, 1991) and combined BIC and AIC (Kong et al., 2016). See Geman and Hwang (1982), Chen and Shen (1998) and Chen (2007) for more theoretical details and an explanation of the implementations of sieves.

Under the sieves assumption, the unknown parameters in the coefficients relationship in (8) or (9) are reduced to a subspace $k = 1, \dots, m_n$. The estimation of the FAR model in (7) is thus transformed to estimation of a finite number of unknown coefficients from the Fourier or B-spline expansions. Under the assumption that the coefficients of the innovation, $\eta_t, \epsilon_{t,k}, \omega_{t,k}$ are IID Gaussian distributed with zero mean and variance σ_k^2 , the relationship of the unknown

coefficients can be solved using MLE or least square estimator (LS) methods with closed-form solution and good asymptotic properties, see for example, Chen and Li (2017).

2.3 | FARX and pFAR for exogenous variables

Functional data or quasi functional data usually exhibit both serial dependence across time and causal dependence on exogenous variables such as economic factors and temperature. To utilize the rich information, the FAR model with exogenous variable (FARX) and partial FAR (pFAR) model extends the classic FAR model by incorporating the effect of exogenous covariates. The FARX takes into account functional type exogenous covariate, and the number of covariates is usually small. Alternatively, the pFAR model enables the consideration of large-dimensional covariates with mixed data types, including both functional covariates such as the lagged curves and scalar covariates such as economic factors. As such, pFAR model can simultaneously handle the serial dependence and causal inference in a unified framework.

2.3.1 | The FARX model

Let $\{X_t^{(\ell)}(\tau), \ell = 1, \dots, q\}$ denote the real continuous exogenous functions defined in \mathcal{H} over the same time domain as $Y_t(\tau)$. The functional covariates are assumed to be stationary, with time invariant mean functions $\mu_x^{(1)}, \dots, \mu_x^{(q)}$. Chen et al. (2018) proposed the FARX model of order p defined as:

$$Y_t(\tau) - \mu_y(\tau) = \sum_{j=1}^p \int_0^1 \phi_{y,j}(\tau - \pi) [Y_{t-j}(\pi) - \mu_y(\pi)] d\pi + \sum_{\ell=1}^q \int_0^1 \phi_{x,\ell}(\tau - \pi) [X_{t-1}^{(\ell)}(\pi) - \mu_x^{(\ell)}(\pi)] d\pi + \varepsilon_t(\tau), \quad (10)$$

where $\phi_{y,j}$ and $\phi_{x,\ell}$ are the kernel operators for $j = 1, \dots, p$, and $\ell = 1, \dots, q$ with $\phi_{(\cdot, \cdot)} \in L_2$ and $\|\phi_{(\cdot, \cdot)}\|_2 < 1$ representing the impact of the p lagged curves and q exogenous variables respectively. Without loss of generality, we adopt even functions for the kernels.

The FARX model in (10) is a generalized version of the FAR(p) model in (7) with q number of exogenous functional predictors. Following the similar expansion as in the classic FAR model derived in Section 2.2, Chen et al. (2018) obtain the relationship of expansion coefficients under the sieve Θ_{m_n} , which yields a closed form solution for the FARX model. We refer to Chen et al. (2018) for the detailed procedure for the FARX estimator and its asymptotic properties.

2.3.2 | The pFAR model

In the big data era, there are often large scale features with mixed data types. The dynamics of the functional response can be described not only with the lagged functional covariates, but there may also be some causal dependence with ultra-high dimensional exogenous scalar covariates. Let $\{z_{t\ell}, \ell = 1, \dots, d\}_{t=1}^n$ denote d exogenous scalar covariates with mean values $\mu_z^{(1)}, \dots, \mu_z^{(d)}$. The number of scalar covariates, d , is usually large. Chen, Koch, and Xu (2019) introduced the pFAR(p) model defined as:

$$Y_t(\tau) - \mu_y(\tau) = \sum_{j=1}^p \int_0^1 \phi_{y,j}(\tau - s) [Y_{t-j}(s) - \mu_y(s)] ds + \sum_{\ell=1}^d \phi_{z,\ell}(\tau) [z_{t-1,\ell} - \mu_z^{(\ell)}] + \varepsilon_t(\tau), \quad (11)$$

where the kernel operators $\phi_{y,j} \in L^2([0, 1])$ control the serial dependence with $j = 1, \dots, p$ and the kernel operators $\phi_{z,\ell} \in L^2([0, 1])$ measure the impact of the large-scale scalar covariates with $\ell = 1, \dots, d$. The pFAR model in (11) is one of the generalized versions of the FAR(p) model in (7) with a large number of scalar exogenous variables.

Direct estimation of pFAR easily causes overparametrization and overfitting due to the curse of dimensionality, making interpretation difficult and prediction inaccurate. Regularized estimation provides a remedy. Instead of

assuming all the covariates are relevant, we could select key elements and groups (i.e., a combination of elements) that are essentially driving the dynamics of the response variable. Among others, Tibshirani (1996) proposed Lasso for variable selection under sparse regularity. See also adaptive Lasso (Ivanoff, Picard, & Rivoirard, 2016), smoothly clipped absolute deviation (SCAD; Aneiros, Ferraty, & Vieu, 2015), and least angle regression selection (LARS; Efron, Hastie, Johnstone, & Tibshirani, 2004). In addition to single variable selection, grouping structures often arise where a number of variables belong to a particular predetermined group. Group Lasso (Friedman, Hastie, & Tibshirani, 2010; Xu & Ghosh, 2015) and group SCAD (Huang, Breheny, & Ma, 2012) focus on group selection, which, however, does not yield sparsity within a group. On the other hand, sparse group Lasso (Simon, Friedman, Hastie, & Tibshirani, 2013) and multivariate sparse group Lasso (Li, Nan, & Zhu, 2015) are able to detect both the active groups and the active elements within the group simultaneously. In the pFAR framework, Chen, Koch, and Xu (2019) assume that there exist certain group structures among the large number of exogenous variables and only a few of them are significant. The sparse group Lasso penalty (Friedman et al., 2010; Li et al., 2015) is applied to achieve the group and individual variable selection. The regularized estimation is conducted without knowing the number and location of the active groups and elements.

Following a similar expansion as in the FAR model in Section 2.2 under the approximating sieves Θ_{m_n} , the relationship of coefficients for model (11) is obtained. Next, Chen, Koch, and Xu (2019) adopt the two-layer sparsity assumption for both groups and elements given the high-dimensional mixed-type covariates. For this more challenging problem, there is no closed-form estimator. The regularized pFAR model is estimated via least squares with the two-layer sparsity as penalty function:

$$\min_B \left\{ \frac{1}{2n} \left\| \mathbf{Y} - \sum_{g \in G} \mathbf{X}^{(g)} \mathbf{B}_g \right\|_F^2 + \lambda \sum_{g \in G} \eta_g \|\mathbf{B}_g\|_2 + \alpha \sum_{i,j} |\mathbf{B}_{i,j}| \right\}$$

where $\|\cdot\|_F$ is the Frobenius norm. $\mathbf{Y} = (y_1^\top, \dots, y_n^\top)^\top$ and $\mathbf{X} = (x_1^\top, \dots, x_n^\top)^\top$ with $y_t = (a_{t,0}, b_{t,1}, a_{t,1}, \dots, b_{t,m_n}, a_{t,m_n})^\top$, $x_t = (y_{t-1}^\top, \dots, y_{t-p}^\top, z_{t-1,1}, \dots, z_{t-1,d})^\top$. \mathbf{B} denotes the unknown Fourier coefficients of the kernel functions $\{\phi_{y,j}, j = 1, \dots, p\}$ and $\{\phi_{z,\ell}, \ell = 1, \dots, d\}$. Assume \mathbf{B} contains G groups, and \mathbf{B}_g denotes the group of g with $g \in 1, \dots, G$. \mathcal{G} denotes the group structure set, that is $\mathbf{B}_g \in \mathcal{G}$ for group $g \in G$. $\mathbf{X}^{(g)}$ and \mathbf{B}_g refer to the submatrix of \mathbf{X} and \mathbf{B} corresponding to group g respectively. η_g is a positive group weight and a default choice of η_g is the square root of the group size; see Huang et al. (2012) and Simon et al. (2013). The tuning parameter $\lambda \geq 0$ is for groups. When $\lambda = 0$, the penalty reduces to Lasso. When λ increases, the group sparsity increases and becomes more important. The other tuning parameter $\alpha \geq 0$ controls the sparsity for the individual variables. When $\alpha = 0$, the penalty becomes group Lasso. When α increases, the element sparsity involves a larger weight and becomes more important. Various criteria have been proposed to choose the penalty parameters, for example, cross-validation (Rice & Silverman, 1991) and BIC (Wang, Li, & Tsai, 2007) and forward-looking criterion by optimizing the out-of-sample forecast accuracy (Chen, Koch, & Xu, 2019). We refer to Chen, Koch, and Xu (2019) for the detailed procedure for the pFAR estimator and its asymptotic properties.

2.4 | Multivariate models: VFAR and CFPC

In the literature the univariate FAR models is extended to multivariate framework where multiple functional time series $Y_t^{(k)}(\tau)$ are observed simultaneously, $k = 1, \dots, K$. The multiple functional time series, which can also be referred as multiple groups (each sequence of functional time series is referred as one group), exhibit both serial dependence and lead-lag cross-dependence among each other. Among others, we discuss the vector FAR (VFAR) model for moderate dimensions of functional time series and also the CFPC approach that can be used to reduce the dimensionality of ultra-high dimensional functional time series.

2.4.1 | The VFAR model

Without loss of generality, we consider two functional time series. Nevertheless, the derivation can be easily extended to higher dimensions. Denote the bivariate series of curves by $Y_t^{(1)}(\tau)$ and $Y_t^{(2)}(\tau)$ at $t = 1, \dots, n$. Chen, Chua, and Härdle (2019) proposed the vector FAR (VFAR) model of order p for the bivariate time series defined as:

$$\begin{bmatrix} Y_t^{(1)} - \mu_1 \\ Y_t^{(2)} - \mu_2 \end{bmatrix} = \sum_{k=1}^p \begin{bmatrix} \rho^{11,k} & \rho^{12,k} \\ \rho^{21,k} & \rho^{22,k} \end{bmatrix} \begin{bmatrix} Y_{t-k}^{(1)} - \mu_1 \\ Y_{t-k}^{(2)} - \mu_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \end{bmatrix} \quad (12)$$

where $\rho^{11,k}, \rho^{12,k}, \rho^{21,k}, \rho^{22,k}$ are the operators that show the serial cross-dependence among the curves on their k th lagged values. Again, the operators are bounded linear operator from \mathcal{H} to \mathcal{H} . The innovation processes $\varepsilon_t^{(1)}$ and $\varepsilon_t^{(2)}$ are strong \mathcal{H} -white noise as defined before and not necessarily cross-independent. VFAR model in (12) is one of the generalized versions of the FAR(p) model in (7) for two dimensional functional time series. As before, every ρ is written in the form of a convolution kernel Hilbert-Schmidt operator as $\phi_{11}, \phi_{12}, \phi_{21}$ and ϕ_{22} respectively with $\phi_{xy} \in L^2([0, 1])$ and $\|\phi_{xy}\|_2 < 1$ for $xy = 11, 12, 21$ and 22 . Following the similar expansion as in the classic univariate model under sieve Θ_{m_n} , for example, the B-spline expansion as in Chen, Chua, and Härdle (2019), the relationship of the B-spline coefficients are derived for the VFAR. We refer to Chen, Chua, and Härdle (2019) for the detailed procedure for the VFAR estimator and its asymptotic properties.

2.4.2 | Common functional principal component

When the dimensions of functional time series is high, factor models based on CFPC technique can be used to reduce dimensionality. Zhang et al. (2017) considered common FPC (CFPC) for multiple functional time series which requires the same eigen-structure across all groups of curves and makes the projected basis comparable among multiple groups. In specific, based on the functional eigenbasis expansion in (4), in CFPC the covariance operators Υ_g have shared orthonormal eigenfunctions ϕ_j for G groups, where $\phi_{g_1,j} = \phi_{g_2,j}$ for $1 \leq g_1, g_2 \leq G$, while the eigenvalues $\lambda_{g,j}$ are different indicating the heterogeneity among the groups. After extracting the CFPCs across all the groups of functional time series, the classic time series modeling is applied to each individual functional time series, that is, the series of the common functional principal scores. For group g the covariance function is written as:

$$\nu_g(\pi, \tau) = \sum_{j=1}^{\infty} \lambda_{g,j} \phi_j(\pi) \phi_j(\tau).$$

The eigendecomposition in the CFPC framework becomes:

$$\int_0^1 \nu_g(\pi, \tau) \phi_j(\pi) d\pi = \lambda_{g,j} \phi_j(\tau) \Leftrightarrow (\Upsilon_g \phi_j)(\tau) = \lambda_{g,j} \phi_j(\tau), \text{ subject to } \langle \phi_j, \phi_l \rangle = \delta_{jl}. \quad (13)$$

Approximating the integral of the covariance function using quadrature techniques, the eigenequation in (13) can be approximated as:

$$\begin{aligned} \int_0^1 \nu_g(\pi, \tau) \phi_j(\pi) d\pi &\approx \sum_{\ell=1}^L w_\ell \nu_g(\pi_\ell, \tau) \phi_j(\pi_\ell) \\ &= \bar{\nu}_g(\tau)^T w^* \bar{\phi}_j, \end{aligned}$$

where $\bar{\nu}_g(\tau) = (\nu_g(\pi_1, \tau), \dots, \nu_g(\pi_L, \tau))^T$, $w = (w_1, \dots, w_L)^T$ are quadrature weights, $\bar{\phi}_j = (\phi_j(\pi_1), \dots, \phi_j(\pi_L))^T$ and $*$ is a Hadamard product and L is the number of discrete arguments of quadrature expansion. As the covariance function is symmetric, it yields to:

$$\Upsilon \phi_j = V_g W \bar{\phi}_j,$$

where W is a diagonal $(L \times L)$ matrix with w_l in the elements and $V_g = (\nu_g(\pi_\ell, \pi_k))_{\ell, k}$ is a $(L \times L)$ matrix with the covariance function values at the quadrature points, $\bar{\phi}_j = (\phi_j(\pi_1), \dots, \phi_j(\pi_L))^T$. Therefore, the eigenequation in (13) is rewritten as

$$V_g W \bar{\phi}_j = \lambda_{g,j} \bar{\phi}_j, \text{ subject to the condition of orthonormality } \bar{\phi}_j^T W \bar{\phi}_\ell = \delta_{j\ell}.$$

Assuming the weights to be positive, the approximated eigenequation is obtained in the following form,

$$W^{1/2}V_g W^{1/2}u_j = \lambda_{g,j}u_j, \text{ subject to } \langle u_j, u_\ell \rangle = \delta_{j\ell}, \quad (14)$$

where $u_j = W^{1/2}\bar{\phi}_j$.

After selecting p factors corresponding to the largest component scores, denoted as $\xi_{t,j}^{(g)}$ for $j = 1, \dots, p$ defined in (5), we can adopt the classic time series modeling to estimate and forecast the component scores, for example, AR model and VAR model. We refer to Zhang et al. (2017) for the detailed procedure for the CFPC estimator and its application.

2.5 | Under nonstationarity: WFAR, VC-FAR And AFAR

The above models assume time homogeneity where the dynamics of time series is stable. It means that the structure of the serial- and cross-dependence as well as the association to exogenous variables are always constant. Although stationarity probably holds in some cases, it is more realistic to ask, “what if there exists a structural change due to seasonality?” Or even more challenging, “what if there is an unforeseeable regime shift on either mean or cross dependence?” Seasonal variations and unstable turbulence take place in practice, especially in economic and finance spheres. In this section, we present the warping FAR (WFAR) model, the varying-coefficient FAR (VC-FAR) and adaptive FAR (AFAR) for nonstationary functional time series. The WFAR model is used to separate seasonal variation from the time series. The VC-FAR model is designed to handle nonstationarity with smooth changes whereby the operators (parameters) vary smoothly over time and can be locally approximated by a function of time. The AFAR model is more flexible and can model nonstationary time series with both smooth or abrupt structural changes of the underlying process.

2.5.1 | Warping FAR

In practice, functional time series may exhibit seasonal variations and time evolution which are driven by different dynamic movements. Seasonal variations are defined as the existence of phase swings in some curves at every period of seasons. For example, in the energy sector, phases of hourly electricity prices within a day may be different on weekends compared to weekdays. A separation of the amplitude and phase variations is of great interest in order to discover a broad common diurnal pattern of the curves without the impact of seasons. In addition, it helps understanding of the seasonal changes in curves and will enhance the forecast accuracy. Chen, Marron, and Zhang (2019) proposed the Warping FAR (WFAR) model to simultaneously account for phase and amplitude variations of functional time series with seasonality, where the seasonal phase variations in functional curves are separated from the amplitude changes using a warping (seasonal adjustment) methods, afterwards the seasonally-adjusted curves are used in FAR framework.

We denote the warping functions $\gamma^{(s)}$, where type s refers to the group of curves repeating in each season, for example, week days in weekly seasonality. Chen, Marron, and Zhang (2019) proposed the WFAR model of order p defined as follows:

$$X_t^{(s)}(\tau) = Y_t^{(s)}(\tau) \circ \gamma_t^{(s)} \quad (15)$$

$$X_t(\tau) - \mu_x(\tau) = \sum_{j=1}^p \rho_j (X_{t-j}(\tau) - \mu_x(\tau)) + \varepsilon_t(\tau), \quad (16)$$

where $X_t^{(s)}(\tau)$ denotes the seasonally-adjusted curve obtained in a warping process $Y_t^{(s)}(\tau) \circ \gamma_t^{(s)}$. $\mu_x(\tau)$ is the mean function of $X_t(\tau)$. The warping function γ_t is estimated following the nonparametric approach proposed by Srivastava, Wu, Kurtek, Klassen, and Marron (2011), where the amplitude evolution of the curves remains unchanged in the seasonal adjustment. The serial dependence of the warped curves in (16) is then modeled with the classic FAR model with kernel operator $\rho_j \in L^2([0, 1])$.

The forecasts of the curves are obtained by warping back the deseasonalized curve forecasts with the respective warping functions of the same type. For an h -step-ahead forecast, we have

$$\hat{Y}_{t+h}^{(s)}(\tau) = \hat{X}_{t+h}^{(s)}(\tau) \circ [\gamma^{(s)}]^{-1}(\tau), \quad (17)$$

where $[\gamma^{(s)}]^{-1}$ is the inverse of $\gamma^{(s)}$. We refer to Chen, Marron, and Zhang (2019) for the detailed estimation procedure of the warping function and numerical analysis of the WFAR model.

2.5.2 | Varying-coefficient FAR

There are nonstationarity cases with changing time variations. The varying coefficient FAR (VC-FAR) and the adaptive FAR (AFAR), are defined with time dependent Hilbert–Schmidt operator (kernel function) as:

$$Y_t(\tau) - \mu(\tau) = \sum_{j=1}^p \int_0^1 \phi_{t,j}(\tau - \pi) [Y_{t-j}(\tau) - \mu(\pi)] d\pi + \varepsilon_t(\tau), \quad (18)$$

where $\phi_{t,j}(\tau) \in L^2([0, 1])$ is the kernel function for a certain time point t with $\|\phi_{t,j}\|_2 < 1$ and $\|\cdot\|_2$ denotes the standard L_2 norm. Both VC-FAR and AFAR are designed for nonstationary functional time series. However, the VC-FAR relies on smoothing transition depending on time t and its estimation is performed with time varying kernels. The Adaptive FAR, on the other hand, does not make any assumption on the type of structural change and its estimation is conducted based on sequential testing procedure under local homogeneity assumption.

In the VC-FAR framework proposed by Xu et al. (2017), the kernel $\phi_{t,j}(\tau)$ is time dependent and is required to be a smoothing function over time. Therefore, after functional expansion in (8), the coefficients of the functional basis are time dependent. Here, we give an illustration when the functional terms are expanded with Fourier basis, when $p = 1$.

After expanding the curves with the Fourier basis under the sieve assumptions, the unknown Fourier coefficients assumed to be continuous with time t , that is, $\theta(t) = (\theta_0(t)^\top, \theta_1(t)^\top, \dots, \theta_{m_n}(t)^\top)^\top$ for $k = 1, \dots, m_n$, where $\theta_0(t) = (\omega_0(t), c_0(t))^\top$ and $\theta_k(t) = (\omega_k(t), \eta_k(t), c_k(t), d_k(t))^\top$. A first order Taylor expansion of the time dependent Fourier coefficients of the kernel and intercept function yields

$$\begin{aligned} c_0(t) &\approx c_0(s) + c_0^{(1)}(s)(t-s), c_k(t) \approx c_k(s) + c_k^{(1)}(s)(t-s), d_k(t) \approx d_k(s) + d_k^{(1)}(s)(t-s), \\ \omega_0(t) &\approx \omega_0(s) + \omega_0^{(1)}(s)(t-s), \omega_k(t) \approx \omega_k(s) + \omega_k^{(1)}(s)(t-s), \eta_k(t) \approx \eta_k(s) + \eta_k^{(1)}(s)(t-s), \end{aligned}$$

for s in a local neighborhood of t . We denote the k th order derivatives of the coefficient functions $\theta(s)$ as $\theta^{(k)}(s)$. The estimation of the unknown parameters can be estimated by minimizing the following local least square:

$$\left(\hat{\theta}(s), \hat{\theta}^{(1)}(s) \right) = \operatorname{argmax}_{\theta, \theta^{(1)}} \sum_{t=1}^n l(\theta(s), t, Y) J_h(t-s), \quad (19)$$

where Y is the set of all the transformed data series $\{a_{t,0}, a_{t,k}, b_{t,k}\}_{t=1}^n$ for $k = 1, 2, \dots, m_n$, and

$$\begin{aligned} l(\theta(s), t, Y) &= \left(a_{t,0} - \left(\omega_0(s) + \omega_0^{(1)}(s)(t-s) \right) - \left(c_{t,0} + c_0^{(1)}(s)(t-s) \right) a_{t-1,0} \right)^2 \\ &+ \sum_{k=1}^{m_n} \left[\left(a_{t,k} - \left(\omega_k(s) + \omega_k^{(1)}(s)(t-s) \right) - \frac{1}{\sqrt{2}} \left(c_k(s) + c_k^{(1)}(s)(t-s) \right) a_{t-1,k} \right. \right. \\ &+ \left. \left. \frac{1}{\sqrt{2}} \left(d_k(s) + d_k^{(1)}(s)(t-s) \right) b_{t-1,k} \right)^2 + \left(b_{t,k} - \left(\eta_{t,k} + \eta_k^{(1)}(s)(t-s) \right) \right. \right. \\ &\left. \left. - \frac{1}{\sqrt{2}} \left(c_{t,k} + c_k^{(1)}(s)(t-s) \right) b_{t-1,k} + \frac{1}{\sqrt{2}} \left(d_{t,k} + d_k^{(1)}(s)(t-s) \right) a_{t-1,k} \right)^2 \right]. \end{aligned}$$

$J_h(t-s)$ denotes the rescaled kernel function $J(\frac{t-s}{h})/h$, and $J(\cdot)$ is set to be the Gaussian kernel function in the numerical analysis of this review study.

To facilitate notations, Xu et al. (2017) define $A_0 = (a_{1,0}, a_{2,0}, \dots, a_{n,0})^\top$, $A_k = (a_{1,k}, a_{2,k}, \dots, a_{n,k})^\top$ and $B_k = (b_{1,k}, b_{2,k}, \dots, b_{n,k})^\top$,

$$Y_0 = \begin{pmatrix} 1 & a_{0,0} & 1-s & a_{0,0}(1-s) \\ 1 & a_{1,0} & 2-s & a_{1,0}(2-s) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & a_{n-1,0} & n-s & a_{n-1,0}(n-s) \end{pmatrix},$$

$$Y_k^1 = \begin{pmatrix} 1 & 0 & a_{0,k} & -b_{0,k} & 1-s & 0 & a_{0,k}(1-s) & -b_{0,k}(1-s) \\ 1 & 0 & a_{1,k} & -b_{1,k} & 2-s & 0 & a_{1,k}(2-s) & -b_{1,k}(2-s) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & a_{n-1,k} & -b_{n-1,k} & n-s & 0 & a_{n-1,k}(n-s) & -b_{n-1,k}(n-s) \end{pmatrix},$$

$$Y_k^2 = \begin{pmatrix} 0 & 1 & b_{0,k} & a_{0,k} & 0 & 1-s & b_{0,k}(1-s) & a_{0,k}(1-s) \\ 0 & 1 & b_{1,k} & a_{1,k} & 0 & 2-s & b_{1,k}(2-s) & a_{1,k}(2-s) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & b_{n-1,k} & a_{n-1,k} & 0 & n-s & b_{n-1,k}(n-s) & a_{n-1,k}(n-s) \end{pmatrix},$$

$$Y_k = (Y_k^1 \ Y_k^2)^\top,$$

$$W = \text{diag}\{J_h(1-s), J_h(2-s), \dots, J_h(n-s)\}, W_2 = \text{diag}\{W, W\}.$$

Furthermore, elements of $\theta(s)$ and $\theta^{(1)}(s)$ are partitioned and grouped into two components, and denoted as

$$\beta_0(s) = (\omega_0(s), c_0(s), \omega_0^{(1)}(s), c_0^{(1)}(s))^\top,$$

$$\beta_k(s) = (\omega_k(s), \eta_k(s), c_k(s)/\sqrt{2}, d_k(s)/\sqrt{2}, \omega_k^{(1)}(s), \eta_k^{(1)}(s), c_k^{(1)}(s)/\sqrt{2}, d_k^{(1)}(s)/\sqrt{2})^\top.$$

We obtain the following estimators which minimize (19),

$$\hat{\beta}_0(s) = (Y_0^\top W Y_0)^{-1} Y_0^\top W A_0, \hat{\beta}_k(s) = (Y_k^\top W_2 Y_k)^{-1} Y_k^\top W_2 \begin{pmatrix} A_k \\ B_k \end{pmatrix}.$$

Eventually we find the local linear estimators for kernel functions of Hilbert–Schmidt operator at time s as follows,

$$(\hat{\omega}_0(s), \hat{c}_0(s))^\top = l_1 \hat{\beta}_0(s), (\hat{\omega}_k(s), \hat{\eta}_k(s))^\top = l_2^1 \hat{\beta}_k(s), (\hat{c}_k(s), \hat{d}_k(s))^\top = \sqrt{2} \times l_2^2 \hat{\beta}_k(s),$$

with the selection matrices defined by $l_1 = (I_2, 0_{2 \times 2})$, $l_2^1 = (I_2, 0_{2 \times 6})$, and $l_2^2 = (0_{2 \times 2}, I_2, 0_{2 \times 4})$ and I_2 denotes the identity matrix with a degree of 2. We refer to Xu et al. (2017) for the detailed procedure for the VC-FAR estimation and the theoretical properties of its estimator.

2.5.3 | Adaptive FAR

In a more general setup where both the location and the form of structural changes are unknown, AFAR proposed by Chen and Li (2017) is an appropriate choice. The AFAR framework is the same as defined in model (18), but the kernel operator $\phi_t(\tau)$ is time dependent and does not need to qualify as a smoothing function.

Under time-homogeneity with the constant parameters, the AFAR model is obviously reduced to the classic FAR model in (7). In a more realistic situation with $p = 1$, where the unknown coefficient denoted by $\theta_t = (c_{t0}, c_{tk}, \omega_{t0}, \omega_{tk}, \eta_{tk}, \sigma_{t0}, \sigma_{tk})$ are time dependent, the estimation of AFAR is conducted under the assumption of local homogeneity. That is, at a fix time point t there exists a local interval $I_t = [t - \ell_t, t]$ with $1 < \ell_t < t$ over which all the included observations can be well described by a local FAR model with approximately constant parameters, that is, $\theta_t \approx \text{constant}$. Simultaneously we require that the modeling bias under this local parametric assumption is small, that is, the small modeling bias condition (Belomestny, Spokoiny, et al., 2007). Over interval I_t , the abovementioned FAR model under stationarity can be safely used for parameter estimation. The estimated parameter over I_t denoted as $\hat{\theta}_t$ is called adaptive estimator.

In reality, the local homogeneity interval I_t is unknown and the number of possible candidates is large, for example, as many subsamples as there are past sample periods, and it also can be computationally expensive to search among all possible interval candidates in the samples. Belomestny et al. (2007) shows that an optimal choice of the interval of local homogeneity can be obtained via an adaptive procedure based on likelihood ratio testing. Likewise, Chen and Li (2017) adopted a sequential testing to detect such interval from S candidate intervals $\mathbf{I}_t = \{I_t^{(1)}, \dots, I_t^{(S)}\}$ with $I_t^{(1)} \subset \dots \subset I_t^{(S)}$ to alleviate the computational burden in practice. To each interval there exists a corresponding local estimator obtained using method at Section 2.2, denoted by $\tilde{\theta}_t^{(s)}$, which is called weak estimator.

A sequential testing procedure is adopted to select the longest one from the S candidates that does not contain any breaks and obtain $\hat{\theta}_t$. The process of parameter estimation starts from the smallest interval $I_t^{(1)}$, where the local homogeneity is assumed to be true. The corresponding adaptive estimator is referred as $\hat{\theta}_t^{(1)} = \tilde{\theta}_t^{(1)}$. With the next iteration, a longer interval $I_t^{(s)}$ for $s = 2, \dots, S$ is selected, and checked using a test statistic as follows:

$$T_t^{(s)} = \left| L(I_t^{(s)}; \tilde{\theta}_t^{(s)}) - L(I_t^{(s)}; \hat{\theta}_t^{(s-1)}) \right|^{1/2},$$

where $L(I_t^{(s)}; \tilde{\theta}_t^{(s)})$ and $L(I_t^{(s)}; \hat{\theta}_t^{(s-1)})$ are the local log-likelihood over interval $I_t^{(s)}$ using the weak estimator to be tested for local homogeneity in the current step and using the adaptive estimator accepted in the previous step, respectively. The statistic $T_t^{(s)}$ measures the divergence of the hypothetical AFAR model and the time varying model accepted from the previous step. If the test statistics is significantly large, it means the local interval $I_t^{(s)}$ does not satisfy the local homogeneity condition and the hypothetical FAR model considerably diverges from the time varying model with the accepted parameter in the previous step. It leads us to reject the null hypothesis and stop the algorithm process. If the null hypothesis is not rejected, we accept the local homogeneity in the local interval and accept the adaptive estimator to be $\hat{\theta}_t^{(s)} = \tilde{\theta}_t^{(s)}$. The iteration stops when it detects a change or reaches the longest candidate interval. The final selected interval will be the last accepted interval before the stop. After the identification of the local interval of homogeneity, the FAR model and its variation under stationarity can be used to describe the local dynamics. We refer to Chen and Li (2017) for the detailed algorithm of the sequential testing procedure and theoretical properties of the AFAR model.

3 | DATA: NATURAL GASFLOWS

Natural gas, as a key energy resource for Germany and Europe, is distributed over the country with a high-pressure gas pipeline network of more than 11,000 km in length. Gas flows at the more than 1,000 entry- and exit-points of the network are recorded every hour and are available as daily gas curves. The future demand and supply of the natural gas flows are potentially influenced by many factors, for example, the cycles of working routines, the temperatures in different locations of the gas transmission network, the market prices, and the supply of renewable energy resources. A number of gas markets are shifting towards more short-term planning, namely, day-ahead contracts, which can be challenging for the operators in the dispatching centre (Chen et al., 2018). In order to increase the efficiency of the network operation, it is important to predict future natural gas consumption with high accuracy.

We consider the natural gas flows at two municipal energy supplier-nodes in the German high-pressure natural gas transmission network. The data cover a 2-year period from 1 October Y1 to 30 September Y3, comprising 730 days,

where $Y1$ and $Y3$ denotes the years of the dataset. The years are hidden for confidentiality reasons. The gas flow is recorded as supply and demand with an hourly time resolution over 24 hr, 7 days a week. The two nodes (locations), labeled as $M1$ and $M2$ respectively, are demand nodes, that is, they are gas outflows from the network and serve residential and small commercial constituents. Both nodes are located near the city of Frankfurt, Germany. To provide an interpretable comparison between the two nodes, the gas flows are normalized such that each node has zero mean and unit variance. The hourly gas flow data are converted to daily flow curves by smoothing over the 24-hourly observations on each day using the Fourier expansion, motivated by the periodic characteristics of the energy data. Figure 1 displays the flow curves for each node. A strong annual seasonality can be observed in the two nodes, driven by the working routines of households. Despite differing trends and seasonality, Figure 2 displays the sample autocorrelations of the hourly gas flow at 9 a.m. and sample cross correlation surfaces of the daily gas flow curves at nodes $M1$ and $M2$. As shown, there exist significant serial dependence and positive correlation across all hours with a gradual decreasing trend as lag increases for both nodes.

Besides the gas flow data, we also have a large number of exogenous variables, namely 85 environmental and economic variables. These include four daily market price values in different networks, that is, NCG, GASPOOL, TTF, and Zeebrugge, three hourly renewable energy variables, that is, solar, wind-onshore and wind-offshore; and hourly temperatures in 78 locations, which are further split into four geographic zones and each of which is separately considered as a group. Because the variables vary in scale - the maximum value of the variable TTF, for example, is 2.90 and the maximum value of the variable wind-on-shore is 30, 158—the exogenous scalars are also normalized. Figure 3 presents the normalized observations of the 4 price variables, 3 renewable energy variables, and the average temperature variables within the 4 geographical zones covering 78 locations respectively.

4 | FORECASTING RESULTS

In this section, we demonstrate the application results of the FAR models in energy forecasting of natural gas flows described in Section 3. We demonstrate how to use the different FAR models to obtain forecasts based on the fitted dynamics of the high resolution natural gas flows. Although our implementation focuses on energy functional time series, the models and methods can be easily adapted to analyze other functional data in diverse areas.

4.1 | Model setup

We consider the classic FAR, pFAR, FPC, CFPC, WFAR, and AFAR models in the real data implementation to demonstrate the modeling of serial dependence, high-dimensional exogenous covariates, multivariate dimensional curves and

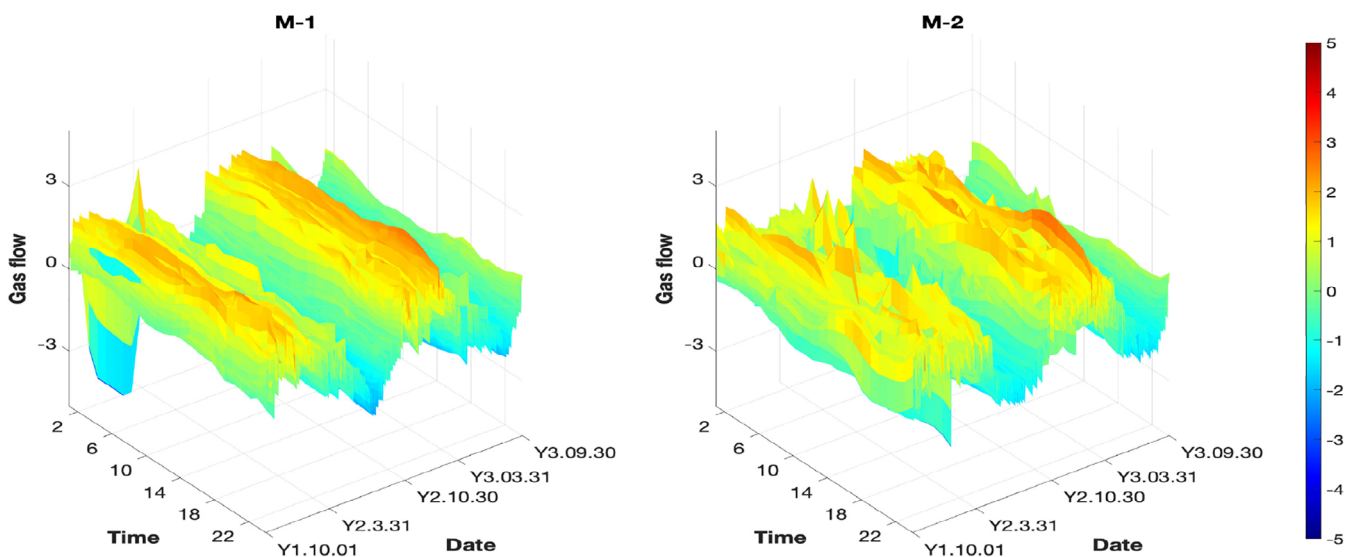


FIGURE 1 Daily gas flow curves at two representative nodes from 1 October $Y1$ to 30 September $Y3$

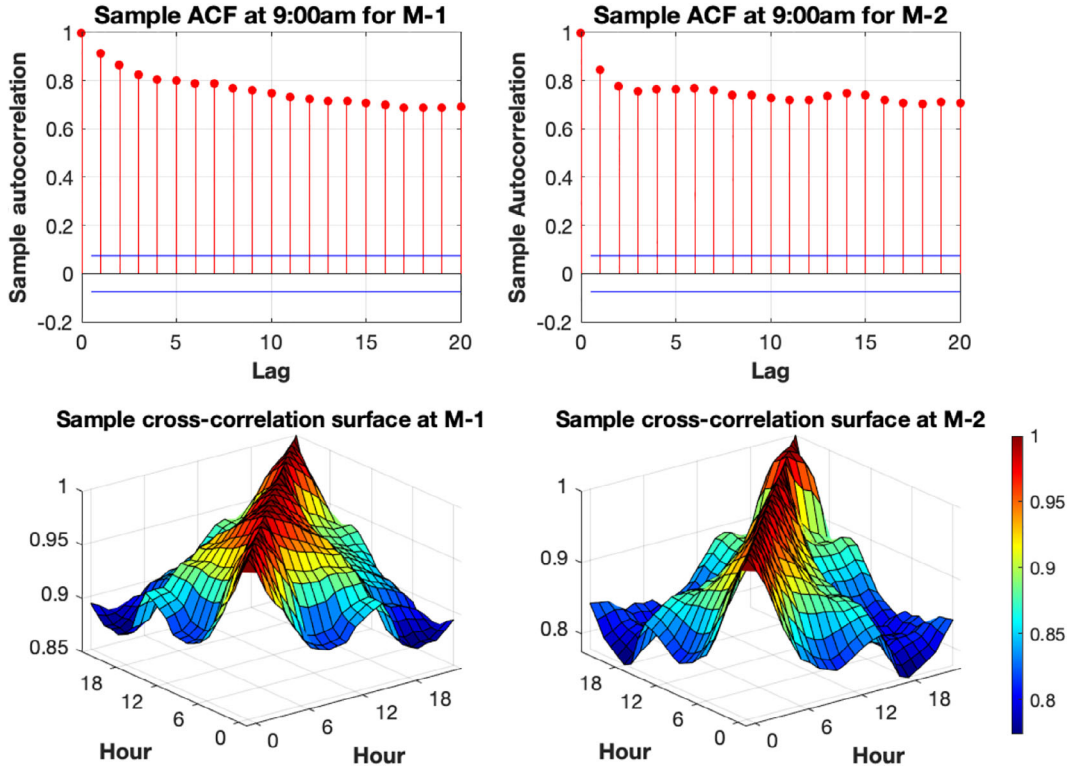


FIGURE 2 Sample autocorrelations of the gas flows at 9 a.m. and sample cross-correlations surfaces of gas flows among 24 hr at M1 and M2 from 1 October Y1 to 30 September Y3

nonstationarity in functional time series analysis. We use the first 548 daily natural gas curves from 1 October Y1 to 31 March Y3 to initialise the out-of-sample prediction. The forecast exercise starts from the 549th curve dated on 1 April Y3, and ends at the last curve on 30 September Y3 in a total of 182 forecast days. For each forecast point, we always use the period consisting of all historical data for the FAR, pFAR, FPC, and CFPC models, while applying the seasonal adjustment technique for the WFAR model and adaptive technique to identify the interval of local homogeneity for the AFAR model. The parameters are estimated over the adopted periods at each time point and the fitted model is then used to compute 1- and 14-step-ahead forecast of the natural gas curves which refers to short and long term forecast respectively. For the dimensionality of sieve m_n , we set $m_n = 1$ for pFAR model as in Chen, Koch, and Xu (2019) and use $m_n = 23$ for the FAR and AFAR models as in Chen and Li (2017). We choose seven factors for the FPC and CFPC techniques. For the pFAR model with exogenous variables, we use all the 85 exogenous covariates where sparse group Lasso and Lasso are adopted respectively for variable selection.

We measure forecast accuracy in terms of level and the out-of-sample goodness-of-fit. To facilitate comparison, we convert back to the original gas flow data when computing the forecast accuracy. Specifically, we compute the hourly mean absolute percentage error (MAPE) and the out-of-sample R^2 for h -step-ahead forecast as

$$\text{MAPE}_s = \frac{1}{|T-t_0|} \sum_{t=t_0}^T \left| \frac{Y_{t+h}(\tau_s) - \hat{Y}_{t+h}(\tau_s)}{Y_{t+h}(\tau_s)} \right|, s=1, \dots, 24,$$

$$R^2 = 1 - \frac{\sum_{s=1}^{24} \sum_{t=t_0}^T (Y_{t+h}(\tau_s) - \hat{Y}_{t+h}(\tau_s))^2}{\sum_{s=1}^{24} \sum_{t=t_0}^T (Y_{t+h}(\tau_s) - \bar{Y}_{t+h}(\tau_s))^2},$$

where interval $[t_0, T]$ indicates the forecast period. $\hat{Y}_{t+h}(\tau_s)$ is the forecast of $Y_{t+h}(\tau_s)$ at time $t+h$ and hour s . $\bar{Y}_{t+h}(\tau_s)$ is the historical average of the hourly flows up to time t .

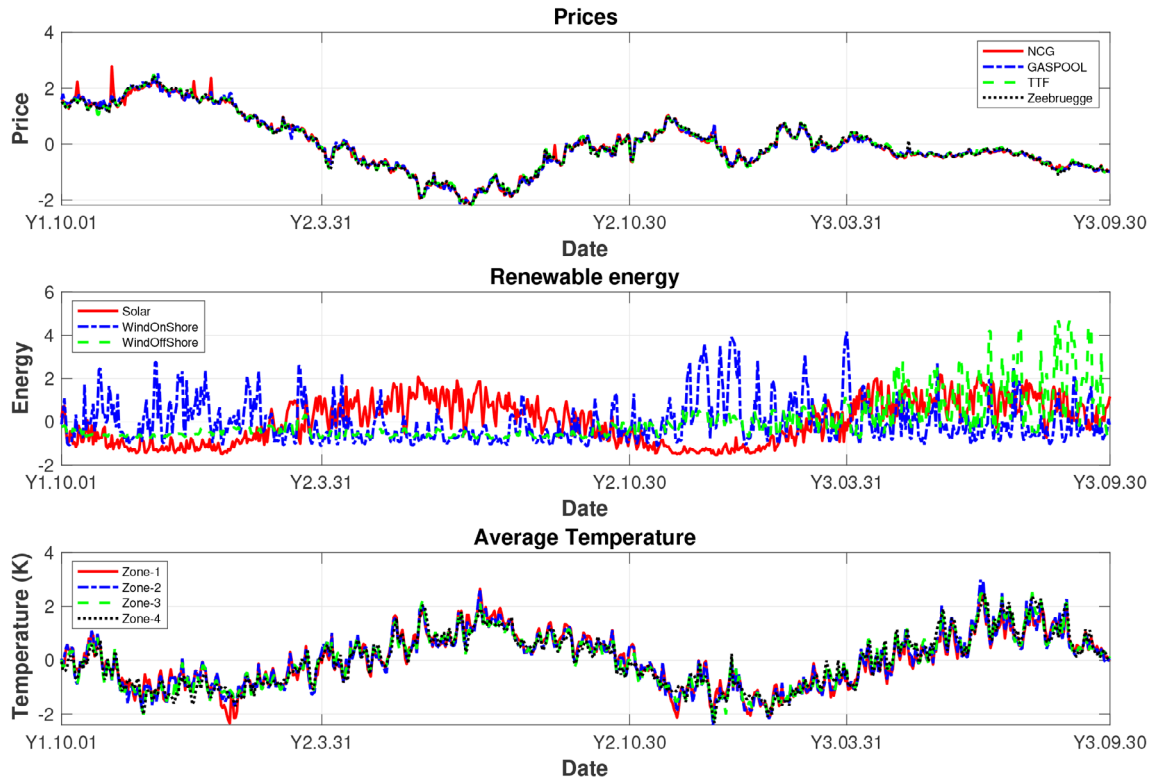


FIGURE 3 Exogenous scalar variables: four prices, that is, NCG, GASPOOL, TTF, and Zeebrugge; three renewable energy sources, that is, solar, wind-onshore and wind-offshore, and average temperature of four zones from 1 October Y1 to 30 September Y3, where Y1 and Y3 denote the years

4.2 | Result

Tables 2 and 3 report the MAPE and out-of-sample R^2 of 1-day and 14-days-ahead forecast at the two nodes for all FAR type models. The best performing model for each node is underlined and the second best is marked in bold.

The results show that pFAR(7) performs best and the improvement becomes more significant when the forecast horizon increases from 1-day to 2 weeks (14 days). For example, for 1-day-ahead forecast, pFAR(7) gives the best forecast results 13 out of 24 times compared to other models at node M1 and 15 out of 24 at node M2. However, the difference in forecast accuracy between the best and worst performing models is not that much with MAPE varying from 8.75% to 9.51%, and out-of-sample R^2 ranges from 91.50% to 88.86%, which indicates that all models have very good performance. While for a longer term forecast, pFAR(7) improves the performance of MAPE from the worst 21.30% (FPC) to 16.87% at M1, and from 34.47% (CFPC) to 24.70% at M2. This indicates that some environmental and economic variables truly affect the demand and supply of natural gas flows, and incorporating the essential lagged gas flow curves and exogenous factors could help improve forecasting performance. In the pFAR model, selected via sparse group Lasso, the active factors are found to be economically essential. For example, we find that the price effects are weak in both nodes, but solar energy has a more significantly negative effect on M1, while the wind variable has a greater positive effect on M2. Gas flows at both nodes are independent to the zone-2 temperature.

FAR(7) is the second best compared to the rest of models for 1-day-ahead forecast. For example, at node M1, FAR(7) beats other models 4 times out of 24, and comes second best with 9 out of 24. It also has the second best out-of-sample R^2 at node M1. These values become 9 out of 24 and 15 out of 24 for node M2. It is worth noting that the prediction accuracy of the FAR with lag order 7 is better than FAR with lag order 1. This indicates that incorporating higher order lags of natural gas flow curves improves the forecast results at both nodes, which is consistent to the findings in sample autocorrelation analysis in Section 3.

However, when h moves to 14, the AFAR model is the second best for most of the 24 hr forecast, and shows promising prediction performance in forecasting gas flow at both nodes. Although for $h = 1$ the forecast results of the AFAR model is only slightly better than the constant FAR(1), it shows more significant improvement when $h = 14$. Given that

TABLE 2 1-Day-ahead forecast: mean absolute percentage error (MAPE) and out-of-sample R^2 of M1 and M2 gas nodes evaluated at 24 hr

Node	M1							M2						
	FAR(1)	FAR(7)	pFAR(7)	FPC	CFPC	WFAR	AFAR(1)	FAR(1)	FAR(7)	pFAR(7)	FPC	CFPC	WFAR	AFAR(1)
1	7.77%	7.12%	7.16%	8.08%	7.97%	6.99%	7.14%	15.96%	15.36%	14.55%	16.79%	16.61%	15.23%	15.36%
2	8.19%	<u>7.62%</u>	7.69%	7.66%	7.96%	8.02%	7.63%	18.29%	16.55%	<u>15.68%</u>	18.62%	20.03%	18.30%	17.53%
3	7.65%	7.30%	7.09%	7.71%	7.88%	7.93%	7.75%	19.17%	17.83%	16.84%	19.10%	20.60%	19.27%	18.68%
4	7.04%	6.80%	<u>6.58%</u>	7.24%	7.34%	7.39%	7.59%	19.46%	18.22%	<u>17.27%</u>	19.38%	20.51%	19.61%	19.26%
5	7.21%	6.97%	6.96%	7.37%	7.50%	7.46%	7.54%	18.50%	16.51%	15.60%	18.35%	19.17%	18.64%	17.43%
6	10.23%	9.81%	<u>9.65%</u>	10.38%	10.51%	10.31%	10.53%	17.99%	15.39%	<u>14.23%</u>	17.98%	18.67%	18.10%	16.42%
7	7.89%	7.51%	7.15%	7.91%	8.08%	7.73%	7.68%	18.83%	16.21%	15.04%	18.80%	19.28%	18.78%	16.98%
8	8.57%	8.13%	<u>7.86%</u>	8.52%	8.69%	8.23%	8.12%	18.59%	15.95%	<u>15.60%</u>	18.32%	18.54%	18.53%	17.47%
9	9.35%	8.90%	<u>8.50%</u>	9.42%	9.49%	9.10%	8.75%	19.20%	16.77%	<u>16.63%</u>	19.07%	19.13%	19.30%	18.33%
10	9.92%	9.25%	9.12%	10.05%	10.08%	9.72%	<u>9.06%</u>	20.24%	<u>17.89%</u>	18.16%	19.94%	19.95%	20.30%	18.94%
11	10.52%	9.76%	<u>9.61%</u>	10.68%	10.74%	10.43%	9.63%	21.88%	<u>18.96%</u>	19.90%	20.89%	21.11%	21.93%	19.91%
12	10.85%	10.16%	10.11%	10.96%	11.03%	10.80%	<u>9.98%</u>	21.57%	<u>18.71%</u>	19.35%	23.54%	23.95%	21.44%	18.92%
13	11.03%	10.39%	<u>10.10%</u>	11.11%	11.13%	10.88%	10.27%	22.10%	19.41%	<u>19.02%</u>	21.39%	21.51%	21.82%	19.67%
14	11.22%	10.67%	<u>10.32%</u>	11.28%	11.26%	11.09%	10.52%	22.15%	18.94%	<u>18.76%</u>	21.56%	21.57%	21.75%	19.13%
15	10.81%	10.33%	10.19%	11.04%	10.85%	10.63%	<u>10.12%</u>	21.53%	<u>18.69%</u>	18.83%	20.95%	20.65%	21.08%	19.17%
16	10.43%	9.97%	9.96%	10.75%	10.55%	10.32%	<u>9.93%</u>	19.53%	<u>17.19%</u>	17.62%	18.97%	18.73%	19.14%	17.97%
17	9.07%	<u>8.77%</u>	8.86%	9.76%	9.50%	9.16%	9.20%	17.39%	<u>15.76%</u>	16.04%	17.50%	17.06%	17.22%	17.29%
18	8.62%	8.40%	8.48%	9.78%	9.16%	8.72%	8.89%	15.39%	<u>14.36%</u>	14.72%	16.26%	15.71%	15.52%	17.10%
19	8.97%	<u>8.71%</u>	8.93%	9.98%	9.41%	8.95%	9.13%	13.90%	<u>13.46%</u>	13.87%	15.06%	14.71%	14.32%	16.16%
20	9.63%	9.34%	<u>9.33%</u>	10.14%	9.85%	9.48%	9.53%	15.40%	<u>14.66%</u>	14.74%	16.42%	16.10%	15.69%	16.87%
21	9.19%	8.98%	9.07%	9.41%	9.25%	<u>8.93%</u>	9.02%	16.45%	15.90%	<u>15.82%</u>	17.25%	17.07%	16.39%	18.04%
22	9.21%	9.13%	8.92%	9.17%	9.02%	8.78%	8.78%	16.73%	16.11%	<u>15.81%</u>	17.40%	17.21%	16.24%	17.97%
23	9.54%	9.31%	<u>9.11%</u>	9.53%	9.33%	9.12%	9.14%	17.28%	16.14%	<u>15.30%</u>	17.47%	17.21%	16.76%	18.10%
24	9.94%	9.73%	<u>9.30%</u>	10.36%	9.86%	9.58%	9.90%	18.70%	17.00%	<u>16.26%</u>	18.47%	18.11%	18.38%	18.90%
Ave.	9.29%	8.88%	<u>8.75%</u>	9.51%	9.44%	9.16%	9.07%	18.59%	16.75%	<u>16.49%</u>	18.63%	18.78%	18.49%	17.98%
R^2	90.64%	91.10%	<u>91.50%</u>	90.41%	90.46%	88.86%	90.60%	86.39%	88.22%	<u>88.71%</u>	86.48%	86.32%	84.18%	86.14%

Note: The best performance for each hour is underlined and the second best is marked with boldface for each node.

Abbreviations: AFAR, adaptive FAR model; CFPC, common functional principal component; FAR, functional autoregressive; FPC, functional principle component; pFAR, partial FAR; WFAR, warping FAR.

TABLE 3 14-Days-ahead forecast: mean absolute percentage error (MAPE) and out-of-sample R^2 of M1 and M2 gas nodes evaluated at 24 hr

Node	M1							M2						
	FAR(1)	FAR(7)	pFAR(7)	FPC	CFPC	WFAR	AFAR(1)	FAR(1)	FAR(7)	pFAR(7)	FPC	CFPC	WFAR	AFAR(1)
1	21.30%	20.74%	<u>16.87%</u>	21.84%	21.79%	20.97%	19.81%	30.91%	28.70%	<u>25.40%</u>	31.68%	31.71%	30.29%	27.27%
2	22.35%	21.73%	<u>17.66%</u>	22.43%	22.66%	22.12%	21.09%	35.69%	33.11%	<u>28.67%</u>	36.18%	37.17%	35.31%	31.54%
3	22.16%	21.64%	<u>17.95%</u>	22.18%	22.47%	21.97%	21.26%	36.49%	34.27%	<u>28.67%</u>	36.89%	38.02%	36.26%	32.45%
4	21.55%	21.08%	<u>17.28%</u>	21.53%	21.79%	21.38%	20.64%	36.59%	34.50%	<u>28.08%</u>	36.75%	37.61%	36.37%	33.39%
5	20.93%	20.44%	<u>16.88%</u>	20.94%	21.13%	20.85%	19.99%	34.57%	32.27%	<u>25.82%</u>	34.72%	35.29%	34.38%	29.82%
6	22.74%	22.09%	<u>18.46%</u>	22.79%	22.89%	22.73%	21.70%	32.49%	30.35%	<u>23.85%</u>	32.76%	33.16%	32.39%	28.95%
7	19.66%	19.06%	<u>16.17%</u>	19.95%	19.98%	19.79%	19.23%	31.91%	29.41%	<u>23.81%</u>	32.33%	32.63%	31.80%	27.54%
8	19.68%	19.37%	<u>16.86%</u>	20.03%	20.05%	19.79%	19.57%	29.90%	27.92%	<u>23.78%</u>	30.25%	30.29%	29.91%	25.88%
9	20.75%	20.34%	<u>17.29%</u>	21.10%	21.13%	20.80%	19.70%	29.11%	27.68%	<u>24.25%</u>	29.43%	29.24%	29.26%	25.88%
10	21.22%	20.67%	<u>17.11%</u>	21.66%	21.68%	21.19%	19.75%	30.12%	28.85%	<u>25.39%</u>	30.42%	30.18%	30.22%	27.08%
11	21.83%	21.34%	<u>17.96%</u>	22.19%	22.27%	21.71%	20.23%	31.34%	30.24%	<u>26.81%</u>	31.64%	31.75%	31.33%	29.38%
12	22.63%	21.96%	<u>18.10%</u>	22.95%	23.01%	22.37%	20.24%	31.70%	30.12%	<u>26.23%</u>	31.82%	31.93%	31.51%	27.62%
13	22.81%	22.08%	<u>17.60%</u>	23.02%	23.05%	22.43%	19.81%	31.39%	30.28%	<u>25.26%</u>	31.39%	31.42%	30.90%	27.74%
14	23.37%	22.73%	<u>18.02%</u>	23.52%	23.52%	22.98%	20.63%	31.54%	30.65%	<u>24.64%</u>	31.51%	31.43%	30.83%	27.09%
15	23.58%	23.03%	<u>18.21%</u>	23.90%	23.80%	23.27%	20.53%	31.33%	30.45%	<u>25.53%</u>	31.10%	30.67%	30.83%	28.62%
16	23.23%	22.73%	<u>17.73%</u>	23.58%	23.47%	23.06%	20.83%	29.31%	28.33%	<u>24.71%</u>	29.00%	28.69%	29.02%	27.32%
17	20.81%	20.47%	<u>16.64%</u>	21.38%	21.21%	20.81%	19.61%	26.90%	25.95%	<u>23.78%</u>	27.33%	26.96%	27.12%	26.06%
18	19.26%	18.99%	<u>15.75%</u>	20.12%	19.81%	19.31%	18.94%	24.38%	23.96%	<u>21.62%</u>	25.31%	24.73%	24.65%	27.24%
19	18.79%	18.48%	<u>15.82%</u>	19.74%	19.40%	18.82%	18.77%	24.18%	24.01%	<u>21.80%</u>	25.18%	24.70%	24.28%	25.95%
20	18.35%	18.13%	<u>15.47%</u>	19.16%	18.97%	18.29%	18.23%	24.80%	25.22%	<u>21.82%</u>	25.72%	25.35%	24.68%	26.58%
21	18.52%	18.12%	<u>15.41%</u>	19.22%	19.12%	18.44%	18.09%	26.68%	27.05%	<u>23.37%</u>	27.41%	27.39%	26.50%	28.27%
22	18.35%	17.97%	<u>14.93%</u>	19.00%	18.91%	18.19%	17.60%	28.10%	27.95%	<u>24.36%</u>	28.93%	28.90%	27.70%	27.64%
23	18.54%	18.29%	<u>15.06%</u>	19.21%	19.05%	18.31%	17.27%	26.57%	26.05%	<u>23.16%</u>	27.10%	26.84%	26.05%	26.35%
24	18.93%	18.54%	<u>15.64%</u>	19.76%	19.34%	18.63%	17.87%	25.60%	24.36%	<u>21.87%</u>	25.99%	25.12%	24.91%	23.85%
Ave.	20.89%	20.42%	<u>16.87%</u>	21.30%	21.27%	20.26%	19.64%	30.07%	28.82%	<u>24.70%</u>	30.45%	30.47%	29.86%	27.90%
R^2	59.96%	62.65%	<u>70.75%</u>	59.03%	59.03%	50.68%	61.02%	73.50%	73.90%	<u>80.47%</u>	71.78%	71.73%	67.67%	72.26%

Note: The best performance for each hour is underlined and the second best is marked with boldface for each node.

Abbreviations: AFAR, adaptive FAR model; CFPC, common functional principal component; FAR, functional autoregressive; FPC, functional principle component; pFAR, partial FAR; WFAR, warping FAR.

the AFAR model is flexible and can be safely used for both stationary and nonstationary functional data, the superior prediction performance of the AFAR over the FAR models indicates that the gas flow data is likely to be nonstationary, a situation which the stationary models are not capable of handling. Moreover, when the forecast horizon increases, the deviation of reality (nonstationary data) from the modeling assumption (stationary FAR models) becomes severe and leads to poor forecast accuracy as we observe in this illustration. The AFAR, though delivering lower accuracy given the increase of forecast horizon, is taking advantage of being able to identify local homogeneous intervals in the forecasting experiment and thus still provides relatively stable accuracy. As such the relative forecast accuracy of AFAR over the FAR models improved for longer term forecasting.

Overall, we can see some trade-off between the complexity of the modeling and the accuracy of the forecasting results. The models with complex but heuristic structure such as pFAR, AFAR outperform simple models such as FAR (1) models. Furthermore, the prediction results obtained with the WFAR model are close to the FAR(1) model at both nodes, as there is no evidence of seasonal variations in natural gas curves and the warping function is estimated to be close to one. The factor modeling, FPC and CFPC, do not show outstanding performance of gas flows at two nodes for both forecasting horizons. This is possibly because the two models rely on independence assumption when extracting factors from covariance decomposition, whereas the data are not independent but serially correlated. Also, there is neither significant improvement in modeling with two series of curves (CFPC) compared to modeling with individual series (FPC). This may be explained by the insignificant joint dynamic and cross-dependence within the two series.

In summary, the functional time series models produce accurate and stable out-of-sample forecast accuracy for short term gas curve prediction across the two nodes with various features. For long term forecast, the performances of all models become weaker. Despite that FAR model with large-scale exogenous variables of mixed data type, pFAR, shows the most successful forecast performance in both nodes and both forecast horizons. Outstanding performance of the pFAR model benefited from the consideration of serial dependence of stochastic processes and the incorporation of essential functional and scalar covariates simultaneously. Furthermore, adaptively selecting the intervals with time-varying lengths contributes to the success of AFAR model for stationary as well as nonstationary functional data, and the benefit becomes more significant for longer term forecasting. It is advantageous to consider proper lags for FAR type models. Although factor model based on (C)FPC techniques as well as WFAR models only provide competitive performance in natural gas forecasting, it is a data-driven method and can be applied in other applications of functional time series with higher dimensions and seasonal variation.

The relative forecast accuracy of the alternative models depends on the data's features. For example, while the functional time series models generally provide better prediction accuracy for M1 than M2, the pFAR seems superior for M2. This is because the pFAR model benefits from incorporating some exogenous variables—such as temperature and renewable energies—that are more relevant to predict the movement of gas in M2 than that in M1, whereas other FAR models do not take into account the exogenous variables.

It is worth noting that the prediction results in this review paper do not give a general ranking for the FAR based models. The best-performing model depends on the particular application. We would like to remind readers that the relative performance of the models may also vary for different data and different time windows. Meanwhile, we noticed that in the previous works of Chen and Li (2017), the impact of data sampling has been investigated and the similar outperformance of FAR is observed.

5 | CONCLUSION

In this paper, we provide a review study on the recently developed FAR modeling for univariate and multivariate functional data with serial dependence in both stationary and nonstationary framework. Specifically, we present the FAR models and their variations that can be applied under different scenarios. These models include the classic FAR model under stationarity, the FARX and pFAR model dealing with multiple exogenous functional variables, and large-scale mixed-type exogenous variables respectively, the VFAR model and CFPC technique to handle multiple dimensional functional time series, and the WFAR, VC-FAR and AFAR model to solve seasonal variations, slow varying effects and the more challenging cases of structural changes or breaks respectively. We introduce three smoothing (expansion) methods and elaborate the statistical techniques of projecting the series of curves onto finite parameter space and estimating the dynamics with sieve. We provide details of the model setup and estimation procedure with a closed-form estimator or penalized estimator for different models.

We discuss the models' applicability and illustrate the numerical performance along with real data of natural gas flows in the high-pressure gas pipeline network in Germany. We conduct 1- and 14-days-ahead out-of-sample forecasts of the gas flow curves at two municipal energy supplier-nodes. We measure forecast accuracy in terms of MAPE and out-of-sample R^2 . We find that the functional time series models generally produce competitive and stable out-of-sample forecast accuracy for the gas flow curves at both nodes. In particular, the FAR model with large-scale exogenous variables of mixed data type, pFAR, shows good forecast performance benefiting from the consideration of serial dependence of stochastic processes and the incorporation of essential functional and scalar covariates simultaneously. Furthermore, it is also advantageous to consider proper lags for FAR type models, and adaptively select the local homogeneity intervals for both stationary and nonstationary functional series using the AFAR model. Although factor models based on (C)FPC techniques and WFAR models only provide competitive performance in natural gas forecasting, it is data-driven and can be applied in different applications of high-dimensional functional time series with seasonal variation.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS

Ying Chen: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; resources; supervision; validation; visualization; writing-original draft; writing-review and editing. **Thorsten Koch:** Data curation; formal analysis; resources; writing-original draft; writing-review and editing. **Kian Guan Lim:** Conceptualization; resources; writing-original draft; writing-review and editing. **Xiaofei Xu:** Conceptualization; investigation; methodology; visualization; writing-original draft; writing-review and editing. **Nazgul Zakiyeva:** Conceptualization; visualization; writing-original draft; writing-review and editing.

ORCID

Nazgul Zakiyeva  <https://orcid.org/0000-0001-9106-9916>

RELATED WIREs ARTICLES

[Time series factor models](#)

[Autoregressive processes](#)

REFERENCES

- Aneiros, G., Ferraty, F., & Vieu, P. (2015). Variable selection in partial linear regression with functional covariate. *Statistics*, *49*, 1322–1347.
- Antoniadis, A., Paparoditis, E., & Sapatinas, T. (2006). A functional wavelet-kernel approach for time series prediction. *Journal of the Royal Statistical Society, Series B*, *68*, 837–857.
- Antoniadis, A., & Sapatinas, T. (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *Journal of Multivariate Analysis*, *87*, 133–158.
- Belomestny, D., & Spokoiny, V. (2007). Spatial aggregation of local likelihood estimates with applications to classification. *The Annals of Statistics*, *35*, 2287–2311.
- Besse, P., Cardot, H., & Stephenson, D. (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, *27*, 673–687.
- Bosq, D. (1991). Nonparametric functional estimation and related topics. In G. Roussas (Ed.), *NATO Science Series C* (pp. 509–529). Berlin: Springer.
- Bosq, D. (2000). *Linear processes in function spaces: Theory and applications* (Vol. 149). New York, NY: Springer.
- Bosq, D. (2007). General linear processes in Hilbert spaces and prediction. *Journal of Statistical Planning and Inference*, *137*, 879–894.
- Brockwell, P., & Davis, R. (1991). *Time series: Theory and methods*. New York, NY: Springer.
- Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*, *6*, 5549–5632.
- Chen, X., & Shen, X. (1998). Sieve extremum estimates for weakly dependent data. *Econometrica*, *66*, 289–314.
- Chen, Y., Chua, W. S., & Härdle, W. K. (2019). Forecasting limit order book liquidity supply–demand curves with functional autoregressive dynamics. *Quantitative Finance*, *19*, 1473–1489.
- Chen, Y., Chua, W. S., & Koch, T. (2018). Forecasting day-ahead high-resolution natural-gas demand and supply in Germany. *Applied Energy*, *228*, 1091–1110.
- Chen, Y., Koch, T., & Xu, X. (2019). *Regularized partially functional autoregressive model*. Retrieved from <http://ssrn.com/abstract=3482262>
- Chen, Y., & Li, B. (2017). An adaptive functional autoregressive forecast model to predict electricity price curves. *Journal of Business & Economic Statistics*, *35*(3), 371–388.

- Chen, Y., Marron, J., & Zhang, J. (2019). Modeling seasonality and serial dependence of electricity price curves with warping functional autoregressive dynamics. *The Annals of Applied Statistics*, 13(3), 1590–1616.
- Damon, J., & Guillas, S. (2002). The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics*, 13, 759–774.
- Dedecker, J., & Merlevède, F. (2002). Necessary and sufficient conditions for the conditional central limit theorem. *Annals of Probability*, 30, 1044–1081.
- Dedecker, J., & Merlevède, F. (2003). The conditional central limit theorem in Hilbert spaces. *Stochastic Processes and their Applications*, 108, 229–262.
- Didericksen, D., Kokoszka, P., & Zhang, X. (2012). Empirical properties of forecasts with the functional autoregressive model. *Computational Statistics*, 27, 285–298.
- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics*, 32, 407–499.
- Farindon, P. (2011). Linear processes for functional data. In F. Ferraty & Y. Romain (Eds.), *The Oxford handbook of functional data* (pp. 47–71). Verlag: Springer.
- Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis: Theory and practice*. Berlin: Springer.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). *A note on the group Lasso and a sparse group Lasso*. Retrieved from arXiv:1001.0736.
- Geman, S., & Hwang, C.-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *The Annals of Statistics*, 10, 401–414.
- Grenander, U. (1981). *Abstract inference*. New York, NY: Wiley.
- Hörmann, S. (2013). A functional version of the ARCH model. *Econometric Theory*, 29, 267–288.
- Hörmann, S., Kokoszka, P., & Nisol, G. (2018). Testing for periodicity in functional time series. *The Annals of Statistics*, 46, 2960–2984.
- Hörmann, S., & Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, 38(3), 1845–1884.
- Horváth, L., Kokoszka, P., & Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, 179, 66–82.
- Huang, J., Breheny, P., & Ma, S. (2012). A selective review of group selection in high-dimensional models. *Statistical Science*, 27, 481–499.
- Ivanoff, S., Picard, F., & Rivoirard, V. (2016). Adaptive Lasso and group-Lasso for functional Poisson regression. *The Journal of Machine Learning Research*, 17, 1903–1948.
- Kargin, V., & Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, 99, 2508–2526.
- Kokoszka, P., & Reimherr, M. (2013). Determining the order of the functional autoregressive model. *Journal of Time Series Analysis*, 34, 116–129.
- Kong, D., Xue, K., Yao, F., & Zhang, H. (2016). Partially functional linear regression in high dimensions. *Biometrika*, 103, 147–159.
- Kosiorowski, D., Mielczarek, D., Rydlewski, J., & Snarska, M. (2014). *Aspects of functional data analysis in short term prediction of non-stationary economic time series*. Pisa: ERCIM.
- Li, Y., Nan, B., & Zhu, J. (2015). Multivariate sparse group Lasso for the multivariate multiple linear regression with an arbitrary group structure. *Biometrics*, 71, 354–363.
- Liu, X., Xiao, H., & Chen, R. (2016). Convolutional autoregressive models for functional time series. *Journal of Econometrics*, 194, 263–282.
- Mas, A. (2007). Weak convergence in the functional autoregressive model. *Journal of Multivariate Analysis*, 98, 1231–1261.
- Mas, A., & Menneteau, L. (2003). Large and moderate deviations for infinite-dimensional autoregressive processes. *Journal of Multivariate Analysis*, 87, 241–260.
- Mas, A., & Pumo, B. (2018). Linear processes for functional data. In *The Oxford handbook of functional data analysis*. Oxford, England: Oxford University Press.
- Matsui, H., Kawano, S., & Konishi, S. (2009). Regularized functional regression modeling for functional response and predictors. *Journal of Mathematics for Industry*, 1, 17–25.
- Matsui, H., & Konishi, S. (2011). Variable selection for functional regression models via the l1 regularization. *Computational Statistics & Data Analysis*, 55, 3304–3310.
- Menneteau, L. (2005). Some laws of the iterated logarithm in Hilbertian autoregressive models. *Journal of Multivariate Analysis*, 92, 405–425.
- Mokhtari, F., & Mourid, T. (2003). Prediction of continuous time autoregressive processes via the reproducing kernel spaces. *Statistical Inference for Stochastic Processes*, 6, 247–266.
- Mourid, T., & Bensmain, N. (2006). Sieves estimator of the operator of a functional autoregressive process. *Statistics & Probability Letters*, 76, 93–108.
- Müller, H.-G., & Stadtmüller, U. (2005). Generalized functional linear models. *The Annals of Statistics*, 33, 774–805.
- Programm, C. C. (1982). *Canadian climate normals 1981–1980*. Ottawa: Environment Canada.
- Pumo, B. (1998). Prediction of continuous time processes by $C[0, 1]$ -valued autoregressive process. *Statistical Inference for Stochastic Processes*, 1, 297–309.
- Ramosay, J., & Dalzell, C. (1991). Some tools for functional data analysis (with discussion). *Journal of the Royal Statistical Society, Series B*, 53, 539–572.
- Ramsay, J. O., & Silverman, B. W. (2002). *Applied functional data analysis: Methods and case studies*. Berlin: Springer.
- Ramsay, J. O., & Silverman, B. W. (2005). Functional data analysis. In *Springer series in statistics* (2nd ed.). New York, NY: Springer.
- Rice, J. A., & Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B*, 53, 233–243.

- Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2013). A sparse-group Lasso. *Journal of Computational and Graphical Statistics*, 22, 231–245.
- Srivastava, A., Wu, W., Kurtek, S., Klassen, E., & Marron, J. S. (2011). *Registration of functional data using Fisher–Rao metric*. Retrieved from arXiv:1103.3817.
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58, 267–288.
- Tuddenham, R. D. (1954). Physical growth of California boys and girls from birth to eighteen years. *University of California Publications in Child Development*, 1, 183–364.
- Ullah, S., & Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology*, 13, 43.
- Wang, H., Li, R., & Tsai, C.-L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, 94, 553–568.
- Wang, J.-L., Chiou, J.-M., & Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3, 257–295.
- Xu, M., Li, J., & Chen, Y. (2017). Varying coefficient functional autoregressive model with application to the US treasuries. *Journal of Multivariate Analysis*, 159, 168–183.
- Xu, X., & Ghosh, M. (2015). Bayesian variable selection and estimation for group Lasso. *Bayesian Analysis*, 10, 909–936.
- Yao, F., Müller, H.-G., & Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100, 577–590.
- Zhang, J., Chen, Y., Klotz, S., & Lim, K. G. (2017). International yield curve prediction with common functional principal component analysis. In *Robustness in econometrics* (pp. 287–304). Berlin: Springer.