MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://monash.edu/business/ebs/research/publications

Seasonal functional autoregressive models

Atefeh Zamani, Hossein Haghbin, Maryam Hashemi, and Rob J Hyndman

August 2019

Working Paper 16/19

AACSB ACCREDITED EFMD EQUIS ACCREDITED ASSOCIATION OF MBAS ACCREDITED

Seasonal functional autoregressive models

Atefeh Zamani
Department of Statistics, Shiraz University, Shiraz, Iran.

Hossein Haghbin
Department of Statistics, Persian Gulf University, Boushehr, Iran.

Maryam Hashemi
Department of Statistics, University of Khansar, Khansar, Iran.

Rob J Hyndman
Department of Econometrics and Business Statistics,
Monash University, Melbourne, Australia.
Email: rob.hyndman@monash.edu

17 August 2019

JEL classification: C32,C14

Abstract

Functional autoregressive models are popular for functional time series analysis, but the standard formulation fails to address seasonal behaviour in functional time series data. To overcome this shortcoming, we introduce seasonal functional autoregressive time series models. For the model of order one, we derive sufficient stationarity conditions and limiting behavior, and provide estimation and prediction methods. Some properties of the general order P model are also presented. The merits of these models are demonstrated using simulation studies and via an application to real data.

Keywords: Functional time series analysis, seasonal functional autoregressive model, central limit theorem, prediction, estimation.

1 Introduction

Improved acquisition techniques make it possible to collect a large amount of high-dimensional data, including data that can be considered functional (Ramsay, 1982). Functional time series arise when such data are collected over time. We are interested in functional time series which exhibit seasonality. The underlying functional process is denoted by $f_t(x)$, where $t = 1, \dots, T$ indexes regularly spaced time and x is a continuous variable.

A seasonal pattern exists when $f_t(x)$ is influenced by seasonal factors (e.g., the quarter of the year, the month, the day of the week, etc.). Usually seasonality is considered to have a fixed and known period. For example, consider satellite observations measuring the normalized difference vegetation index (NDVI) (He, 2018). These are often averaged to obtain monthly observations over the land surface. Here x denotes the two spatial dimensions, while t denotes the month. Seasonal patterns are present due to the natural annual patterns of vegetation variation.

Another example arises in demography where $f_t(x)$ is the mortality rate for people aged x at time t (Hyndman & Ullah, 2007). When such data are collected more frequently than annually, seasonality arises due to deaths being influenced by weather.

In other applications, x may denote a second time variable. For example, Hörmann, Kokoszka, Nisol, et al., 2018 study pollution dat observed every 30 minutes. The long time series is sliced into separate functions, where x denotes the time of day, and t denotes the day. A similar approach has been applied to the El Niño-Southern Oscillation (ENSO) (Besse, Cardot

& Stephenson, 2000), Eurodollar futures rates (Kargin & Onatski, 2008; Horváth, Kokoszka & Reeder, 2013), electricity demand (Shang, 2013), and many other applications.

Although, the term "functional data analysis" was coined by Ramsay (1982), the history of this area is much older and dates back to Grenander (1950) and Rao (1958). Functional data cannot be analyzed using classical statistical tools and need appropriate new techniques in order to be studied theoretically and computationally.

A popular functional time series model is the functional autoregressive (FAR) process introduced by Bosq (2000), and further studied by Hörmann, Kokoszka, et al. (2010), Horváth, Hušková & Kokoszka (2010), Horváth & Kokoszka (2012), Berkes, Horváth & Rice (2013), Hörmann, Horváth & Reeder (2013) and Aue, Norinho & Hörmann (2015).

Although these models are applied in the analysis of various functional time series, they cannot handle seasonality adequately. For example, althouugh it seems that th traffic flow follows a weakly pattern, Klepsch, Klüppelberg & Wei (2017) applied functional ARMA processes for modeling highway traffic data.

A popular model for seasonal univariate time series, $\{X_1, \ldots, X_T\}$, is the class of seasonal autoregressive processes denoted by $\text{SAR}(P)_S$, where S is the seasonal period and P is the order of the autoregression. These models satisfy the following equation:

$$X_t = \phi_1 X_{t-S} + \phi_2 X_{t-2S} + \cdots + \phi_P X_{t-PS} + \varepsilon_t,$$

where $\phi(x) = 1 - \phi_1 x^S - \phi_2 x^{2S} - \cdots - \phi_P x^{PS}$ is the seasonal characteristic polynomial and ε_t is independent of $X_{t-1}, X_{t-2}, \ldots$. For stationarity, the roots of $\phi(x) = 0$ must be greater than 1 in absolute value. This model is a special case of the AR(p) model, which is of order $p = PS$, with nonzero ϕ-coefficients only at the seasonal lags $S, 2S, 3S, \ldots, PS$.

In this paper, we propose a class of seasonal functional AR models, which are analogous to seasonal autoregressive models. We present some notation and definitions in Section 2. Section 3 introduces the seasonal functional AR(1) model and discusses some of its properties. Estimation of the parameters of this model is studied in Section 4 and the prediction problem is considered in Section 5. In Section 6, the more general seasonal functional AR(P) model is introduced and some its basic properties are scrutinized. Section 7 is devoted to simulation studies and real data analysis. We conclude in Section 9.

2 Preliminary notations and definitions

Let $H = L^2([0,1])$ be the separable real Hilbert space of square integrable functions $x : [0,1] \to \mathbb{R}$ with the inner product $\langle x, y \rangle = \int_0^1 x(t)y(t)dt$ and the norm $\|x\| = \left(\int_0^1 x^2(t)dt\right)^{1/2}$. Let $\mathcal{B}$ denote the Borel field in H, and $(\Omega, \mathcal{F}, P)$ stand for a probability space. A random function in H is an $\mathcal{F}/\mathcal{B}$ measurable mapping from Ω into H.

Additionally, let $\mathcal{L}(H)$ denote the space of all continuous linear operators from H to H, with operatorial norm $\|\cdot\|_{\mathcal{L}}$. An important subspace of $\mathcal{L}(H)$ is the space of Hilbert-Schmidt operators, $\mathcal{HS}(H)$, which forms a Hilbert space equipped with the inner product $\langle A, B \rangle_{\mathcal{HS}} = \sum_{k=1}^{\infty} \langle A\phi_k, B\phi_k \rangle$ and the norm $\|A\|_{\mathcal{HS}} = \left\{\sum_{k=1}^{\infty} \|A\phi_k\|^2\right\}^{1/2}$, where $\{\phi_k\}$ is any orthonormal basis on H. The space of nuclear or trace class operators, $\mathcal{N}(H)$, is a notable subclass of $\mathcal{HS}(H)$ and the associated norm is defined as

$$\|A\|_{\mathcal{N}} = \sum_{k=1}^{\infty} \langle |A| \phi_k, \phi_k \rangle = \sum_{k=1}^{\infty} \left\langle (A^* A)^{1/2} \phi_k, \phi_k \right\rangle, \tag{2.1}$$

where A^* is the adjoint of A (Conway, 2000). If A is a self-adjoint nuclear operator with associated eigenvalues λ_k, then $\|A\|_{\mathcal{N}} = \sum_{k=1}^{\infty} |\lambda_k|$. If, in addition, A is non-negative, then $\|A\|_{\mathcal{N}} = \sum_{k=1}^{\infty} \langle A\phi_k, \phi_k \rangle = \sum_{k=1}^{\infty} \lambda_k$. Note that $\|\cdot\|_{\mathcal{L}} \le \|\cdot\|_{\mathcal{HS}} \le \|\cdot\|_{\mathcal{N}}$ (Hsing & Eubank, 2015).

For x and y in H, the tensorial product of x and y, $x \otimes y$, is a nuclear operator and is defined as $(x \otimes y)z := \langle y, z \rangle x$, $z \in H$. Here, we point out some relations which are simple consequences of the definition of $x \otimes y$ and will be applied in the following sections:

$$\begin{aligned}
(x \otimes y)^* &= y \otimes x, \\
(Ax) \otimes y &= A\,(x \otimes y), \\
x \otimes (Ay) &= (x \otimes y)\,A^*,
\end{aligned}$$

where A is an operator and * stands for the adjoint of an operator (Conway, 2000).

Let $\mathbb{Z}$ denote the set of integers. Then, we define a *functional discrete time stochastic process* as a sequence of random functions in H, namely $\boldsymbol{X} = \{X_n,\ n \in \mathbb{Z}\}$. A random function X in H is said to be strongly second order if $\mathbb{E}\|X\|^2 < \infty$. Similarly, a functional discrete time stochastic process $\boldsymbol{X}$ is said to be strongly second order if every X_n is strongly second order. Let $L_H^2(\Omega, \mathcal{F}, P)$ stand for the Hilbert space of all strongly second order random function on the probability space $(\Omega, \mathcal{F}, P)$.

For the random function X_t, the mean function is denoted by $\mu_t := \mathbb{E}(X_t)$ and is defined in terms of Bochner integral. For any $t, t' \in \mathbb{Z}$, the covariance operator is defined as $C^X_{t,t'} := \mathbb{E}\left[(X_t - \mu_t) \otimes (X_{t'} - \mu_{t'})\right]$. Besides, as an integral operator, $C^X_{t,t'}$ can be represented as

$$C^X_{t,t'} h(s) = \int_0^1 C^X_{t,t'}(s,s')\, h(s')\, ds', \quad s, s' \in [0,1] \quad \text{and} \quad t, t' \in \mathbb{Z},$$

where $C^X_{t,t'}(s,s') := \mathbb{E}\left[(X_t(s) - \mu_t(s))(X_{t'}(s') - \mu_{t'}(s'))\right]$ is the corresponding covariance kernel. When the process is stationary, we will denote $C^X_{t,t'}$ as $C^X_{t-t'}$ and if no confusion arises, we will drop superscript X.

As in any time series analysis, functional white noise processes are of great importance in functional time series analysis.

Definition 2.1. *A sequence $\boldsymbol{\varepsilon} = \{\varepsilon_t,\ t \in \mathbb{Z}\}$ of random functions is called functional white noise if*

(i) $0 < E\,\|\varepsilon_t\|^2 = \sigma^2 < \infty$, $\mathbb{E}(\varepsilon_t) = 0$ *and* $C^\varepsilon_t := C^\varepsilon_0$ *do not depend on* t,

(ii) ε_t *is orthogonal to* $\varepsilon_{t'}$, $t, t' \in \mathbb{Z}$, $t \neq t'$; *i.e.,* $C^\varepsilon_{t,t'} = 0$.

The following definitions will be applied in the subsequent sections.

Definition 2.2. *Let X_n be a sequence of random functions. We say that X_n converges to X in $L^2_H(\Omega, \mathcal{F}, P)$ if $\mathbb{E}\|X_n - X\|^2 \to 0$, as n goes to infinity.*

Definition 2.3. *$\mathcal{G}$ is said to be an $\mathcal{L}$-closed subspace (or hermetically closed subspace) of $L^2_H(\Omega, \mathcal{F}, P)$ if $\mathcal{G}$ is a Hilbertian subspace of $L^2_H(\Omega, \mathcal{F}, P)$ and, if $X \in \mathcal{G}$ and $\ell \in \mathcal{L}$, then $\ell(X) \in \mathcal{G}$. A zero-mean LCS is an $\mathcal{L}$-closed subspace which contains only zero-mean random functions.*

Moreover, let $H^p = \left(L^2([0,1])\right)^p$ denote the product Hilbert space equipped with the inner product $\langle \mathbf{x}, \mathbf{y} \rangle_p = \sum_{i=1}^p \langle x_i, y_i \rangle$ and the norm $\|\mathbf{x}\|_p = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_p}$. We denote by $\mathcal{L}(H^p)$ the space of bounded linear operators on H^p.

3 The SFAR$(1)_S$ model

Following the development of univariate seasonal autoregressions in Harrison (1965), Chatfield & Prothero (1973) and Box et al. (2015), we define the seasonal functional autoregressive process of order one as follows.

Definition 3.1. *A sequence* $\boldsymbol{X} = \{X_t;\ t \in \mathbb{Z}\}$ *of random functions is said to be a pure seasonal functional autoregressive process of order 1 with seasonality S (SFAR(1)$_S$) associated with* $(\mu, \boldsymbol{\varepsilon}, \phi)$ *if*

$$X_t - \mu = \phi\left(X_{t-S} - \mu\right) + \varepsilon_t, \tag{3.1}$$

where $\boldsymbol{\varepsilon} = \{\varepsilon_t;\ t \in \mathbb{Z}\}$ *is a functional white noise process,* $\mu \in H$ *and* $\phi \in \mathcal{L}(H)$*, with* $\phi \neq 0$.

The SFAR(1)$_S$ processes can be studied from two different perspectives. As the first viewpoint, a SFAR(1)$_S$ model is an FAR(S) model with most coefficients equal to zero. This point of view will be applied when dealing with basic properties of these processes in the next subsection. In the other perspective, SFAR(1)$_S$ processes are studied as a special case of an autoregressive time series of order one with values in the product Hilbert space H^S, which will be used while studying the limit theorems of such processes.

For simplicity of notation, let us consider μ as zero. In this case, as there is no intercept, the unconditional mean of the process regardless of the season is equal to zero. However, the conditional mean on the past of X_t depends in S, since $\mathbb{E}\left(X_t | X_{t-1}, \dots\right) = \phi X_{t-S}$ (Ghysels & Osborn, 2001).

3.1 Basic properties

In order to study the existence of the process $\boldsymbol{X}$, the following assumption is required:

Assumption 1: There exists an integer $j_0 \geq 1$ such that $\left\|\phi^{j_0}\right\|_{\mathcal{L}} < 1$.

In the following, we will call an SFAR(1)$_S$ a standard time series if $\mu = 0$ ans Assumption 1 holds.

Theorem 3.1. *Let* X_t *be a standard SFAR(1)$_S$ time series. Then,*

$$X_t = \phi X_{t-S} + \varepsilon_t, \tag{3.2}$$

has a unique stationary solution given by

$$X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-jS}, \quad t \in \mathbb{Z}, \tag{3.3}$$

where the series converges in $L^2_H(\Omega, \mathcal{F}, P)$ *and with probability 1 and* $\boldsymbol{\varepsilon}$ *is the functional white noise process of* X_t.

Proof. As the first step, we will prove that the series in (3.3) converges in $L^2_H(\Omega, \mathcal{F}, P)$. For this purpose, let $X_t^m = \sum_{j=0}^{m} \phi^j \varepsilon_{t-jS}$. By orthogonality of the process ε and for $m' > m$, we have:

$$\mathbb{E}\left\|X_t^{m'} - X_t^m\right\|^2 = \mathbb{E}\left\|\sum_{j=m+1}^{m'} \phi^j \varepsilon_{t-Sj}\right\|^2 = \sum_{j=m+1}^{m'} \mathbb{E}\left\|\phi^j \varepsilon_{t-Sj}\right\|^2.$$

On the other hand,

$$\mathbb{E}\left\|\phi^j \varepsilon_{t-Sj}\right\|^2 \leq \left\|\phi^j\right\|^2_{\mathcal{L}} \mathbb{E}\left\|\varepsilon_{t-Sj}\right\|^2 = \sigma^2 \left\|\phi^j\right\|^2_{\mathcal{L}},$$

and, consequently, $\mathbb{E}\left\|X_t^{m'} - X_t^m\right\|^2 \leq \sigma^2 \sum_{j=m+1}^{m'} \left\|\phi^j\right\|^2_{\mathcal{L}}$, which tends to zero as m and m' tends to infinity (See Lemma 3.1 Bosq, 2000). Therefore, by Cauchy criterion, it can be concluded that (3.3) converges in $L^2_H(\Omega, \mathcal{F}, P)$.

Consider the process $Y_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-Sj}$. Using the boundedness of ϕ, it can be seen that

$$\begin{aligned}
Y_t - \phi Y_{t-S} &= \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-Sj} - \sum_{j=0}^{\infty} \phi^{j+1} \varepsilon_{t-S-Sj} \\
&= \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-Sj} - \sum_{j'=1}^{\infty} \phi^{j'} \varepsilon_{t-Sj'} \\
&= \varepsilon_t,
\end{aligned}$$

which means that Y_t is a solution of (3.1). Conversely, let X_t be a solution of (3.1). It can be shown that

$$\begin{aligned}
X_t &= \phi X_{t-S} + \varepsilon_t \\
&= \phi^2 X_{t-2S} + \phi \varepsilon_{t-S} + \varepsilon_t \\
&= \dots \\
&= \phi^{k+1} X_{t-(k+1)S} + \sum_{j=0}^{k} \phi^j \varepsilon_{t-Sj}
\end{aligned}$$

Therefore, by stationarity of X_t,

$$\mathbb{E}\left\|X_t - \sum_{j=0}^{k} \phi^j \varepsilon_{t-Sj}\right\|^2 \leq \left\|\phi^{k+1}\right\|^2_{\mathcal{L}} \mathbb{E}\left\|X_{t-(k+1)S}\right\|^2$$

which goes to zero as k tends to infinity. This inequality demonstrates that $X_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-jS}$, and proves uniqueness of this solution. $\square$

In time series analysis, the covariance operator is of great importance and plays a crucial role in the analysis of the data. The following theorem demonstrates the structure of covariance operator for the SFAR$(1)_S$ processes.

Theorem 3.2. *If **X** is a standard SFAR(1)$_S$, the following relations hold:*

$$C_0^X = C_S^X \phi^* + C^\varepsilon = \phi C_0^X \phi^* + C^\varepsilon, \tag{3.4}$$

$$C_0^X = \sum_{j=0}^{\infty} \phi^j C^\varepsilon \phi^{*j}, \tag{3.5}$$

where C_S^X is the covariance operator of X at lag S and the series converges in the $\|\cdot\|_{\mathcal{N}}$ sense. Besides, for $t > t'$,

$$C_{t,t'}^X = \begin{cases} \phi^k C_0^X & \text{if } t - t' = kS \\ 0 & \text{Otherwise} \end{cases}, \tag{3.6}$$

and

$$C_{t',t}^X = \begin{cases} C_0^X \phi^{k*} & \text{if } t - t' = kS \\ 0 & \text{Otherwise} \end{cases}, \tag{3.7}$$

where k is some positive integer.

Proof. For each x in H, we have:

$$\begin{aligned} C_0^X &= \mathbb{E}\left(X_t \otimes X_t\right) \\ &= \mathbb{E}\left(X_t \otimes \left(\phi X_{t-S} + \varepsilon_t\right)\right) \\ &= \mathbb{E}\left(X_t \otimes X_{t-S}\right)\phi^* + \mathbb{E}\left(X_t \otimes \varepsilon_t\right) \\ &= C_S^X \phi^* + C^\varepsilon, \end{aligned}$$

and it demonstrates that $C_0^X = C_S^X \phi^* + C^\varepsilon$. Moreover,

$$\begin{aligned} C_0^X(x) &= \mathbb{E}\left(X_t \otimes X_t\right) \\ &= \mathbb{E}\left(\left(\phi X_{t-S} + \varepsilon_t\right) \otimes \left(\phi X_{t-S} + \varepsilon_t\right)\right) \\ &= \mathbb{E}\left(\phi X_{t-S} \otimes \phi X_{t-S}\right) + \mathbb{E}\left(\varepsilon_t \otimes \phi X_{t-S}\right) \\ &\quad + \mathbb{E}\left(\phi X_{t-S} \otimes \varepsilon_t\right) + \mathbb{E}\left(\varepsilon_t \otimes \varepsilon_t\right) \end{aligned}$$

$$= \phi C_0^X \phi^* + C^\varepsilon$$

which implies that $C_0^X = \phi C_0^X \phi^* + C^\varepsilon$.

As demonstrated in the proof of Theorem 3.1, $X_t = \sum_{j=0}^{k} \phi^j \varepsilon_{t-Sj} + \phi^{k+1} X_{t-(k+1)S}$, which results in that:

$$\begin{aligned}
C_0^X &= \mathbb{E}\left(X_t \otimes X_t\right) \\
&= \mathbb{E}\left(\left(\sum_{j=0}^{k} \phi^j \varepsilon_{t-Sj} + \phi^{k+1} X_{t-(k+1)S}\right) \otimes \left(\sum_{j=0}^{k} \phi^j \varepsilon_{t-Sj} + \phi^{k+1} X_{t-(k+1)S}\right)\right) \\
&= \sum_{j=1}^{k} \phi^j C^\varepsilon \phi^{j^*} + \phi^{(k+1)} C_0^X \phi^{(k+1)^*}.
\end{aligned}$$

Consequently, $C_0^X = \sum_{j=1}^{k} \phi^j C_0^\varepsilon \phi^{*j} + \phi^{(k+1)} C_0^X \phi^{*(k+1)}$. Since $\phi^{(k+1)} C_0^X \phi^{*(k+1)}$ is the covariance operator of $\phi^{(k+1)} X_0$, it is a nuclear operator and

$$\begin{aligned}
\left\|\phi^{(k+1)} C_0^X \phi^{*(k+1)}\right\|_{\mathcal{N}} &= \left\|\phi^{(k+1)} A^{1/2} A^{*1/2} \phi^{*(k+1)}\right\|_{\mathcal{N}} \\
&= \left\|\phi^{(k+1)} A^{1/2}\right\|_{\mathcal{HS}}^2 \\
&\leq \left\|\phi^{(k+1)}\right\|_{\mathcal{L}}^2 \|A\|_{\mathcal{HS}},
\end{aligned}$$

where $C_0^X = A^{1/2} A^{*1/2}$. Therefore, $\left\|C_0^X - \sum_{j=1}^{k} \phi^j C^\varepsilon \phi^{*j}\right\|_{\mathcal{N}} \leq \left\|\phi^{(k+1)}\right\|_{\mathcal{L}}^2 \|A\|_{\mathcal{HS}}$, which tends to zero as k goes to infinity, and results in that $C_0^X = \sum_{j=0}^{\infty} \phi^j C^\varepsilon \phi^{*j}$.

Let $t - t' = kS$, for some positive integer k. To demonstrate $C_{t,t'}^X = \phi^k C_0^X$, note that

$$\begin{aligned}
C_{t,t'}^X &= \mathbb{E}\left(X_t \otimes X_{t'}\right) \\
&= \mathbb{E}\left(X_{t'+kS} \otimes X_{t'}\right) \\
&= \mathbb{E}\left(\left(\phi X_{t'+(k-1)S} + \varepsilon_{t'+kS}\right) \otimes X_{t'}\right) \\
&\vdots \\
&= \mathbb{E}\left(\left(\phi^k X_{t'} + \varepsilon_{t'+S}\right) \otimes X_{t'}\right) \\
&= \phi^k C_0^X.
\end{aligned}$$

Concerning (3.7), it suffices to write

$$C_{t',t}^X = \left(C_{t,t'}^X\right)^* = \left[\phi^k C_0^X\right]^* = C_0^X \phi^{k^*},$$

and the proof is completed. □

Note 3.1. *The autocovariance operators characterize all the second-order dynamical properties of a time series and are the focus of time domain analysis of time series data. However, sometimes studying time series in the frequency domain can open new avenues in time series research. In this subsection, we are going to formulate the spectral density operator of* $\{X_t\}$*, based on Panaretos, Tavakoli, et al. (2013). If the autocovariance operators satisfy* $\sum_{t\in\mathbb{Z}} \|C_t^X\|_{\mathcal{N}} < \infty$*, then the spectral density operator of a stationary functional time series,* $\{X_t\}$*, at frequency* λ *(which is nuclear), is defined as*

$$\mathcal{F}_\lambda = \frac{1}{2\pi} \sum_{t\in\mathbb{Z}} e^{-i\lambda t} C_t^X, \tag{3.8}$$

where $\lambda \in \mathbb{R}$ *and the convergence holds in nuclear norm. Similarly, based on Theorem 3.2, for SFAR(1)*$_S$ *time series, if* $\sum_{t\in\mathbb{Z}} \|C_t^X\|_{\mathcal{N}} < \infty$*, then the spectral density operator at frequency* λ*, will be*

$$\begin{aligned} \mathcal{F}_\lambda &= \frac{1}{2\pi} \sum_{t\in\mathbb{Z}} e^{-i\lambda t} C_t^X, \\ &= \frac{1}{2\pi} \left\{ \phi C_0^X \phi^* + C^\varepsilon + \sum_{k=1}^{\infty} e^{-i\lambda k S} \phi^k C_0^X + \sum_{k=-\infty}^{-1} e^{-i\lambda k S} C_0^X \phi^{k*} \right\}. \end{aligned} \tag{3.9}$$

Consider ϕ be a symmetric compact operator. In this case, ϕ admits the spectral decomposition

$$\phi = \sum_{j=1}^{\infty} \alpha_j e_j \otimes e_j, \tag{3.10}$$

where $\{e_j\}$ is an orthonormal basis for H and $\{\alpha_j\}$ is a sequence of real numbers, such that $\lim_{j\to\infty} \alpha_j = 0$. If there exists j such that $\mathbb{E}\left(\langle \varepsilon_0, e_j\rangle\right)^2 > 0$, we can define the operator ϕ_ε, which connects $\boldsymbol{\varepsilon}$ and ϕ, as follows:

$$\phi_\varepsilon = \sum_{j=j_0}^{\infty} \alpha_j e_j \otimes e_j, \tag{3.11}$$

where j_0 is the smallest j satisfying $\mathbb{E}\left(\langle \varepsilon_0, e_j\rangle\right)^2 > 0$. If such a j_0 does not exists, we set $\phi_\varepsilon = 0$.

Theorem 3.3. *Let* ϕ *be a symmetric compact operator over* H *and* $\boldsymbol{\varepsilon}$ *be a functional white noise process. Then, equation* $X_t = \phi X_{t-S} + \varepsilon_t$ *has a stationary solution with innovation* $\boldsymbol{\varepsilon}$ *if and only if* $\|\phi_\varepsilon\|_{\mathcal{L}} < 1$.

Proof. The proof is an easy consequence of Theorem 3.5, Page 83, Bosq (2000). □

Theorem 3.4. *Let $\phi = \sum_{j=1} \alpha_j e_j \otimes e_j$ be a symmetric compact operator on H. Then, X_t is a SFAR(1)$_S$ model if and only if $\langle X_t, e_k \rangle$ is a $SAR(1)_S$ model.*

Proof. Let X_t be a pure SFAR(1)$_S$ process, i.e.,

$$X_t = \phi X_{t-S} + \varepsilon_t,$$

where $\boldsymbol{\varepsilon} = \{\varepsilon_t;\ t \in \mathbb{Z}\}$ is a functional white noise and $\phi \in \mathcal{L}(H)$. Then,

$$\begin{aligned}
\langle X_t, e_k \rangle &= \langle \phi X_{t-S} + \varepsilon_t, e_k \rangle \\
&= \langle X_{t-S}, \phi e_k \rangle + \langle \varepsilon_t, e_k \rangle \\
&= \alpha_k \langle X_{t-S}, e_k \rangle + \langle \varepsilon_t, e_k \rangle, \quad j \geq 1.
\end{aligned} \tag{3.12}$$

If $\alpha_k \neq 0$ and $\mathbb{E} \langle \varepsilon_t, e_k \rangle^2 > 0$, then $\langle X_t, e_k \rangle$ is a $SAR(1)$ model. If $\alpha_k = 0$, $\langle X_t, e_k \rangle = \langle \varepsilon_t, e_k \rangle$ and $\langle X_t, e_k \rangle$ is a degenerate $AR(1)$. Conversely, if (3.12) is holds for each $k \geq 1$, then we have

$$\begin{aligned}
\langle X_t, x \rangle &= \sum_{k=1} \langle X_t, e_k \rangle \langle x, e_k \rangle \\
&= \sum_{k=1} \alpha_k \langle X_{t-S}, e_k \rangle \langle x, e_k \rangle + \sum_{k=1} \langle \varepsilon_t, e_k \rangle \langle x, e_k \rangle \\
&= \sum_{k=1} \langle \phi X_{t-S} + \varepsilon_t, e_k \rangle \langle x, e_k \rangle \\
&= \langle \phi X_{t-S} + \varepsilon_t, x \rangle .
\end{aligned}$$

Consequently, $X_t = \phi X_{t-S} + \varepsilon_t$. □

3.2 Limit theorems

In this subsection, we will focus on some limiting properties of the SFAR(1)$_S$ model. Let us set $\mathbf{Y}_t = (X_t, \ldots, X_{t-S+1})'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_t, 0, \ldots, 0)'$, where 0 appears $S-1$ times. Define the operator $\boldsymbol{\rho}$ on H^S as:

$$\boldsymbol{\rho} = \begin{bmatrix} 0 & 0 & \cdots & 0 & \phi \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}, \tag{3.13}$$

where I denotes the identity operator on H. Consequently, we have the following simple but crucial lemma.

Lemma 3.1. *If $\boldsymbol{X}$ is a SFAR(1)$_S$ process, associated with (ε, ϕ), then $\boldsymbol{Y}$ is an autoregressive time series of order one with values in the product Hilbert space H^S associated with $(\boldsymbol{\varepsilon}, \boldsymbol{\rho})$, i.e.,*

$$\boldsymbol{Y}_t = \boldsymbol{\rho}\boldsymbol{Y}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{3.14}$$

Proof. This lemma is an immediate consequence of Lemma 5.1, Bosq (2000). □

The next theorem demonstrates the required condition for existence and uniqueness of the process $\boldsymbol{X}$ based on the operator $\boldsymbol{\rho}$.

Theorem 3.5. *Let $\mathcal{L}\left(H^S\right)$ denote the space of bounded linear operators on H^S equipped with the norm $\|\cdot\|_{\mathcal{L}^S}$. If*

$$\left\|\boldsymbol{\rho}^{j_0}\right\|_{\mathcal{L}^S} < 1, \quad \textit{for some } j_0 \geq 1, \tag{3.15}$$

then (3.1) has a unique stationary solution given by

$$X_t = \sum_{j=0}^{\infty} (\boldsymbol{\pi}\boldsymbol{\rho})\left(\boldsymbol{\varepsilon}_{t-j}\right), \quad t \in \mathbb{Z}, \tag{3.16}$$

where the series converges in $L^2_{H^S}(\Omega, \mathcal{F}, P)$ and with probability 1 and $\boldsymbol{\pi}$ is the projector of H^S onto H, defined as $\boldsymbol{\pi}\left(x_1, \ldots, x_S\right) = x_1$, $\left(x_1, \ldots, x_S\right) \in H^S$.

By the structure of $\boldsymbol{\rho}$, it can be demonstrated that $\boldsymbol{\rho}^S = \phi \boldsymbol{I}$, where $\boldsymbol{I}$ denotes the identity operator on H^S. Consequently, the following lemma is clear.

Lemma 3.2. *If $\|\phi\| < 1$, then (3.15) holds.*

We may now state the law of large number for the process $\boldsymbol{X}$.

Theorem 3.6. *If $\boldsymbol{X}$ is a standard SFAR(1)$_S$ then, as $n \to \infty$,*

$$\frac{n^{0.25}}{(\log n)^{\beta}} \frac{S_n}{n} \to 0, \quad \beta > 0.5. \tag{3.17}$$

Proof. The proof is an easy consequence of Theorem 5.6 of Bosq (2000). □

The central limit theorem for SFAR(1)$_S$ processes is established in the next theorem, which can be proved using Theorem 5.9 of Bosq (2000).

Theorem 3.7. *Let* $\boldsymbol{X}$ *be a standard SFAR(1)*$_S$ *associated with a functional white noise* $\boldsymbol{\varepsilon}$ *and such that* $I-\phi$ *is invertible. Then*

$$\frac{S_n}{\sqrt{n}} \to \mathcal{N}(0,\Gamma), \tag{3.18}$$

where

$$\Gamma = (I-\phi)^{-1} C^{\varepsilon} (I-\phi^*)^{-1}. \tag{3.19}$$

4 Parameter estimation

In this section, three methods are applied to estimate the parameter ϕ in model (3.1). These methods, namely method of moments, unconditional least squares estimation and the Kargin-Onatski method, are studied in brief in the following subsections.

4.1 Method of moments

Identification of SFAR(1)$_S$ takes place via estimation of parameters. The parameter can be estimated using the classical method of moments. It is easy to show that

$$\mathbb{E}\left(X_t \otimes X_{t-S}\right) = \mathbb{E}\left(\phi X_{t-S} \otimes X_{t-S}\right) + \mathbb{E}\left(\varepsilon_t \otimes X_{t-S}\right),$$

or equivalently, by stationarity of SFAR(1)$_S$,

$$C_S^X = \phi C_0^X + C_{t,t-S}^{\varepsilon,X}.$$

Since $C_S^{X,\varepsilon} = 0$, it can be concluded that

$$C_S^X = \phi C_0^X. \tag{4.1}$$

The operator C_0^X is compact and, consequently, it has the following spectral decomposition:

$$C_0^X = \sum_{m\in\mathbb{N}} \lambda_m \, \nu_m \otimes \nu_m, \quad \sum \lambda_m < \infty, \tag{4.2}$$

where $(\lambda_m)_{m\geqslant 1}$ is the sequence of the positive eigenvalues of C_0^X and $(\nu_m)_{m\geqslant 1}$ is a complete orthogonal system in H.

From equation (4.1), we have

$$C_S^X(\nu_j) = \phi C_0^X(\nu_j) = \lambda_j \phi(\nu_j).$$

Then, for any $x \in H$, the derived equation leads to the following representation:

$$\begin{aligned}\phi(x) &= \phi\left(\sum_{j=1}^{\infty} \langle x, \nu_j \rangle \nu_j\right) \\ &= \sum_{j=1}^{\infty} \langle x, \nu_j \rangle \phi(\nu_j) \\ &= \sum_{j=1}^{\infty} \frac{C_S^X(\nu_j)}{\lambda_j} \langle x, \nu_j \rangle \end{aligned} \tag{4.3}$$

Equation (4.3) gives a core idea for the estimation of ϕ. We thus estimate C_S^X, λ_j and ν_j from our empirical data and they will be further plugged into equation (4.3).

The estimated eigen-elements $(\hat{\lambda}_j, \hat{\nu}_j)_{1 \leqslant j \leqslant n}$ will be obtained from the empirical covariance operator $\hat{C}_0^X = \frac{1}{n}\sum_{j=1}^{n} X_j \otimes X_j$. Note that from the finite sample, we cannot estimate the entire sequence (λ_j, ν_j), rather we have to work with a truncated version. This leads to

$$\hat{\phi}(x) = \sum_{j=1}^{k_n} \frac{\hat{C}_S^X(\hat{\nu}_j)}{\hat{\lambda}_j} \langle x, \hat{\nu}_j \rangle, \tag{4.4}$$

where $\hat{C}_S^X = \frac{1}{n}\sum_{j=S}^{n} X_j \otimes X_{j-S}$ and the choice of k_n can be based on the cumulative percentage of total variance (CPV). .

Note 4.1. *For a sequence of zero mean $SFAR(1)_S$ processes, the Yule-Walker equations can be obtained as:*

$$C_h^X = \begin{cases} \phi C_{h-S}^X & \text{if } h = kS \\ 0 & \text{Otherwise} \end{cases}, \tag{4.5}$$

and

$$C_0^X = C_S^X \phi^* + C^{\varepsilon}, \tag{4.6}$$

where k is some positive integer. If the covariance operators are estimated using their empirical counterparts, the Yule-Walker equations can be applied for estimating the unknown parameters of the model.

4.2 Unconditional least square estimation method

In this section we obtain the least square estimation of autocorrelation parameter. Let C_0^X be the covariance operator of X_i, with the sequence of $\{(\lambda_i, \nu_i)\}$ as its eigen-values and eigen-functions. The idea is that the functional data can be represented by their coordinates with respect to the functional principal components (FPC) of the X_t, e.g. $X_{tk} = \langle X_t, \nu_k \rangle$, which is the projection of the tth observation onto the kth largest FPC. Therefore,

$$\begin{aligned} \langle X_t, \nu_k \rangle &= \langle \phi X_{t-S}, \nu_k \rangle + \langle \varepsilon_t, \nu_k \rangle \\ &= \sum_{j=1}^{\infty} \langle X_{t-S}, \nu_j \rangle \langle \phi(\nu_j), \nu_k \rangle + \langle \varepsilon_t, \nu_k \rangle \\ &= \sum_{j=1}^{p} \langle X_{t-S}, \nu_j \rangle \langle \phi(\nu_j), \nu_k \rangle + \delta_{tk}, \end{aligned} \tag{4.7}$$

where $\delta_{tk} = \langle \varepsilon_t, \nu_k \rangle + \sum_{j=p+1}^{\infty} \langle X_{t-S}, \nu_k \rangle \langle \phi(\nu_j), \nu_k \rangle$. Consequently, for any $1 \leq k \leq p$, we have

$$X_{tk} = \sum_{j=1}^{p} \phi_{kj} X_{t-S,j} + \delta_{tk} \tag{4.8}$$

where $\phi_{kj} = \langle \phi(\nu_j), \nu_k \rangle$. Note that the δ_{tk} are not iid. Setting

$$\boldsymbol{X}_t = (X_{t1}, \ldots, X_{tp})', \ \boldsymbol{\delta}_t = (\delta_{t1}, \ldots, \delta_{tp})',$$

$$\boldsymbol{\phi} = (\phi_{11}, \ldots, \phi_{1p}, \phi_{21}, \ldots, \phi_{2p}, \ldots, \phi_{p1}, \ldots, \phi_{pp})',$$

we rewrite (4.8) as

$$\boldsymbol{X}_t = \boldsymbol{Z}_{t-S} \boldsymbol{\phi} + \boldsymbol{\delta}_t, \quad t = 1, 2, \ldots, N,$$

where each $\boldsymbol{Z}_t$ is a $p \times p^2$ matrix

$$\boldsymbol{Z}_t = \begin{bmatrix} \boldsymbol{X}_t' & \mathbf{0}_p' & \ldots & \mathbf{0}_p' \\ \mathbf{0}_p' & \boldsymbol{X}_t' & \ldots & \mathbf{0}_p' \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{0}_p' & \mathbf{0}_p' & \ldots & \boldsymbol{X}_t' \end{bmatrix}, \tag{4.9}$$

with $\mathbf{0}_p = (0, \ldots, 0)'$.

Finally, defining the $Np \times 1$ vectors $\boldsymbol{X}$ and δ and the $Np \times p^2$ matrix $\boldsymbol{Z}$ by

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_N \end{bmatrix}, \quad \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_{1-S} \\ \boldsymbol{Z}_{2-S} \\ \vdots \\ \boldsymbol{Z}_{N-S} \end{bmatrix},$$

we obtain the following linear model

$$\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{\phi} + \delta \tag{4.10}$$

Representation (4.10) leads to the formal least square estimator for $\boldsymbol{\phi}$:

$$\hat{\boldsymbol{\phi}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{X} \tag{4.11}$$

which can be computed if the eigenvalues and eigenfunctions of C_0^X are known. However, in practical problems, since C_0^X is estimated by its empirical estimator, the ν_k must be replaced by its empirical counterpart, called $\hat{\nu}_k$.

4.3 The Kargin-Onatski method

Based on the Kargin and Onatski method, the main idea of predictive factors is to focus on the estimation of those linear functionals of the data that can most reduce the expected square error of prediction. In fact, we would like to find an operator A, approximating ϕ, that minimizes $\mathbb{E}\|X_{t+1} - A(X_{t-S+1})\|^2$. Based on Kargin and Onatski method, we obtain a consistent estimator as follows.

Let $\hat{C}_0$ and $\hat{C}_S$ be the empirical covariance and empirical covariance of lag S, respectively, i.e.,

$$\hat{C}_0(X) = \frac{1}{n}\sum_{t=1}^{n} \langle X_t, x\rangle\, X_t, \qquad \hat{C}_S(X) = \frac{1}{n-S}\sum_{t=S}^{n} \langle X_t, x\rangle\, X_{t-S}.$$

Define $\hat{C}_\alpha$ as $\hat{C}_0 + \alpha I$, where α is a positive real number. Let $\{v_{\alpha,i}, i = 1, 2, \ldots, k_n\}$ be the k_n eigenfunctions of the operator $\hat{C}_\alpha^{-1/2}\hat{C}_S'\hat{C}_S\hat{C}_\alpha^{-1/2}$, corresponding to the first largest eigenvalues $\hat{\tau}_{\alpha,1} > \cdots > \hat{\tau}_{\alpha,k_n}$. Based on these eigenfunctions, we construct the empirical estimator of ϕ, $\hat{\phi}_{\alpha,k_n}$, as follows:

$$\hat{\phi}_{\alpha,k_n} = \sum_{i=1}^{k_n} \hat{C}_{\alpha}^{-1/2} v_{\alpha,i} \otimes \hat{C}_S \hat{C}_{\alpha}^{-1/2} v_{\alpha,i}. \tag{4.12}$$

Besides, it can be demonstrated that that if $\{\alpha_n\}$ is a sequence of positive real numbers such that $\alpha_n \sim n^{-1/6}$, as n goes to infinity, and $\{k_n\}$ is any sequence of positive integers such that $Kn^{-1/4} \leq k_n \leq n$, for some $K > 0$, then $\hat{\phi}_{\alpha,k_n}$ is a consistent estimator of ϕ.

5 Prediction

In this section, we present the h-step functional best linear predictor (FBLP) $\hat{X}_{n+h}$ of X_{n+h} based on $X_1, \ldots, X_n$. For this purpose, we will follow the method of Bosq (2014). In the last section, this prediction method will be compared with the Hyndman-Ullah method and multivariate predictors.

Let us define $\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)'$. It can be demonstrated that the $\mathcal{L}-$closed subspace generated by $\boldsymbol{X}_n$, called G, is the closure of $\{\ell_0 \boldsymbol{X}_n;\ \ell_0 \in \mathcal{L}(H^n, H)\}$, (see Bosq, 2000, Theorem 1.8 for the proof). The h-step FBLP of X_{n+h} is defined as the projection of X_{n+h} on G; i.e., $\hat{X}_{n+h} = P_G X_{n+h}$. The following proposition, which is a modified version of (see Bosq, 2014, Proposition 2.2), presents the necessary and sufficient condition for the existence of FBLP in terms of bounded linear operators and determines its form.

Proposition 5.1. *For $h \in \mathbb{N}$ the following statements are equivalent:*

(i) There exists $\ell_0 \in \mathcal{L}(H^n, H)$ such that $C_{\boldsymbol{X}_n, X_{n+h}} = \ell_0 C_{\boldsymbol{X}_n}$.

(ii) $P_G X_{n+h} = \ell_0 \boldsymbol{X}_n$ for some $\ell_0 \in \mathcal{L}(H^n, H)$.

Based on this proposition, to determine the FBLP of X_{n+h}, it is required to find $\ell_0 \in \mathcal{L}(H^n, H)$ such that $C_{\boldsymbol{X}_n, X_{n+h}} = \ell_0 C_{\boldsymbol{X}_n}$. Let $h = aS + c \in \mathbb{Z}$, $a \geq 0$ and $0 \leq c < S$. Then , for any $\boldsymbol{x} = (x_1, \ldots, x_n) \in H^n$, we have:

$$\begin{aligned}
C_{\boldsymbol{X}_n, X_{n+h}}(\boldsymbol{x}) &= \mathbb{E}\left(\langle \boldsymbol{X}_n, \boldsymbol{x}\rangle_{H^n} X_{n+h}\right) \\
&= \mathbb{E}\left(\langle \boldsymbol{X}_n, \boldsymbol{x}\rangle_{H^n} \left(\phi^{a+1} X_{n-S+c} + \sum_{k=0}^{a} \phi^k \varepsilon_{n-(k-a)S+c}\right)\right) \\
&= \mathbb{E}\left(\langle \boldsymbol{X}_n, \boldsymbol{x}\rangle_{H^n} \phi^{a+1} X_{n-S+c}\right) \\
&= \mathbb{E}\left(\langle \boldsymbol{X}_n, \boldsymbol{x}\rangle_{H^n} \Phi_{n-S+c}^{a+1} \boldsymbol{X}_n\right) \\
&= \Phi_{n-S+c}^{a+1} \mathbb{E}\left(\langle \boldsymbol{X}_n, \boldsymbol{x}\rangle_{H^n} \boldsymbol{X}_n\right)
\end{aligned}$$

$$= \Phi^{a+1}_{n-S+c} C_{\boldsymbol{X}_n}(\boldsymbol{x}),$$

where Φ^i_j is in $\mathcal{L}(H^n, H)$ and is defined as an $n \times 1$ vector of zeros with ϕ^i in the j^{th} position. Consequently, $\hat{X}_{n+h} = P_G X_{n+h} = \Phi^{a+1}_{n-S+c} \boldsymbol{X}_n = \phi^{a+1} X_{n-S+c}$.

Note 5.1. *Based on The Kargin-Onatski estimation method, the 1-step ahead predictor of X_{n+1} is:*

$$\hat{X}_{n+1} = \sum_{i=1}^{k_n} < X_{n-S+1}, \hat{z}_{\alpha,i} > \hat{C}_S(\hat{z}_{\alpha,i}), \tag{5.1}$$

where

$$\hat{z}_{\alpha,i} = \sum_{j=1}^{q} \hat{\tau}_j^{-1/2} \left\langle v_{\alpha,i}, \hat{v}_j \right\rangle \hat{v}_j + \alpha v_{\alpha,i}, \tag{5.2}$$

The method depends on a selection of q and k_n. We selected q by the cumulative variance method and set $k_n = q$.

6 The $SFAR(P)_S$ model

Although having interesting properties, SFAR(1)$_S$ processes have some limitations in practical problems. Therefore, in this section, we extend SFAR processes to include seasonal functional autoregressive process of order P.

Definition 6.1. *A sequence $X = \{X_n; n \in \mathbb{Z}\}$ of random functions variables is said to be a pure seasonal functional autoregressive process of order P with seasonality S ($SFAR(P)_S$) associated with $(\varepsilon, \phi_1, \ldots, \phi_P)$ if*

$$X_n - \mu = \phi_1 (X_{n-S} - \mu) + \phi_2 (X_{n-2S} - \mu) + \cdots + \phi_P (X_{n-PS} - \mu) + \varepsilon_n, \tag{6.1}$$

where $\varepsilon = \{\varepsilon_n, n \in \mathbb{Z}\}$ is a functional white noise and $\phi_1, \ldots, \phi_P \in \mathcal{L}(H)$, with $\phi_P \neq 0$.

6.1 Basic properties

For $n \in \mathbb{Z}$, let $Y_n = (X_n, X_{n-S}, \ldots, X_{n-PS+S})'$, and $\varepsilon'_n = (\varepsilon_n, 0, \ldots, 0)'$, where 0 appears $P-1$ times. Let us define the operator $\boldsymbol{\phi}$ on H^P as

$$\boldsymbol{\phi} = \begin{bmatrix} \phi_1 & \phi_2 & \ldots & \phi_P \\ I & 0 & \ldots & 0 \\ 0 & I & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}, \tag{6.2}$$

where I and 0 denote the identity operator and zero operator on H, respectively.

Lemma 6.1. *If X is a SFAR$(P)_S$ associated with associated with $(\varepsilon, \phi_1, \ldots, \phi_P)$, then Y is a SFAR$(1)_S$ with values in the product Hilbert space H^P associated with $(\varepsilon', \boldsymbol{\phi})$.*

Proof. It can easily be shown that ε' is a H^P white noise process and $\boldsymbol{\phi}$ is an $\mathcal{L}\left(H^P\right)$ operator. Moreover, it is clear that $Y_n = \boldsymbol{\phi} Y_{n-S} + \varepsilon'$, which demonstrates that Y is a SFAR$(1)_S$ process with values in the product Hilbert space H^P. □

Theorem 6.1. *Let X_n be a SFAR$(P)_S$ zero-mean process associated with $(\varepsilon, \phi_1, \phi_2, \ldots, \phi_P)$. Suppose that there exist $\nu \in H$ and $\alpha_1, \ldots, \alpha_P \in \mathbb{R}$, $\alpha_P \neq 0$, such that $\phi_j(\nu) = \alpha_j \nu_j$, $j = 1, \ldots, P$ and $\mathbb{E}\langle \varepsilon_0, \nu \rangle^2 > 0$. Then, $(\langle X_n, \nu \rangle,\ n \in \mathbb{Z})$ is a SAR(P) process, i.e.,*

$$\langle X_n, \nu \rangle = \sum_{j=1}^{P} \alpha_j \langle X_{n-jS}, \nu \rangle + \langle \varepsilon_n, \nu \rangle, \qquad n \in \mathbb{Z}. \tag{6.3}$$

Theorem 6.2. *If X is a standard SFAR$(P)_S$ process, then*

$$C_h^X = \sum_{j=1}^{P} \phi_j C_{h-jS}^X, \qquad h = 1, 2, \ldots, \tag{6.4}$$

$$C_0^X = \sum_{j=1}^{P} \phi_j C_{jS}^X + C^\varepsilon, \tag{6.5}$$

where C^ε is the covariance operator of the innovation process ε.

7 Simulation results

Let $\{X_t\}$ follows a SFAR$(1)_S$ model, i.e.,

$$X_t(\tau) = \phi X_{t-S}(\tau) + \varepsilon_t(\tau), \quad t = 1, \ldots, n, \tag{7.1}$$

where ϕ is an integral operator with *parabolic* kernel

$$k_\phi(\tau, u) = \gamma_0 \left(2 - (2\tau - 1)^2 - (2u - 1)^2\right).$$

The value of γ_0 is chosen such that $\|\phi\|_{\mathcal{S}}^2 = \int_0^1 \int_0^1 |k_\phi(\tau, u)|^2 \, dsdu = 0.9$. Moreover, the white noise terms $\varepsilon_t(\tau)$ are considered as independent standard Brownian motion process on $[0, 1]$ with variance 0.05. Then $B = 1000$ trajectories are simulated from the Model (7.1) and the model parameter is estimated using the method of moments (MME), Unconditional Least Square (ULSE) and Kargin-Onatski (KOE) estimators. As a measure to check the adequacy of the fit of the model, the root mean-squared error (RMSE), defined as

$$\text{RMSE} = \sqrt{\frac{1}{B} \sum_{i=1}^{B} \|\hat{\phi}_i - \phi\|_{\mathcal{S}}^2}, \tag{7.2}$$

is applied.

The MSE results are reported in Table 1. To better understand these methods, a trajectory path of length $n = 200$ from the model (7.1) is generated by considering $\|\phi\|_{\mathcal{S}} = 0.9$ and $k_n = 1$. Then three estimation methods are compared using the generated path and the results are shown in Figure 1. Moreover, the figures related to MME are shown in Figure 2 for $k_n = 1, 2, 3$. As can be seen in these figures, the method of moments and unconditional least squares have a better performance than the Kargin-Onatski method. Besides, increasing the number of k_n in the estimation methods, decreases the accuracy of estimation.

		$\|\phi\|_{\mathcal{S}} = 0.1$			$\|\phi\|_{\mathcal{S}} = 0.5$			$\|\phi\|_{\mathcal{S}} = 0.9$		
n	k_n	MME	ULSE	KOE	MME	ULSE	KOE	MME	ULSE	KOE
50	1	0.1750	0.1645	0.0951	0.2403	0.2838	0.3716	0.1986	0.2323	0.5096
	2	0.5484	0.5189	0.0959	0.4931	0.7000	0.3720	0.4387	1.0381	0.5099
	3	1.0239	0.9657	0.0961	0.9988	1.1478	0.3721	1.0435	1.7282	0.5099
	4	1.5573	1.4934	0.0962	1.5340	1.6513	0.3721	1.6382	2.4725	0.5099
100	1	0.1222	0.1183	0.0861	0.2050	0.2579	0.3539	0.1387	0.1709	0.4134
	2	0.3662	0.3598	0.0866	0.3325	0.6087	0.3541	0.2743	0.9728	0.4136
	3	0.6830	0.6798	0.0868	0.6661	0.8723	0.3541	0.6694	1.3925	0.4136
	4	1.0645	1.0243	0.0868	1.0245	1.1973	0.3542	1.0193	1.9377	0.4136
150	1	0.1033	0.1027	0.0825	0.1946	0.2505	0.3460	0.1205	0.1517	0.3735
	2	0.2903	0.2900	0.0830	0.2666	0.5704	0.3462	0.2149	0.9478	0.3737
	3	0.5533	0.5449	0.0831	0.5387	0.7601	0.3462	0.5272	1.2493	0.3736
	4	0.8560	0.8237	0.0831	0.8256	1.0040	0.3462	0.8106	1.6683	0.3736
200	1	0.0917	0.0935	0.0798	0.1879	0.2457	0.3393	0.1114	0.1419	0.3496
	2	0.2490	0.2610	0.0803	0.2285	0.5568	0.3394	0.1818	0.9411	0.3497
	3	0.4790	0.4745	0.0804	0.4684	0.7047	0.3394	0.4542	1.1896	0.3497
	4	0.7438	0.7127	0.0804	0.7199	0.9042	0.3394	0.7040	1.5134	0.3497

Table 1: *The NMSE of the parameter estimators.*

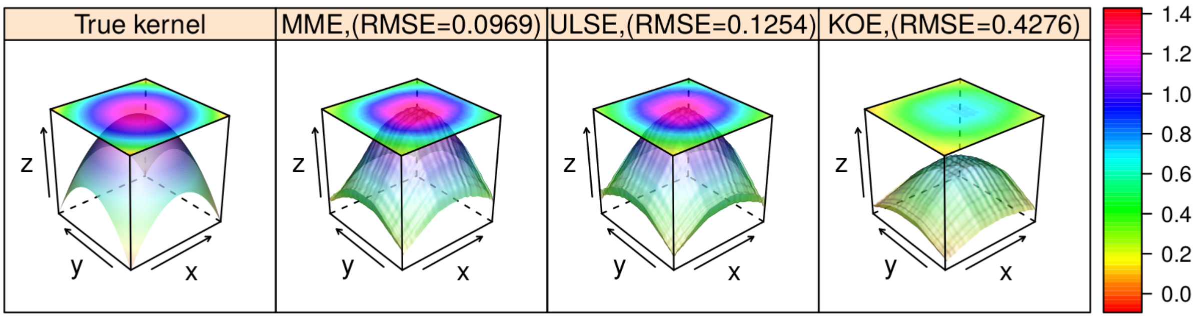

Figure 1: *Comparison of all methods of estimations when $n = 200$, $\|\phi\|_{\mathcal{S}} = 0.9$, and $k_n = 1$.*

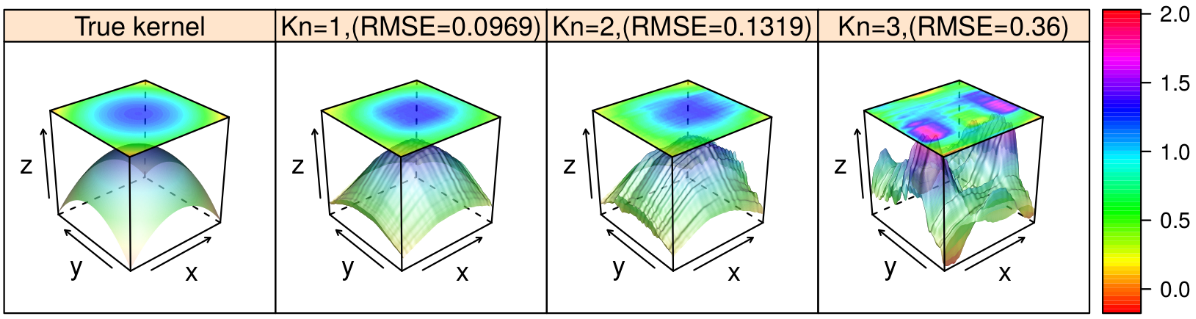

Figure 2: *Method of moment estimation of the model parameter kernel by considering $n = 200$, $\|\phi\|_{\mathcal{S}} = 0.9$, and $k_n = 1, 2, 3$.*

8 Application to pedestrian traffic

In this section, we study pedestrian behavior in Melbourne, Australia. These data are counts of pedestrians captured at hourly intervals by 43 sensors scattered around the city. The dataset can shed light on people's daily rhythms, and assist the city administration and local businesses with event planning and operational management. Our dataset consists of measurements at Flagstaff Street Station from 00:00 1 Jan 2016 to 23:00 31 Dec 2016.[1] In Figure 3, we show the pedestrian behavior by day of the week, which are converted to functional objects using 11 Fourier basis functions.

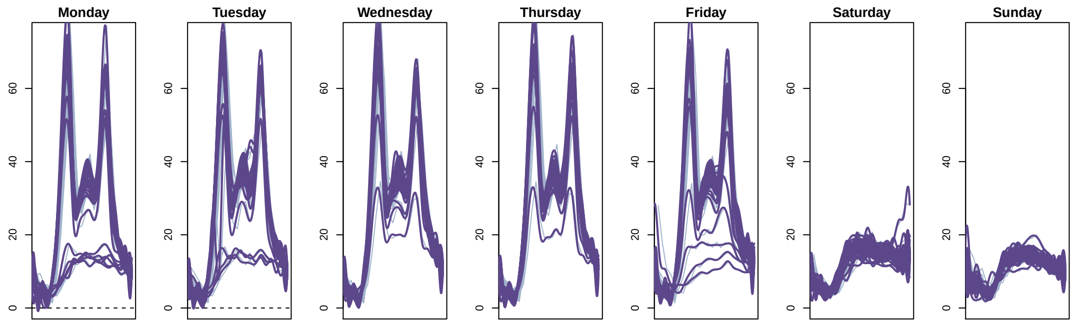

Figure 3: *Square root of functional pedestrian hourly counts at Flagstaff Street Station, by day of the week.*

Although Figure 3 shows the existence of a weekly pattern in the data, the different pattern in weekends can put stationarity of data in doubt. Therefore, we do not consider weekends in data alnaysis and SFAR$(1)_7$ model is applied to the rest of the data. The number of bases is chosen using a generalized cross-validation criteria. Figure 4 shows the estimation of the autocorrelation kernel using different estimation methods, as described in Section 4.

A rolling forecast origin for one day ahead throughout the data set was computed. The Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) of the forecasts are tabulated in Table 2, comparing the results when using MME and ULSE estimation methods. The table also include the seasonal naive (SN) forecast for comparison. As can be seen in this table, the best predictions come with small values of k_n (similar to the results in the last section) and using MME. Using this method of prediction, the one-day-ahead forecasts are shown in Figure 5, along with the functional pedestrian data.

[1]The data are available using the rwalkr package (Wang, 2019) in R.

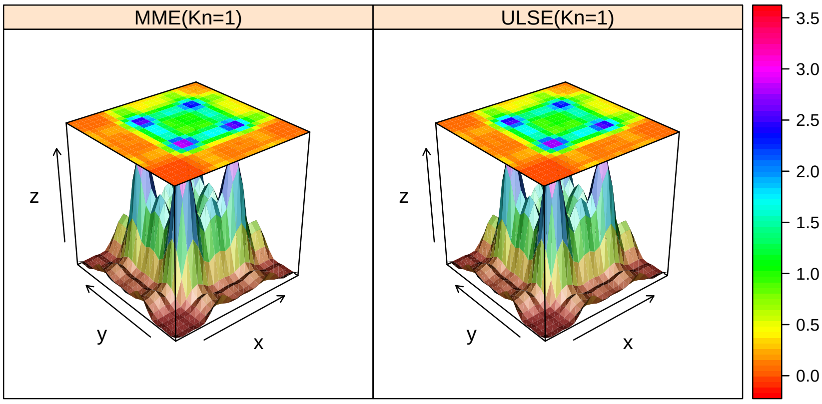

Figure 4: *The estimated kernel of the autocorrelation operator using MME and ULSE methods.*

	MAE		RMSE	
k_n	MME	ULSE	MME	ULSE
1	3.04	3.025	6.08	6.073
2	135.19	24.55	138.09	24.70
3	139.46	27.65	142.09	27.88
4	155.52	30.07	157.62	30.41
5	165.34	32.60	168.06	33.23
6	180.86	33.20	186.92	34.01
SN	3.67		6.46	

Table 2: *The MAE and RMSE of the 1-step predictor when the autocorrelation operator is estimated by MME and ULSE methods.*

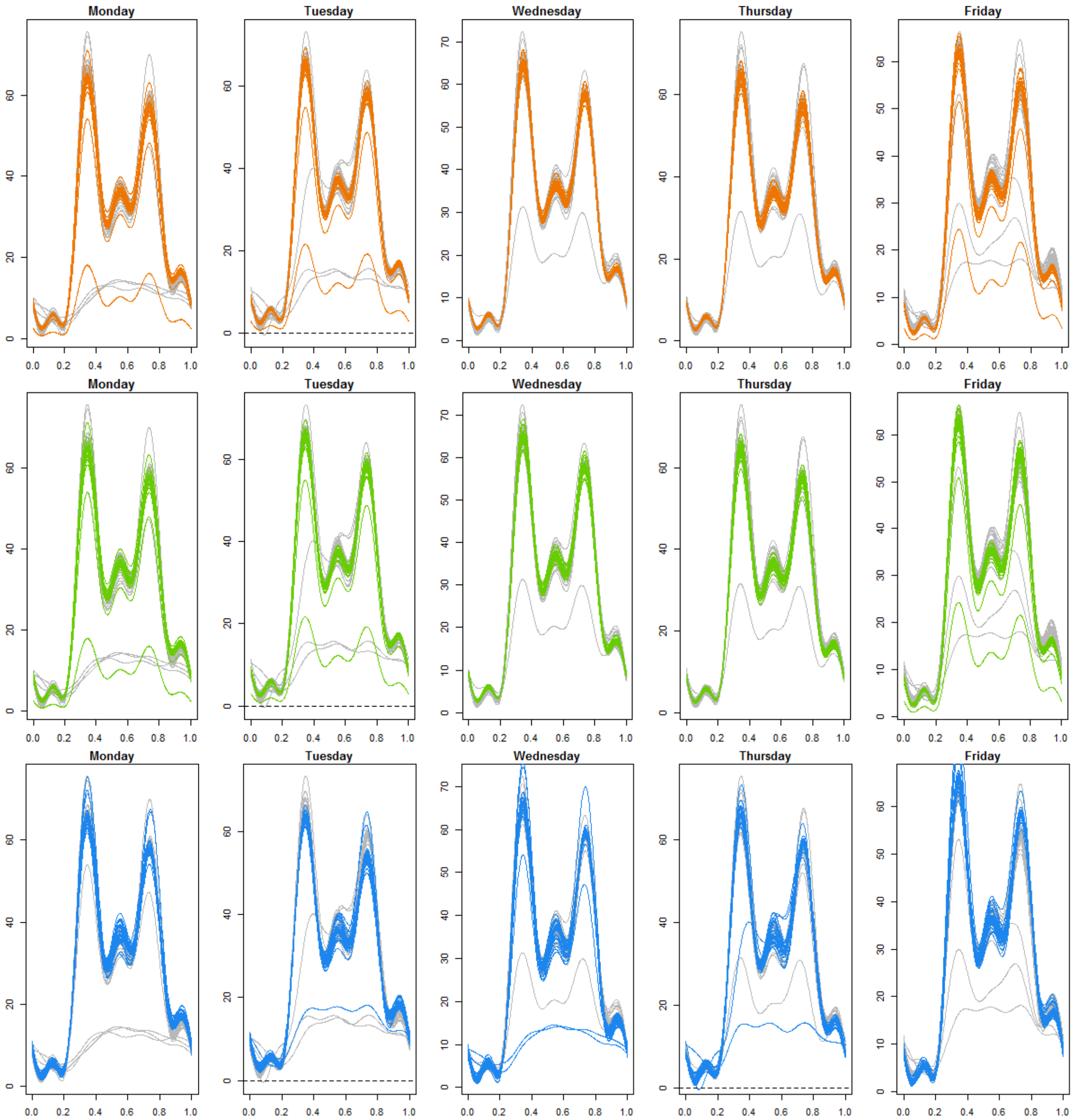

Figure 5: *Rolling forecast origin of all dataset for one-day-ahead using MME method with $k_n = 1$ ("orange" curves), ULSE method with $k_n = 1$ ("green" curves) and Seasonal naive method ("blue curves") . The "grey" curves indicate original data.*

9 Conclusion

In this paper, we have focused on seasonal functional autoregressive time series. We have presented conditions for the existence of a unique stationary and causal solution. Furthermore, we have derived some basic properties and limiting behaviour. In FSAR$(1)_S$, we proposed three estimation methods, namely methods of moment, unconditional least square and Kargin-Onatski. Furthermore, for arbitrary $h \in \mathbb{N}$, we have investigated the h-step predictor based on Kargin-Onatski (2008) and Bosq (2014). The performance of the estimation methods are compared using a simulation study, which demonstrates that the MME has the best performance among mentioned estimation methods. Finally, a real data set is analyzed using the SFAR$(1)_S$ model.

References

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

Berkes, I, L Horváth & G Rice (2013). Weak invariance principles for sums of dependent random functions. *Stochastic Processes and their Applications* **123**(2), 385–403.

Besse, PC, H Cardot & DB Stephenson (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics* **27**(4), 673–687.

Bosq, D (2000). *Linear processes in function spaces: theory and applications*. Vol. 149. Lecture Notes in Statistics. Springer Science & Business Media.

Bosq, D (2014). Computing the best linear predictor in a Hilbert space. Applications to general ARMAH processes. *Journal of Multivariate Analysis* **124**, 436–450.

Box, GE, GM Jenkins, GC Reinsel & GM Ljung (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Chatfield, C & D Prothero (1973). Box-Jenkins seasonal forecasting: problems in a case-study. *Journal of the Royal Statistical Society, Series A*, 295–336.

Conway, JB (2000). *A course in operator theory*. American Mathematical Society.

Ghysels, E & DR Osborn (2001). *The econometric analysis of seasonal time series*. Cambridge University Press.

Grenander, U (1950). Stochastic processes and statistical inference. *Arkiv för matematik* **1**(3), 195–277.

Harrison, PJ (1965). Short-term sales forecasting. *Applied Statistics*, 102–139.

He, F (2018). “Statistical Modeling of CO2 Flux Data”. PhD thesis. The University of Western Ontario.

Hörmann, S, L Horváth & R Reeder (2013). A functional version of the ARCH model. *Econometric Theory* **29**(2), 267–288.

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

Hörmann, S, P Kokoszka, G Nisol, et al. (2018). Testing for periodicity in functional time series. *The Annals of Statistics* **46**(6A), 2960–2984.

Horváth, L, M Hušková & P Kokoszka (2010). Testing the stability of the functional autoregressive process. *Journal of Multivariate Analysis* **101**(2), 352–367.

Horváth, L & P Kokoszka (2012). *Inference for functional data with applications*. Vol. 200. Springer Science & Business Media.

Horváth, L, P Kokoszka & R Reeder (2013). Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**(1), 103–122.

Hsing, T & R Eubank (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons.

Hyndman, RJ & S Ullah (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis* **51**(10), 4942–4956.

Kargin, V & A Onatski (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis* **99**(10), 2508–2526.

Klepsch, J, C Klüppelberg & T Wei (2017). Prediction of functional ARMA processes with an application to traffic data. *Econometrics and Statistics* **1**, 128–149.

Panaretos, VM, S Tavakoli, et al. (2013). Fourier analysis of stationary time series in function space. *The Annals of Statistics* **41**(2), 568–603.

Ramsay, J (1982). When the data are functions. *Psychometrika* **47**(4), 379–396.

Rao, CR (1958). Some statistical methods for comparison of growth curves. *Biometrics* **14**(1), 1–17.

Shang, HL (2013). Functional time series approach for forecasting very short-term electricity demand. *Journal of Applied Statistics* **40**(1), 152–168.

Wang, E (2019). *rwalkr: API to Melbourne Pedestrian Data*. R package version 0.4.0. http://pkg.earo.me/rwalkr.