Functional time series model identification and diagnosis by means of auto- and partial autocorrelation analysis

Guillermo Mestre[a,*], José Portela[b], Gregory Rice[c], Antonio Muñoz San Roque[a], Estrella Alonso[d]

[a] *Universidad Pontificia Comillas. Escuela Técnica Superior de Ingeniería ICAI. Instituto de Investigación Tecnológica. Madrid, Spain*
[b] *Universidad Pontificia Comillas. Facultad de Ciencias Económicas y Empresariales ICADE. Madrid, Spain*
[c] *Department of Statistics and Actuarial Science, University of Waterloo, Canada*
[d] *Universidad Pontificia Comillas. Escuela Técnica Superior de Ingeniería ICAI. Departamento de Matemática Aplicada. Madrid, Spain*

This is a published, author-produced PDF version of an article appearing in Computational Statistics & Data Analysis following peer review. The published version "G. Mestre, J. Portela, G. Rice, A.M. San Roque, E. Alonso (2021), Functional time series model identification and diagnosis by means of auto- and partial autocorrelation analysis, *Computational Statistics and Data Analysis,* 155, 107108". Available online at: https://doi.org/10.1016/j.csda.2020.107108

[*] Correspondence to: Santa Cruz de Marcenado, 26 - 28015 Madrid, Spain. Tel. (+34) 91 542 2800 ext. 2749.
Email address: guillermo.mestre@comillas.edu (Guillermo Mestre)

Functional time series model identification and diagnosis by means of auto- and partial autocorrelation analysis

Guillermo Mestre[a,*], José Portela[b], Gregory Rice[c], Antonio Muñoz San Roque[a], Estrella Alonso[d]

[a] *Universidad Pontificia Comillas. Escuela Técnica Superior de Ingeniería ICAI. Instituto de Investigación Tecnológica. Madrid, Spain*
[b] *Universidad Pontificia Comillas. Facultad de Ciencias Económicas y Empresariales ICADE. Madrid, Spain*
[c] *Department of Statistics and Actuarial Science, University of Waterloo, Canada*
[d] *Universidad Pontificia Comillas. Escuela Técnica Superior de Ingeniería ICAI. Departamento de Matemática Aplicada. Madrid, Spain*

Abstract

Quantifying the serial correlation across time lags is a crucial step in the identification and diagnosis of a time series model. Simple and partial autocorrelation functions of the time series are the most widely used tools for this purpose with scalar time series. Nevertheless, there is a lack of an established method for the identification of functional time series (FTS) models. Functional versions of the autocorrelation and partial autocorrelation functions for FTS based on the L^2 norm of the lagged autocovariance operators of the series are proposed. Diagnostic plots of these functions coupled with prediction bounds derived from large sample results for the autocorrelation and partial autocorrelation functions estimated from a strong functional white noise series are proposed as fast and efficient tools for selecting the order and assessing the adequacy of functional SARMAX models. These methods are studied in numerical simulations with both white noise and serially correlated functional processes, which show that the structure of the processes can be well identified using the proposed techniques. The applicability of the method is illustrated with two real-world datasets: Eurodollar futures contracts and electricity price profiles.

Keywords: Autocorrelation, Partial autocorrelation, Functional time series, Model diagnosis
2010 MSC: 62M10

*Correspondence to: Santa Cruz de Marcenado, 26 - 28015 Madrid, Spain. Tel. (+34) 91 542 2800 ext. 2749.
Email address: guillermo.mestre@comillas.edu (Guillermo Mestre)

1. Introduction

Due to recent technological advances in science and industry, the collection and analysis of high-dimensional data has become an important topic in the field of applied statistics. Many of these observations arise from continuous processes observed, for example, over time, space or frequency domains. These observations can be naturally expressed as functions, and thus the name functional data is commonly given to such observations. The need for statistical methodologies to explore and analyze the structure of these high-dimensional data has led to the development of the Functional Data Analysis (FDA) framework. The monographs Ramsay and Silverman (2005), Ramsay et al. (2009) and Kokoszka and Reimherr (2017) provide excellent introductions to the key topics of functional data and its applications.

If these functional data are collected sequentially over time, it is natural to expect a time dependence between functional observations. This motivates the consideration of functional time series (FTS); see e.g. Bosq (2000); Mas and Pumo (2010). Due to its wide range of real world applications, there has been growing research interest in the analysis and identification of such functional processes, as presented in Horváth and Kokoszka (2012) where several tests are introduced to measure the serial dependence structure and other properties of FTS.

To present the main ideas, let $\{Y_t(u)\,;\, t = 1, \ldots, T,\ u \in [0,1]\}$ denote an observed stretch of length T of a functional time series. Here we assume that each observation $Y_t(u)$ is a stochastic process indexed by $[0,1]$ whose sample paths are in $L^2([0,1])$, which is the space of square integrable functions defined on the unit interval. Since typically functional data are only observed at some discrete collection of points, we are assuming that the elements of the FTS under consideration have been obtained after a pre-processing step, such as linear interpolation or basis smoothing; see e.g. Ramsay and Silverman (2005). We could consider instead FTS taking values in more general function spaces, but proceed with $L^2([0,1])$ in this presentation given the data applications we present below.

As with the majority of scalar and vector-valued time series models, most FTS models can be expressed as

$$Y_t = f(Y_{t-1}, \ldots, Y_{t-p}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}) + \varepsilon_t, \tag{1}$$

where $\{\varepsilon_t\,;\, t \in \mathbb{Z}\}$ denotes a sequence of independent and identically distributed random functional processes with zero mean, and f is an operator of the p past values of the time series and the q

past values of the innovation processes. Functional models such as the functional autoregressive model (Bosq, 2000; Aue et al., 2015; Martínez-Hernández et al., 2019), the functional ARHX model (Damon and Guillas, 2002, 2005), the functional moving average model (Turbillon et al., 2007), the functional ARMA model (Klepsch et al., 2017),the SARMAHX model (Portela et al., 2018; Mestre et al., 2020) and the functional ARCH model (Hörmann et al., 2013) all follow the general equation (1).

Identifying relevant features of the time series such as serial correlation, trend or seasonality is of utmost importance when identifying a time series model. Several methods have been proposed for revealing these attributes. Graphical tools can be found in Hyndman and Shang (2010), where the rainbow plot was proposed for visualizing large time indexed functional data sets and performing outlier detection. Curves that exhibit an odd shape or magnitude can be identified with this method. In addition, it provides some information about the trend of the series. Other tools can be found in Canale and Vantini (2016), where the authors use visual representations of the autocorrelation functions of the curves to obtain insights about the dependence across functions observed at different times.

For example, Figures 1 and 2 illustrate these tools for a FTS composed of daily electricity price profiles from the Spanish electricity market in 2014 (Portela et al., 2018). In particular, Figure 1 shows the rainbow plot for the price profiles time series. Increases or changes in trend of the daily price profiles are clearly visible in the summer months, as well as in January (red curves). It is however difficult to decipher from this plot the nature of the serial dependence between the curves. Figure 2 illustrates the autocovariance surfaces, or kernels for different time lags; see equation (5) below for a formal definition. It can be seen by analyzing these carefully that the electricity prices for hours 12 to 18 are correlated with past curves, and that hours 19 to 22 are less influenced by the former price curves. These graphical tools provide some insights on the trend and outliers of the FTS, but in terms of describing the nature of serial dependence between the curves are lacking in many respects, especcially in that they are difficult to interpret at a glance.

Moreover, once a model of the form (1) has been fitted to a given FTS, the plausibility of the main assumptions made by the model should be investigated. In particular, the residuals of the model should behave as independent, identically distributed curves. Several tests have been proposed in order to check the validity of the serial independence assumption for functional data.

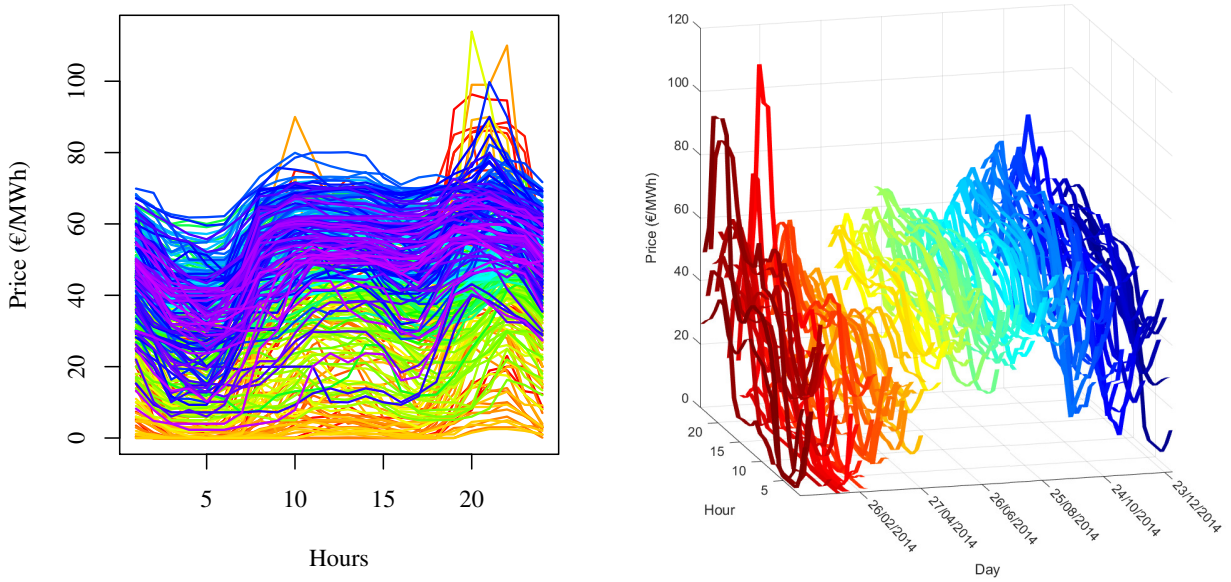

Figure 1: Left: Rainbow plot of the FTS of daily electricity price profiles from the Spanish electricity market for the year 2014. Color illustrates time, with the oldest curves in red and the most recent data in violet. Right: Rainbow plot with time in a third axis. (The reader is referred to the web version of this article for interpretation of the references to color in this figure).

A Functional Principal Component (FPCA)-based Portmanteau test is proposed in Gabrys and Kokoszka (2007). In Gabrys et al. (2010), the authors project the residuals of the functional linear model on an "optimal" finite dimensional subspace to test for model goodness-of-fit. The white noise test proposed in Horváth et al. (2013) is based on the sum of the L^2 norms of the autocovariance operators of the FTS, rather than on summaries of the FPCA projections. Kokoszka et al. (2017) also develop portmanteau tests and inferential procedures for the autocovariance operators of a FTS that are valid in the presence of conditional heteroscedasticity. In addition, frequency-domain tests have been proposed recently in Zhang (2016), Bagchi et al. (2018) and Characiejus and Rice (2020), where the authors use the distance between the spectral density operator of the series and its best L^2-approximation by a spectral density operator corresponding to a white noise process as a measure of the deviation of the time series from a white noise process.

The host of aforementioned tests would all benefit from the subsequent use of methods that provide insightful information about the underlying dependence structure of the FTS. In the event

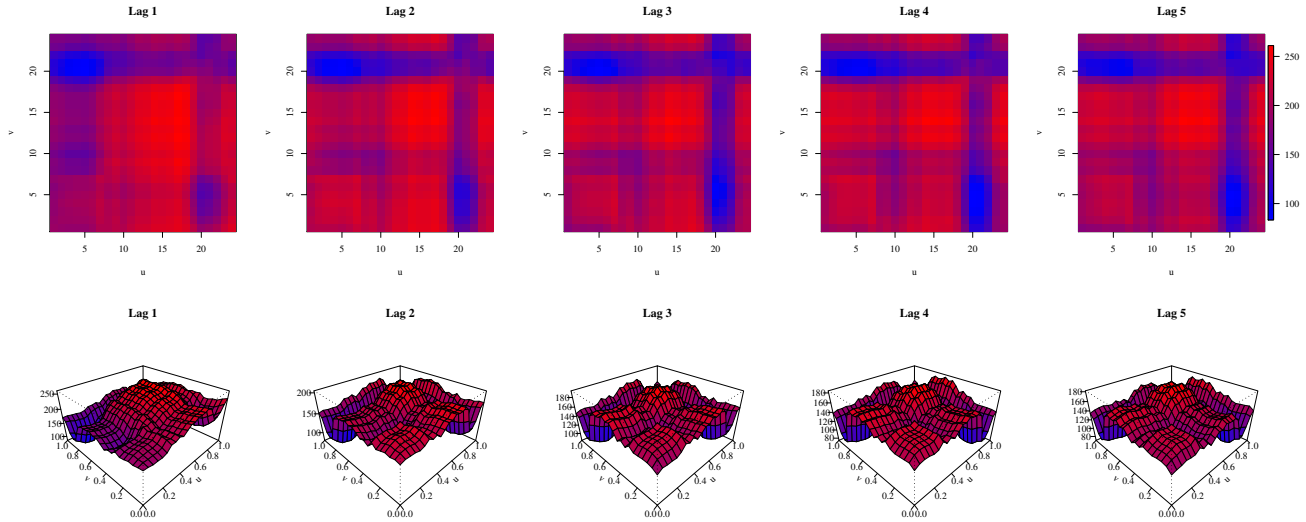

Figure 2: First 5 lagged autocovariance functions for the FTS of daily electricity price profiles (top) and a three-dimensional representation of the surfaces (bottom). These functions provide detailed, though perhaps difficult to interpret, information on covariance structure between curves Y_t and Y_{t-h}.

that these tests suggest that the model errors do not plausibly form a white noise, one often hopes to identify specific lags at which substantial correlation exists in order to fit an improved model to the data. Classical identification methods for scalar time series, such as the Box-Jenkins methodology (Box et al., 2008), use the sample autocorrelation and partial autocorrelation functions of the time series in order to simultaneously test the white noise assumption and to identify the underlying correlation structure of the time series. Identifying the rate of decay of the autocorrelation and specific lags with strong autocorrelation can be used to select the appropriate autoregressive and moving average orders in an ARIMA time series model.

While these kinds of identification methods have been well addressed in the literature of scalar time series, the same cannot be said for FTS. An analog of the autocorrelation function for FTS was proposed in Kokoszka et al. (2017) for the purpose of quantifying conditional heteroscedasticity in functional data, but its use in FTS model selection has not been explored. Further, to our knowledge, the notion of partial autocorrelation with functional data has not been explored to date, nor how such summary information could be used in FTS model building.

This paper proposes methods based on the L^2 norms of lagged covariance operators in order to

quantify the autocorrelation structure of FTS. A functional version of the autocorrelation function is studied in the context of time series model identification. In addition, a generalization of the partial autocorrelation function is proposed, which allows the practitioner to identify functional autoregressive and moving average processes of higher order. We show how these methods can be used in order to identify autocorrelation and seasonality in FTS, which are useful in model selection within the functional SARMAHX family of models introduced in Portela et al. (2018) and Mestre et al. (2020). These models generalize the aforementioned functional linear models, allowing the inclusion of both autoregresive and moving average effects as well as incorporating the seasonal behavior present in the data. As such, identifying the serial correlation structure of the data is of utmost importance to fitting the SARMAHX model.

The paper is organized as follows: Section 2 presents the main contributions and the practical implementation of the proposed identification procedure. Section 3 contains the results of a Monte-Carlo simulation study of the proposed methodology. Demonstrations and applications of these model identification procedures are presented in 4, where two real-world datasets are explored. Section 5 gives some concluding remarks, summarizes the results of the methodology developed in the previous sections, and points to some directions for future research. Appendix A contains details of the main technical results.

2. Functional Autocorrelation Measures

We begin by introducing some notation that will be used throughout this paper and state the main assumptions of the proposed methodology. Throughout this paper, $L^2([0,1]^d)$ denotes the Hilbert space of real valued square integrable functions defined on $[0,1]^d$ with the usual inner product $\langle\,.\,,.\,\rangle$ and the norm $\|.\|$ it generates. The dimension is made clear based on the domain of the input function.

2.1. Autocorrelation function for FTS

Here an FTS $\{Y_t\,;\,t=1,\ldots,T\}$ is assumed to be a realization of length T of a given functional stochastic process $\{Y_t\,;\,t\in\mathbb{Z}\}$, where each random variable Y_t is a square integrable function $\{Y_t(v)\,;\,v\in[0,1]\}$. For the rest of the section, it will be assumed that all the FTS are stationary. The main interest of this work is to develop a statistical tool that can be used to identify the underlying serial dependence structure of a given time series. In order to achieve this objective,

the first step is to test the white noise assumption on a given sequence of functional observations. The definition of functional white noise is given below:

Definition 1. The functional stochastic process $\{\varepsilon_t(v)\,;\, t \in \mathbb{Z}; v \in [0,1]\}$ is a (strong) functional white noise process if the following assumptions are satisfied:

- **A1**: The functional random variables ε_t are independent and identically distributed.
- **A2**: $\mathrm{E}\left[\varepsilon_i(v)\right] = 0$ for all $i = 1, \ldots, T$ and $v \in [0,1]$.

The first assumption implies that such a white noise series does not exhibit any degree of serial dependence, differentiating such a process from others that exhibit serial dependence, such as the functional linear process (Bosq, 2000).

Throughout this paper, it will be assumed that all the functional random variables have finite fourth order moments, which is to say that $E\|Y_t\|^4 < \infty$. This implies that the (lag zero) autocovariance kernel $C_0(u,v)$, defined as

$$C_0(u,v) = \mathrm{E}\left[\left(Y_1(u) - \mathrm{E}Y_1(u)\right)\left(Y_1(v) - \mathrm{E}Y_1(v)\right)\right], \tag{2}$$

is square integrable. Hence, there are real positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$ and orthonormal eigenfunctions $\phi_1, \phi_2, \ldots$ satisfying

$$\lambda_i \phi_i(u) = \int C_0(u,s)\phi_i(s)ds, \qquad i = 1, 2, \ldots \tag{3}$$

The dependence structure of a given stationary FTS can be analyzed via its lagged autocovariance kernels, defined as

$$C_h(u,v) = \mathrm{cov}\left(Y_1(u), Y_{1+h}(v)\right), \tag{4}$$

where $h = 1, \ldots, T-1$ is a given lag parameter. Indeed fourth order moments are more than what is required here to make the above quantities well defined, two would suffice, although we require fourth order moments in order to establish large sample Gaussian approximations for autocovariance operator estimates presented below.

Given $Y_1(v), \ldots, Y_T(v)$ a realization of the FTS, the autocovariance kernel $C_h(u,v)$ can be estimated via the empirical lagged autocovariance kernels:

$$\hat{C}_h(u,v) = \frac{1}{T}\sum_{i=1}^{T-h}(Y_i(u) - \overline{Y}_T(u))(Y_{i+h}(v) - \overline{Y}_T(v)), \tag{5}$$

where

$$\overline{Y}_T(u) = \frac{1}{T}\sum_{i=1}^{T} Y_i(t) \tag{6}$$

denotes the sample mean function. Using right integration as in (3), these kernels may be used to define autocovariance and empirical autocovariance operators, and so we use the terms kernels and operators interchangeably in this setting.

Due to the fact that a functional white noise process exhibits no serial correlation between its terms, the norm of its lagged autocovariance operators should be close to zero for each positive lag h. Horváth et al. (2013) for example thus propose to use the sums of the L^2-norms of the lagged autocovariance operators up to a user specified maximum lag H in order to obtain a white noise test for functional data.

However, such a test does not provide the practitioner with specific information about the lagged dependence structure of the FTS. This information may be summarized simply by considering the magnitudes and plots of the empirical covariance operators as a function of the lag. Evaluating whether the magnitudes $\|\hat{C}_h\|$ are consistent with the series following a white noise process, and further identifying specific lags where strong correlation is present, is facilitated by comparing $\|\hat{C}_h\|$ to an estimate of its distribution assuming the data are indeed a white noise. The following result describes the large sample behaviour of $\|\hat{C}_h\|$ under assuming the data follow a white noise; see Kokoszka et al. (2017) for a proof.

Theorem 1. If the FTS $\{Y_t \ : \ t \in \mathbb{Z}\}$ is independent and identically distributed with $E\|Y_t\|^4 < \infty$, then

$$\hat{Q}_{T,h} = \iint T\hat{C}_h^2(u,v)dudv \xrightarrow{D} Q = \sum_{j=1}^{\infty}\sum_{l=1}^{\infty} \lambda_j\lambda_l\chi^2_{j,l}(1), \qquad h = 1,2,\dots \tag{7}$$

as $T \to \infty$. The $\lambda_i's$ are the eigenvalues of the covariance operator of Y_t and $\{\chi^2_{j,l}(1) \ , \ j,l \in \mathbb{N}\}$ are independent random variables following a chi-squared distribution $\chi^2(1)$.

Following Kokoszka et al. (2017), the functional autocorrelation coefficient at lag h can be defined as

$$\rho_h = \frac{\|C_h\|}{\int C_0(u,u)du} \ , \ \|C_h\|^2 = \iint C_h^2(u,v)dudv. \tag{8}$$

It follows from the Cauchy-Schwarz inequality applied pointwise to $C_h(u,v) = \operatorname{cov}\left(Y_1(u), Y_{1+h}(v)\right)$ that $0 \leq \rho_h \leq 1$, and gives a scale free measure of the serial correlation in the FTS at lag h. One

could also define ρ_h as $\|C_h\|/\|C_0\|$, which would also give a scale free measure of the autocorrelation structure of a given FTS. Both definitions are equally useful in identifying the lagged dependence structure encoded by $\|C_h\|$. Henceforth, in this paper the definition in (8) is used, as this definition is more easily adapted to a functional version of a partial autocorrelation function, which we pursue below.

ρ_h admits the plug in estimate

$$\hat{\rho}_h = \frac{\|\hat{C}_h\|}{\int \hat{C}_0(u,u)du} = \frac{\sqrt{Q_{T,h}}}{\sqrt{T}\int \hat{C}_0(u,u)du}. \tag{9}$$

Based on Theorem 1, a procedure to produce a prediction bound for $\hat{\rho}_h$ under the assumption that the data are a white noise can be summarized as:

- **Step 1**: Estimate $\hat{\lambda}_1, \ldots, \hat{\lambda}_d$ from

$$\hat{C}_0(u,v) = \frac{1}{T}\sum_{i=1}^{T}(Y_i(u) - \overline{Y}_T(u))(Y_i(v) - \overline{Y}_T(v)) \tag{10}$$

  for a fixed value of d selected by the practitioner, so that

$$\hat{\lambda}_i\hat{\phi}_i(u) = \int \hat{C}_0(u,s)\hat{\phi}_i(s)ds\,. \tag{11}$$

- **Step 2**: Under the assumption of functional white noise, Theorem 1 provides an approximate distribution for the sequence $\{\iint T\hat{C}_h^2(t,s)dtds\,;\, h = 1,\ldots,H\}$. This distribution can be estimated as

$$\hat{Q}(d) = \sum_{j=1}^{d}\sum_{l=1}^{d}\hat{\lambda}_j\hat{\lambda}_l\chi^2_{j,l}(1), \tag{12}$$

  where d is a large number (below we select d as the number of eigenvalues $\hat{\lambda}_i$ of $\hat{C}_0(u,v)$ such that $\hat{\lambda}_i/\hat{\lambda}_1 > 0.0001$. This threshold was observed to give good results in the different empirical tests performed). This distribution is a linear combination of independent chi-squared random variables, and in practice it must be estimated. The method proposed in Imhof (1961) provides a suitable method to do so, providing good approximations in the tails of the distribution.

- **Step 3**: For a given confidence level $\alpha \in (0,1)$, and under the white noise assumption, an upper prediction bound for the sequence $\{\hat{\rho}_h\,;\, h = 1,\ldots,H\}$ can be set at

$$\frac{\sqrt{\hat{Q}(d)_{(1-\alpha)}}}{\sqrt{T}\int \hat{C}_0(u,u)du}, \tag{13}$$

where $\hat{Q}(d)_{(1-\alpha)}$ denotes the $(1-\alpha)$ critical value of the distribution of $\hat{Q}(d)$.

For small α, significant deviations of $\{\hat{\rho}_h\,;\, h = 1, \ldots, H\}$ beyond the boundary (13) can be used to identify structural dependence in the FTS, and also for rejecting the white noise hypothesis at a given lag.

2.2. Partial autocorrelation function for FTS

In classical time series analysis, the autocorrelation function is often accompanied by the partial autocorrelation function, which measures the correlation between two lags of the time series after removing the influence of the intermediate terms. By plotting these two functions together, autoregressive and moving average processes of any order can be identified by comparing them to the theoretical behavior of these plots when the order of the process is known. This section is devoted to developing the notion of a partial autocorrelation function for functional data, which can be used in conjunction with the proposed FACF to identify the correlation structure of a given FTS. There are a number of challenges in extending the notion of partial autocorrelation to a FTS $\{Y_t \,:\, t = 1, ..., T\}$, although here we propose and establish some of the large sample properties of such a quantity that is calculated analogously to the functional autocorrelation function defined in Section 2.1. It is similarly simple to interpret and calculate, and is useful in FTS model selection and diagnostic checking.

In plain language, the partial autocovariance between Y_t and Y_{t+h} is the autocovariance of the residuals of Y_t and Y_{t+h} produced by "regressing" them linearly on the intermediate variables $Y_{t+1}, ..., Y_{t+h-1}$. Towards a more formal definition of such a regression operation, first let $\boldsymbol{Y}_{t,h} = (Y_{t+1}, ..., Y_{t+h-1})$ denote the collection of these intermediary variables. Linear regression in this context can be carried out by selecting or estimating appropriate linear operators that relate $\boldsymbol{Y}_{t,h}$ to both Y_t and Y_{t+h}. Let $\mathcal{K}_h$ denote the collection of Hilbert Schmidt kernel integral (linear) operators mapping $(L^2([0,1]))^{\otimes h}$ to $L^2([0,1])$. More precisely, if $\boldsymbol{\Psi}_h \in \mathcal{K}_h$, and if $\boldsymbol{f}_h \in (L^2([0,1]))^{\otimes h}$ with $\boldsymbol{f}_h$ having component functions $\boldsymbol{f}_h = (f_{1,h}, ..., f_{h,h})$, then for almost every $s \in [0,1]$

$$\boldsymbol{\Psi}_h(\boldsymbol{f}_h)(s) = \sum_{j=1}^{h} \int \psi_h^{(j)}(v,s) f_{j,h}(v) dv,$$

for some component kernel functions $\psi_h^{(j)},\ j = 1, ..., h-1$ satisfying $\|\psi_h^{(j)}\| < \infty$. The collection of such operators may be made into its own Hilbert space endowed with the Hilbert-Schmidt norm

$$\|\mathbf{\Psi}_h\|_{HS} = \left(\sum_{j=1}^{h} \|\psi_h^{(j)}\|^2 \right)^{1/2}.$$

For example, $\{Y_t \ : \ t \in \mathbb{Z}\}$ follows a functional autoregressive model of order p, denoted ARH(p), if there exists an operator $\mathbf{\Psi}_p \in \mathcal{K}_p$ and an innovation sequence $\{\epsilon_t \ : \ t \in \mathbb{Z}\}$ so that

$$Y_t(u) = \mathbf{\Psi}_p(\mathbf{Y}_{t,p+1})(u) + \epsilon_t(u) = \sum_{j=1}^{p} \int \psi^{(j)}(v,u) Y_{t-j}(v) dv + \epsilon_t(u).$$

In this case the kernels $\psi^{(j)}$ $j = 1, ..., p$ are referred to as the autoregressive kernels. Now let

$$Proj(X, \mathbf{Y}_{t,h}) = \mathbf{\Psi}(\mathbf{Y}_{t,h}), \tag{14}$$

where $\mathbf{\Psi}$ satisfies

$$E\|X - \mathbf{\Psi}(\mathbf{Y}_{t,h})\|^2 = \inf_{\mathbf{\Psi}' \in \mathcal{K}_{h-1}} E\|X - \Psi'(\mathbf{Y}_{t,h})\|^2. \tag{15}$$

With this defined an intuitive definition of the partial autocovariance kernel at lag h of the series is

$$C_{h,h}(u,v) = \text{Cov}[Y_t(u) - Proj(Y_t, \mathbf{Y}_{t,h})(u), Y_{t+h}(v) - Proj(Y_t, \mathbf{Y}_{t,h})(v)].$$

The functional partial autocorrelation function (FPACF) may then be defined as

$$\rho_{h,h} = \frac{\|C_{h,h}\|}{\gamma_{1,h}\gamma_{2,h}},$$

where

$$\gamma_{1,h} = \left(\int E\{[Y_t(u) - Proj(Y_t, \mathbf{Y}_{t,h})(u)]\}^2 du \right)^{1/2},$$

and

$$\gamma_{2,h} = \left(\int E\{[Y_{t+h}(u) - Proj(Y_{t+h}, \mathbf{Y}_{t,h})(u)]\}^2 du \right)^{1/2}.$$

It follows by the Cauchy-Schwarz inequality that $0 \leq \rho_{h,h} \leq 1$.

Although this may serve as a logical analog to the scalar PACF, a number of theoretical and practical challenges are faced here stemming from the fact that estimating the projections of Y_t and Y_{t+h} onto $\mathbf{Y}_{t,h}$, or equivalently estimating the operators $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$ so that

$$E\|Y_t - \mathbf{\Psi}_{1,h}(\mathbf{Y}_{t,h})\|^2 = \inf_{\mathbf{\Psi}' \in \mathcal{K}_{h-1}} E\|Y_t - \Psi'(\mathbf{Y}_{t,h})\|^2,$$

and

$$E\|Y_{t+h} - \mathbf{\Psi}_{2,h}(\mathbf{Y}_{t,h})\|^2 = \inf_{\mathbf{\Psi}' \in \mathcal{K}_{h-1}} E\|Y_t - \Psi'(\mathbf{Y}_{t,h})\|^2, \tag{16}$$

in general presents an ill-posed inverse problem. Nonetheless, these operators may still be estimated in practice using standard regularization techniques, see e.g. Chapter 9 of Bosq (2000). For any fixed h, let $\hat{\mathbf{\Psi}}_{1,h}$ denote the operator obtained by performing a regularized functional linear regression of Y_t on $\mathbf{Y}_{t,h}$, and let $\hat{\mathbf{\Psi}}_{2,h}$ denote the operator obtained by performing a regularized functional linear regression of Y_{t+h} on $\mathbf{Y}_{t,h}$. We note that in the second case this may be viewed as fitting a functional autoregressive model of order $h-1$ to the series. Define

$$\hat{C}_{h,h}(u,v) = \frac{1}{T-h}\sum_{t=1}^{T-h}[Y_t(u) - \hat{\Psi}_{h,1}(\mathbf{Y}_{t,h})(u)][Y_{t+h}(v) - \hat{\Psi}_{h,2}(\mathbf{Y}_{t,h})(v)],$$

$$\hat{\gamma}_{1,h} = \left(\int \frac{1}{T-h}\sum_{t=1}^{T-h}[Y_t(u) - \hat{\Psi}_{h,1}(\mathbf{Y}_{t,h})(u)]^2 du\right)^{1/2},$$

and

$$\hat{\gamma}_{2,h} = \left(\int \frac{1}{T-h}\sum_{t=1}^{T-h}[Y_{t+h}(u) - \hat{\Psi}_{h,2}(\mathbf{Y}_{t,h})(u)]^2 du\right)^{1/2}.$$

Then, the empirical partial autocorrelation may then be defined as

$$\hat{\rho}_{h,h} = \frac{\|\hat{C}_{h,h}\|}{\hat{\gamma}_{1,h}\hat{\gamma}_{2,h}}.$$

It may be shown that under mild conditions on the estimators $\hat{\mathbf{\Psi}}_{1,h}$ and $\hat{\mathbf{\Psi}}_{2,h}$ that $\hat{\rho}_{h,h}$ has the same large sample properties as $\hat{\rho}_h$ when the data are a white noise, as described in the following theorem.

Theorem 2. If $\{Y_t : t \in \mathbb{Z}\}$ is independent and identically distributed with $E\|Y_t\|^4 < \infty$, $\|\hat{\Psi}_{h,1}\|_{HS} = o_P(T^{-1/4})$, and $\|\hat{\Psi}_{h,2}\|_{HS} = o_P(T^{-1/4})$, then

$$\hat{Q}_{h,h} = T\|\hat{C}_{h,h}\|^2 \xrightarrow{D} Q = \sum_{j=1}^{\infty}\sum_{l=1}^{\infty}\lambda_j\lambda_l\chi^2_{j,l}(1), \qquad h = 1, 2, \ldots$$

as $T \to \infty$. The $\lambda_i's$ are the eigenvalues of the covariance operator of Y_i. Additionally

$$\hat{\gamma}_{1,h} \xrightarrow{P} \left[\int C_0(u,u)du\right]^{1/2}, \text{ and } \hat{\gamma}_{2,h} \xrightarrow{P} \left[\int C_0(u,u)du\right]^{1/2}.$$

Remark 2.1. Most standard regularized estimators in this setting would satisfy $\|\hat{\mathbf{\Psi}}_{h,1}\|_{HS} = o_P(T^{-1/4})$, and $\|\hat{\mathbf{\Psi}}_{h,2}\|_{HS} = o_P(T^{-1/4})$ under mild conditions, including those derived from principal component regression and Tychonoff regularization. We elaborate on the PCA case, see e.g. Didericksen et al. (2012), Bosq (2000) and Zhang (2016). In case of Tychonoff regularization we refer to Kargin and Onatski (2008). In the case of PCA regression $\hat{\mathbf{\Psi}}_{h,2}$, and similarly $\hat{\mathbf{\Psi}}_{h,1}$, is estimated so that

$$\hat{\mathbf{\Psi}}_{h,2}(\boldsymbol{f}_h) = \sum_{j=1}^{h-1} \int \hat{\psi}_{h,2}^{(j)}(v,s) f_{j,h}(v)dv.$$

The kernels $\hat{\psi}_h^{(j)}$ are constructed so that

$$\hat{\psi}_{h,2}^{(j)}(v,s) = \sum_{r,\ell=1}^{\kappa} \hat{\psi}_{h,2}^{(j)}[r,\ell]\hat{\phi}_r(v)\hat{\phi}_\ell(v),$$

and the coefficient matrix $\hat{\psi}_{h,2}^{(j)}[\cdot,\cdot]$ is estimated to minimize the least squares criterion (16). For example, when $h = 2$, so that regression is performed on a single intermediary variable, this amounts to taking

$$\hat{\psi}_{h,2}^{(1)}[r,\ell] = \frac{\hat{\lambda}_\ell^{-1}}{T-1} \sum_{t=1}^{T-1} \langle Y_t, \hat{\phi}_\ell\rangle\langle Y_{t+1}, \hat{\phi}_r\rangle,$$

where the $\hat{\lambda}_\ell' s$ and $\hat{\phi}_\ell' s$ are the empirical eigenvalues and eigenvectors of the covariance operator defined in (5). κ in this case serves as a truncation level for the number of principal components employed. When κ is a fixed integer, then under the assumption that the data follow a strong white noise and standard eigenvalue spacing conditions $\|\hat{\mathbf{\Psi}}_{h,2}\|_{HS} = O_P(T^{-1/2})$, and even satisfies the Central Limit Theorem when normalized by multiplying by $T^{1/2}$. In order that the regression operator would be consistently estimated assuming that the underlying subspace on which the functional data take their values in $L^2([0,1])$ is infinite dimensional, it is often allowed that $\kappa = \kappa_T$ is a function of the sample size that increases slowly as T increases. In this case under sufficient decay conditions on the eigenvalues of the covariance operator defined in (5), it still holds that $\|\hat{\mathbf{\Psi}}_{h,2}\|_{HS} = o_P(T^{-1/4})$, see Chapter 9 of Bosq (2000) and Zhang (2016). Below all such regressions are carried out using FPCA where κ is chosen to explain more than 95% of the total variation of the data.

Theorem 2 shows that under the hypothesis that the series of interest is a strong white noise having fourth order moments, then the exact same prediction intervals for $\{\hat{\rho}_h, \quad h = 1, ..., H\}$ are asymptotically valid for $\{\hat{\rho}_{h,h}, \quad h = 1, ..., H\}$. Hence comparing $\{\hat{\rho}_{h,h}, \quad h = 1, ..., H\}$ to the threshold

$$\frac{\sqrt{\hat{Q}(d)_{(1-\alpha)}}}{\sqrt{T}\int \hat{C}_0(u,u)du}$$

for a user specified confidence level $(1-\alpha)$ provides a further way of evaluating the whiteness of the series or residuals under study.

Similarly to the standard scalar PACF, the functional PACF (FPACF) will tend to zero at lags larger than the autoregressive order when applied to consistently estimated ARH(p) models. In order to make this statement precise, we assume that $\{Y_t : t \in \mathbb{Z}\}$ follows an ARH(p) model

$$Y_t(u) = \mathbf{\Psi}_p^{ARH}(\boldsymbol{Y}_{t,p+1})(u) + \epsilon_t(u) = \sum_{j=1}^{p} \int \psi_p^{ARH,j}(v,u) Y_{t-j}(v)dv + \epsilon_t(u), \tag{17}$$

for some $\mathbf{\Psi}_p^{ARH} \in \mathcal{K}_p$ with components $\psi_p^{ARH,j}$. So that we may discuss the approximation of this operator as performed in evaluating the FPACF coefficient at lags larger than p, let $\mathbf{\Psi}_{h,p}^{ARH}$ denote the operator in $\mathcal{K}_h$ with component kernels $\psi_h^{ARH,j}$ satisfying

$$\psi_h^{ARH,j} = \begin{cases} \psi_p^{ARH,j} & j \leq p, \\ 0, & j > p. \end{cases} \tag{18}$$

Theorem 3. Suppose that $\{Y_t : t \in \mathbb{Z}\}$ satisfies the ARH(p) model in equation (18) with $0 < E\|\epsilon_0\|^4 < \infty$, and further suppose that the model admits a stationary and causal solution. If $h > p$, $\|\hat{\mathbf{\Psi}}_{2,h} - \mathbf{\Psi}_{h,p}^{ARH}\|_{HS} = o_P(1)$, and there exists an operator $\mathbf{\Psi}_{1,h} \in \mathcal{K}_{h-1}$ so that $\|\hat{\mathbf{\Psi}}_{1,h} - \mathbf{\Psi}_{1,h}\|_{HS} = o_P(1)$, then $\hat{\rho}_{h,h} \xrightarrow{P} 0$ as $T \to \infty$.

Theorem 3 implies that asymptotically the FPACF coefficients cutoff after lag p when estimated based on data following an ARH(p) model, which aligns with the traditional properties of the PACF. Necessary and sufficient conditions for the existence of a stationary and causal solution to a ARH(p) model are discussed in Chapter 5 of Bosq (2000).

3. Simulation study

Methods for estimating the proposed autocorrelation and partial autocorrelation functions for functional data have been implemented in the R package `fdaACF`, available in CRAN (`http:`

//cran.r-project.org/package=fdaACF), which we use in the following simulation study.

3.1. Application to Strong White Noise Processes

The FACF and FPACF introduced above can be used to evaluate the whiteness of a given FTS. In order to validate the proposed method, we consider several synthetic functional white noise processes. This section investigates the finite-sample properties of the method when applied to such processes.

Each functional observation has been discretized in a grid of 100 equi-spaced points in the interval $[0, 1]$. Four different white noise processes have been simulated, and the methods employed to generate each process are described below.

The first 3 series were generated by selecting a functional basis $\phi_k(v)$ and i.i.d. coefficients $b_{t,k}$, thus obtaining the elements of the process as

$$\varepsilon_t(v) = \sum_{k=1}^{K} b_{t,k}\phi_k(v)\,, \quad t = 1, \ldots, T. \tag{19}$$

As the coefficients of the series are i.i.d with respect to t, the FTS should not exhibit any kind of serial correlation. By varying the functional basis ϕ_k and the distributions of the coefficients $b_{t,k}$, different processes have been generated:

- $\varepsilon_t^{(1)}$ is obtained using the first 7 elements of the Fourier basis, due to the fact that this basis is often used when dealing with periodic functional data. The coefficients $b_{t,k}$ follow a standard normal distribution $N(0, 1)$.
- $\varepsilon_t^{(2)}$ is obtained using the first 7 elements of the B-spline basis, due to the fact that this basis is often used when dealing with non-periodic functional data. The coefficients are drawn from a Beta$(2, 5)$ distribution.
- $\varepsilon_t^{(3)}$ is obtained selecting 3 functions and applying the Gram-Schmidt method to obtain 3 orthonormal functions that will be used as a basis in (19). In this case, the functions selected were $\theta_1(v) = \sin(v)$, $\theta_2(v) = \exp v$ and $\theta_3(v) = \cos(v)$. In this case, the coefficients of (19) follow a Exp(1) distribution.

The last white noise process has been obtained simulating a brownian bridge random process, where each innovation function is defined as

$$\varepsilon_t(v) = W_t(v) - vW_t(1),$$

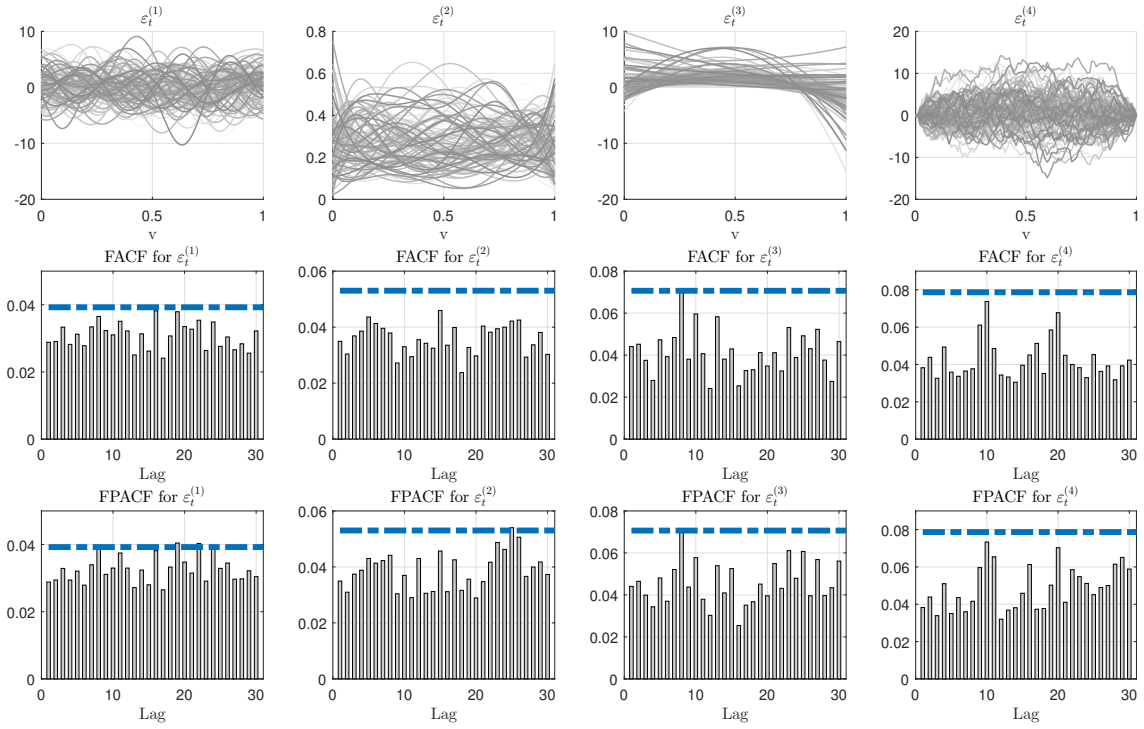

Figure 3: Simulated white noise processes and FACF/FPACF of each series. The dotted line denotes the 99% critical value of the distribution obtained by the proposed method.

where $W_t(v)$ is a Wiener processes defined on the domain $[0, 1]$.

- $\varepsilon_t^{(4)}$ is generated by a brownian bridge process where the variance of the Wiener process is 1.

Figure 3 shows a representative example of the first 50 values of the FACF of each white noise process, including the 99% prediction bound provided by Theorem 1 that has been obtained using the `obtain_FACF` function of the `R` package `fdaACF`. The value of d was selected as the number of eigenvalues $\hat{\lambda}_i$ of $\hat{C}_0(u, v)$ such that $\hat{\lambda}_i/\hat{\lambda}_1 > 0.0001$. Evidently most of the coefficients of the FACF fall below this limit, which is consistent with the white noise hypothesis.

In order to test the empirical properties of the proposed statistic for different confidence values, all 4 white noise series described in this section have been simulated, using different configurations of the length of the time series ($T = 100, 250, 500, 1000, 2000$) and using different lags when obtaining the limit value for the lagged autocorrelations. 50 repetitions of each configuration has been obtained, and Table 1 shows the percentage of out-of-bounds autocorrelations obtained in the previous analysis for the confidence levels $\alpha = 0.1, 0.05, 0.01$. Although it has been observed that the empirical out-of-bound rate seems to be quite conservative for significance level 10% when

increasing the number of lags considered, the overall performance of the test agrees with the selected confidence level α. Similar results are obtained when using the FPACF instead of the FACF, and so we omit those results due to space considerations.

3.2. Application to Serially Correlated Processes

This section present several simulations of functional dependent processes, each one exhibiting some kind of serial correlation. The proposed identification method based on the FACF and the FPACF will be used in order to identify the underlying structure of the simulated FTS; fit a model using that information and check that the residuals of the fitted model plausibly correspond to a white noise process, validating the model.

A crucial step in the modeling of scalar time series is the identification of the structure of the underlying stochastic process. The Box-Jenkins methodology (Box et al., 2008) proposes a standard model identification procedure for scalar time series. Once the stationarity and seasonality of the series has been addressed, the next step is to identify the seasonal and regular order of the AR and MA component of the transformed time series. The main tools used in this identification procedure are the autocorrelation and partial autocorrelation functions, whose plots can be used to obtain useful information about the underlying structure of the stochastic process that has generated the series. These tools can also be used in the diagnosis of the model in order to check the white noise assumption of the residuals.

The purpose of this section is to apply the proposed method in order to obtain a similar tool for FTS. Regarding the identification of FTS models, there has been several attempts to obtain the optimum order of some functional processes. In Kokoszka and Reimherr (2013), a test statistic is derived in order to check if a ARH(p) model fits the data better than a ARH($p-1$) model, by checking if the operator ϕ_p of the formulation of the ARH(p) is significantly different from the null operator. However, up to the authors' knowledge, no general identification procedure analogous to the Box-Jenkins methodology for scalar time series has been proposed for FTS. This section is aimed at showing that the FACF and FPACF proposed in Section 2 can be considered as a first step in this direction, offering a useful tool for the identification of certain patterns in the structure of the residuals of a fitted functional model that should be included in the model in order to increase its accuracy.

In order to illustrate the applicability of this method as a diagnosis tool, several functional

			Significance level						Significance level		
	T	H	10%	5%	1%		T	H	10%	5%	1%
$\varepsilon_t^{(1)}$	100	10	7.11%	3.65%	0.58%	$\varepsilon_t^{(3)}$	100	10	5.77%	3.46%	1.15%
		20	3.84%	1.73%	0.38%			20	6.92%	2.98%	0.57%
		30	3.01%	1.08%	0.12%			30	5.7%	2.88%	0.51%
	250	10	8.46%	3.07%	0.57%		250	10	9.03%	4.23%	1.15%
		20	6.53%	3.26%	0.48%			20	6.82%	3.65%	0.48%
		30	5.7%	2.75%	0.32%			30	6.98%	3.14%	0.32%
	500	25	7.36%	3.28%	0.88%		500	25	10.48%	5.04%	1.12%
		50	6.88%	2.92%	0.56%			50	8.4%	4.12%	0.88%
		100	3.88%	1.92%	0.12%			100	6.96%	3.36%	0.68%
	1000	25	10.08%	4.88%	0.88%		1000	25	9.76%	4.16%	0.56%
		50	7.84%	3.72%	0.96%			50	8.96%	4.44%	0.8%
		100	6.56%	2.98%	0.56%			100	7.66%	3.8%	0.74%
	2000	25	7.76%	3.84%	0.56%		2000	25	11.52%	5.12%	0.8%
		50	9.28%	4.56%	0.88%			50	9.08%	4.64%	1.04%
		100	8.26%	4.1%	0.68%			100	8.74%	4.24%	0.92%
$\varepsilon_t^{(2)}$	100	10	7.69%	3.27%	0.57%	$\varepsilon_t^{(4)}$	100	10	8.46%	4.42%	0.38%
		20	6.53%	3.56%	0.77%			20	6.82%	2.98%	0.46%
		30	5.06%	2.43%	0.44%			30	6.21%	2.75%	0.51%
	250	10	9.23%	4.42%	0.96%		250	10	10.77%	4.42%	0.76%
		20	9.13%	4.32%	0.86%			20	8.46%	3.94%	0.29%
		30	8.78%	4.23%	0.89%			30	8.84%	4.42%	0.06%
	500	25	9.28%	4.32%	0.4%		500	25	9.2%	5.12%	1.28%
		50	7.72%	3.72%	0.64%			50	7.6%	3.88%	0.72%
		100	6.42%	2.9%	0.4%			100	7.58%	3.38%	0.62%
	1000	25	9.28%	4.56%	1.2%		1000	25	12.08%	5.92%	1.36%
		50	9.2%	4.72%	1.04%			50	9.68%	4.6%	0.92%
		100	8.12%	3.84%	0.6%			100	8.38%	3.64%	0.58%
	2000	25	9.04%	4.32%	1.04%		2000	25	10.4%	6.24%	1.44%
		50	10.08%	4.88%	1.16%			50	9.6%	4.36%	0.78%
		100	8.66%	4.1%	0.8%			100	9.26%	4.58%	0.98%

Table 1: Empirical rate of out-of bounds FACF coefficients obtained for several simulated functional white noise processes.

series have been simulated including some specific patterns of temporal dependence. A functional SARMAHX model (Portela et al., 2018; Mestre et al., 2020) will be fitted to each time series, and the method proposed in this paper will be used to test the white noise assumption on the residuals of the fitted model. The SARMAHX model is an extension of the standard seasonal ARMA model to the functional framework by means of functional operators, accounting for autoregressive and moving average effects as well as allowing the inclusion of exogenous variables to the model. Given a stationary and centered FTS $\{Y_t(v)\,;\, t = 1, 2, ..., T\,;\, v \in V\}$ and a set $\{X_t^z(u_z)\,;\, z \in \mathbb{Z}\,;\, t = 1, 2, \ldots, T\,;\, u_z \in V_z\}$ of Z, potentially functional covariates, the expression for the SARMAHX $(p, q) \times (P, Q)_s$ model is defined as:

$$\begin{aligned}
&\left(I - \Psi_1 B - \cdots - \Psi_p B^p\right)\left(I - \Phi_1 B^s - \cdots - \Phi_P B^{P\cdot s}\right) Y_t = \\
&\left(I - \Theta_1 B - \cdots - \Theta_q B^q\right)\left(I - \Upsilon_1 B^s - \cdots - \Upsilon_Q B^{Q\cdot s}\right) \varepsilon_t \\
&+ \Gamma_1(X_t^1) + \cdots + \Gamma_Z(X_t^Z),
\end{aligned} \tag{20}$$

where Ψ_i and Φ_i are the regular and seasonal autoregressive operators, Θ_i and Υ_i the regular and seasonal moving average operators, and Γ_i the operators related to the Z explanatory variables. B^n is the lag operator which is defined as $B^n Y_t = Y_{t-n}$ where $n \in \mathbb{N}$. Finally, I is the identity operator.

Each function of the series will be discretized in a grid of 100 equi-spaced points in the interval $[0, 1]$. The length of the simulated processes will be 1500 functional observations, after leaving 500 observations as a burn-in period in order to stabilize the series.

The first simulated process $Y_t^{(1)}$ is an ARH(1) model, given by the equation

$$Y_t^{(1)}(v) = \Psi(Y_{t-1}^{(1)}(v)) + \varepsilon_t(v) = \int \psi(u, v) Y_{t-1}^{(1)}(u) du + \varepsilon_t(v), \tag{21}$$

where $\Psi(.)$ is an integral operator that defines the process, $\psi(u, v)$ denotes the kernel of said integral operator and $\varepsilon_t(v)$ denotes the white noise innovations of the process. This process, studied in Bosq (2000), is stationary if $\|\Psi\| < 1$, so the simulated series will use the Gaussian kernel

$$\psi(u, v) = 0.6 \exp\left(\frac{u^2 + v^2}{2}\right),\ (u, v) \in [0, 1] \times [0, 1], \tag{22}$$

so $\|\Psi\| \approx 0.7$, ensuring that the resulting series will be stationary. The brownian bridges simulated in the last section will be used as the innovations of the process. The top panels of Figure 4 show the FACF and FPACF of the process $Y_t^{(1)}$ for $H = 20$ lags, together with the 99% i.i.d. bounds

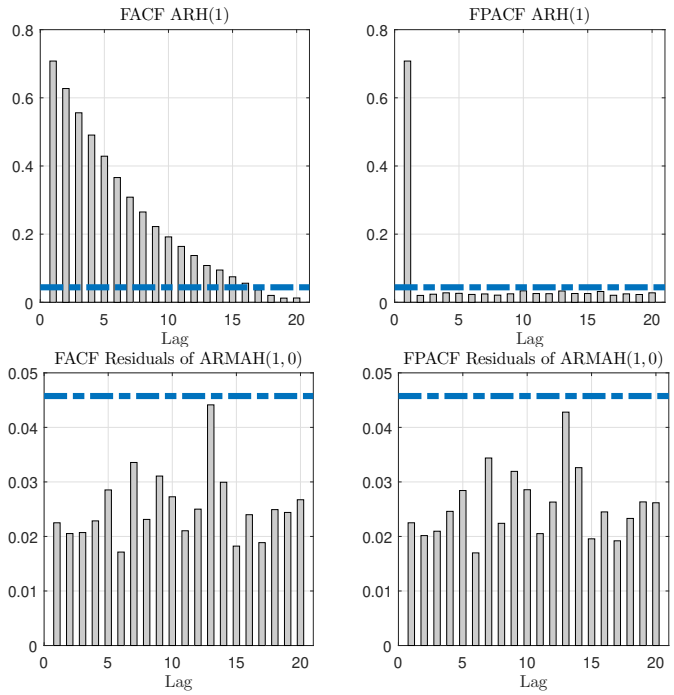

Figure 4: Top: FACF and FPACF for a simulated ARH(1) process. Bottom: FACF and FPACF of the residuals of a fitted ARMAH$(1,0)$ model. The dotted line represents the 99% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

provided by Theorem 1. The coefficients of the FACF exhibit an exponential decay that agrees with the expected behavior of the ARH(1) process.

However, the FACF alone is not enough to identify the order of the autoregressive process– an ARH process of higher order could produce a similar FACF structure. Similarly to the scalar case, the FPACF can be used to better identify the order of autoregressive process, as it aims to remove the influence of the intermediate terms of the series at a given lag. This is illustrated in the top FPACF of Figure 4, where only the first coefficient is significant, which is consistent with the dependence structure of an ARH(1) process.

As most coefficients of the FACF fall above the critical value, the white noise assumption is rejected. The test proposed in Horváth et al. (2013) (referred to as the HHR hereafter) also rejects the white noise hypothesis at levels $\alpha = 0.1, 0.05$ and 0.01. Because of this, a ARMAHX$(1,0)$ model has been fitted to this FTS, and both FACF and FPACF of the residuals of the model are

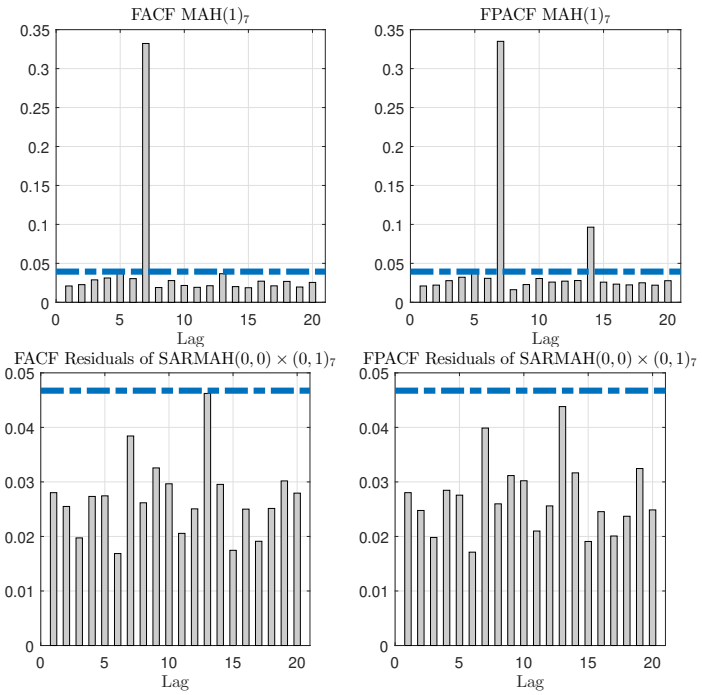

Figure 5: Top: FACF and FPACF for the simulated MAH(1)$_7$ process. Bottom: FACF and FPACF and the residuals of a fitted SARMAHX$(0,0)\times(0,1)_7$ model. The dotted line represents the 99% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

plotted on the bottom panels of Figure 4. As shown in the image, no pattern can be found in the residuals of the model, and all the coefficients fall below the critical value, suggesting the residuals of the model can be treated as a functional white noise. This is corroborated by the HHR test, which does not reject the white noise hypothesis for confidence level $\alpha = 0.1, 0.05$ and 0.01.

The next simulated process $Y_t^{(2)}$ will have the structure of a seasonal MAH(1)$_7$ (Aue et al., 2015), where each observation depends on the 7th lagged term of the series $Y_{t-7}^{(2)}$, following the equation

$$Y_t^{(2)}(v) = \Theta(\varepsilon_{t-7}(v)) + \varepsilon_t(v) = \int \theta(u,v)\varepsilon_{t-7}(u)du + \varepsilon_t(v), \tag{23}$$

where $\Theta(.)$ denotes an integral operator with kernel $\theta(u,v)$, and $\varepsilon_t(v)$ denotes white noise variables. As in the previous series, the innovations will be independent brownian bridges, and the kernel used to define the integral operator will be the Gaussian kernel defined in (22), so $\|\Theta\| \approx 0.7$, resulting in a stationary series. The FACF and FPACF of the simulated series are shown in the top panels of

Figure 5. Due to the innovations being uncorrelated, the FACF plot only shows dependence of the lag 7, whereas the FPACF plot shows a certain, approximately geometric, decay for the seasonal lags 7 and 14. These plots exhibit the shape of a moving average process with seasonality 7. When applied to this data, the HHR test rejects the null hypothesis of functional white noise at levels $\alpha = 0.1, 0.05$ and 0.01.

Once the process has been identified, a ARMAHX$(0,0) \times (0,1)_7$ model is fitted to the data. In order to validate the model, the main hypotheses of the model have to be checked. The values of the lagged autocorrelations of the residuals are obtained as well as the 99% limit distribution of the statistic to test the white noise assumption on the residuals. The HHR test does not reject the null hypothesis of functional white noise at level $\alpha = 0.1$, and when using the method proposed in this paper, similar results were obtained: as shown in the bottom panels of Figure 5; none of the autocorrelation values surpasses the 99% limit, hence the assumption of white noise can not be rejected, and the residuals of the fitted model can be regarded as functional white noise.

These results show the applicability of the proposed method in the diagnosis of FTS models, providing the practitioner with a graphical tool to visualize the structure of the FTS and identify possible deterministic components in the residuals of a model that have not been explained. Figure 6 shows the FACF and FPACF of several functional autoregressive and moving average processes that can be used to identify seasonality, the dependence structure of FTS and the order of the process: an exponential decay on the FACF is associated with an autoregressive process, and significant values of the FPACF indicate the order of the process. On the other hand, moving average processes are associated with a FPACF that exhibits an exponential decay and a FACF that has only a reduced number of values above the i.i.d. upper bound, where the number of significant coefficients of the FACF indicate the order of the moving average process. In the simulations of this section, both the proposed method and the HHR test have proven to provide similar results and rejection rates for FTS that exhibited some kind of time dependency between its terms. Next section will apply this identification and diagnosis method to several real-world datasets, in order to test the usefulness when dealing with real data.

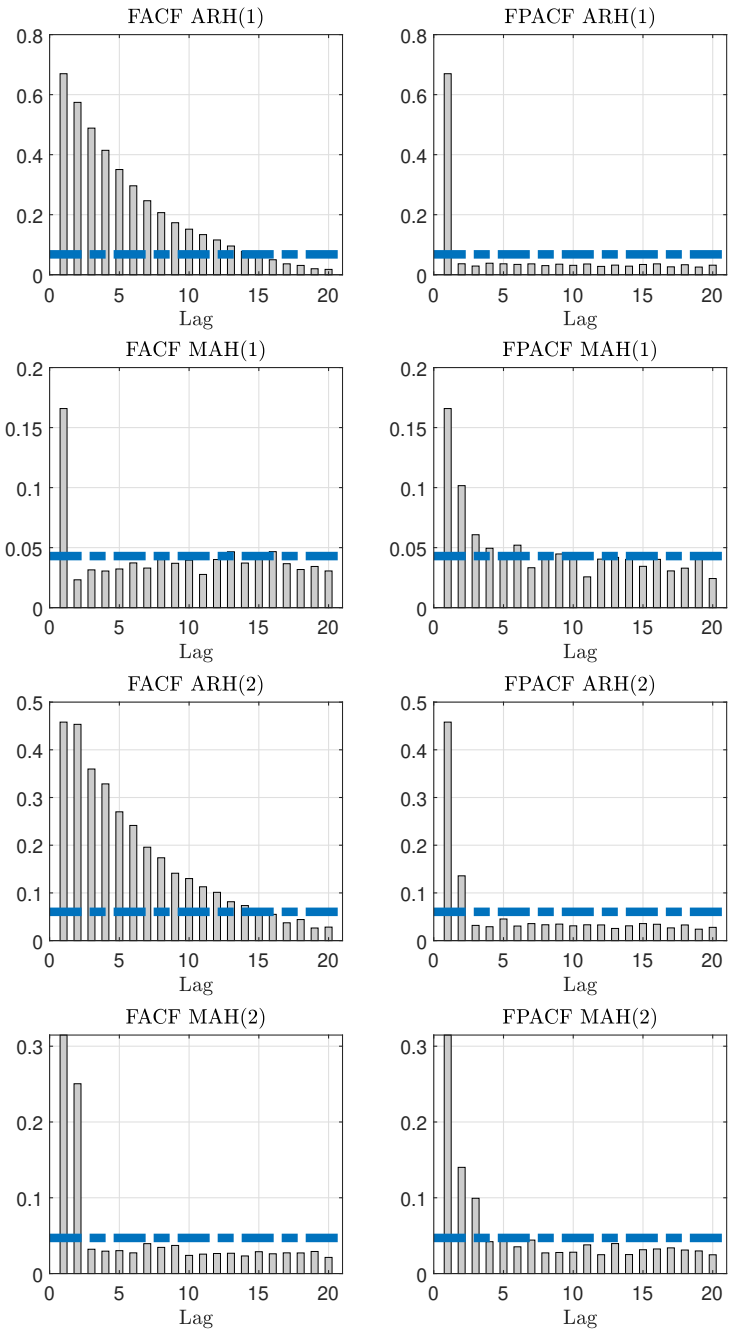

Figure 6: FACF and FPACF of four different simulated functional autoregressive and moving average processes. The shape exhibited by the autocorrelation function of each process can be used to identify each process. The dotted line represents the 99% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

4. Application to real-world datasets

To illustrate the applicability of the proposed method, two widely studied real-world datasets will be analyzed. Firstly, the FACF and FPACF will be obtained to quantify the correlation structure of the FTS in order to check the white-noise assumption on the original time series and to identify AR and MA trends. Secondly, a functional linear model will be fitted to each series and the test will be applied to the residuals of the models, in order to diagnose the models. If the white noise assumption is not rejected on the residuals, it will indicate that the model has extracted all the information of the original time series. In order to validate the proposed test, the results obtained will be compared with the HHR test.

4.1. Eurodollar futures contracts

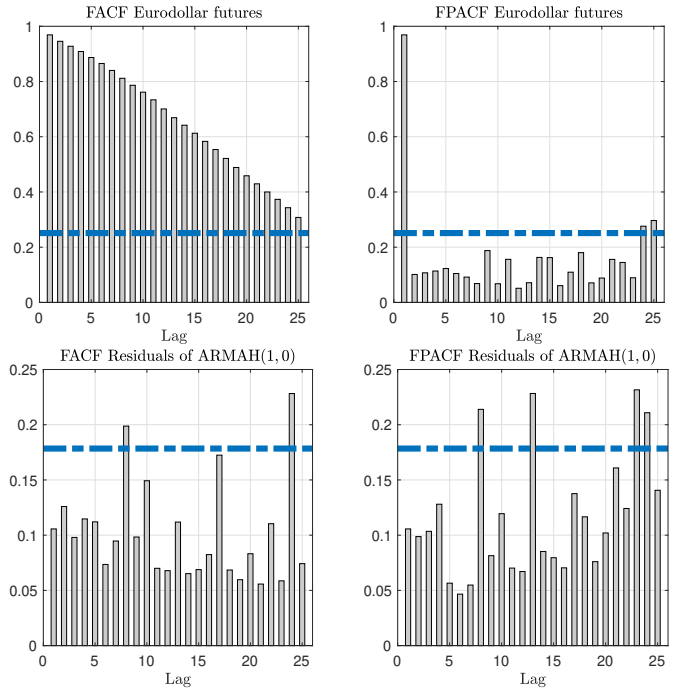

Figure 7: Top: FACF and FPACF of the Eurodollar dataset. Bottom: FACF and FPACF of residuals of a fitted ARMAHX$(1, 0)$ model. The dotted line represents the 95% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

The first real-world dataset to be analyzed consists of Eurodollar future rates (Kargin and

Onatski, 2008; Horváth et al., 2013). As stated in those works, each Eurodollar futures contract is an obligation to deliver a 3-month deposit of $1,000,000 to a bank account outside of the United States at a specified time. Given v a scalar value, if one defines $Y_t^E(v)$ as the price of a contract with closing date v months from day t, data from this dataset can be regarded as a FTS. The Eurodollar dataset construct those curves using 114 points for each day, and in order to replicate the results of the HHR test (Horváth et al., 2013), the sample considered in this analysis consists of 100 days of data taken from January 24 to June 17,1997.

An ARH(1) model was identified in Kargin and Onatski (2008), where the authors develop a technique to estimate the autoregression operator of the model, and illustrate this technique to the Eurodollar dataset. However, no identification procedure was used to check if the data can be modeled as an ARH(1). The top panels of Figure 7 show the FACF and FPACF of the Eurodollar series. The shape of the FACF plot shows an exponential decay, whereas only the first coefficient of the FPACF is significant. The shape exhibited by the correlogram is similar to the simulated ARH(1) process shown in Figure 4; hence validating the selection of an autoregressive model to model the data. The presence of structural dependence between the data is also confirmed by the 95% critical line displayed in the plot: all of the coefficients of the FACF fall above of the critical bound, hence the assumption of functional white noise must be dropped. This was also observed in Horváth et al. (2013), where the authors applied the HHR test to the same data and rejected the white noise hypothesis at 99% confidence level.

Once the structure of the process has been identified, an ARMAHX$(1, 0)$ model has been fitted to the data, and the FACF and FPACF of the model's residuals are displayed on the bottom panels of Figure 7. The HHR test does not reject the white noise hypothesis at confidence level $\alpha = 0.1$. This result is consistent with the outcome of the diagnosis methodology developed in this paper: most of the coefficients of the FACF and FPACF obtained for the residuals fall below the 95% critical bound and no structure can be identified in the plots, hence assuming that the residuals of the model are white noise, validating the selected model.

4.2. Spanish electricity daily price profiles

The second dataset was studied in Portela et al. (2018), where the authors analyze the series of hourly Spanish electricity spot prices provided by the Spanish electricity Market Operator (`www.omie.es`). As explained in the paper, the spot price series $\{y_{t,h}\}$ is transformed into a FTS $\{Y_t^P(v)\,;\, t = 1,\ldots,T\,;\, v \in [1,24]\}$ where each function Y_t is observed at discrete hours $v_i \in \{1,\ldots,24\}$, thus $Y_t^P(v_i) = y_{t,v_i}$. Hence each observation is a daily price profile. In order to replicate the results of that work, the sample considered in this analysis consists of 1 year of data taken from January 1, 2014 to December 31, 2014. The daily price profiles are shown in Figure 1. We here note that all below functional regressions models fitted to obtain the FPACF are estimated using FPCA regularization with level chosen by so that 95% of the total variance is explained by the selected number of FPC's.

In order to model the electricity price series, Portela et al. (2018) proposed a SARMAHX$(1,0)\times(0,1)_7$ model, using the daily demand X_t^D and wind profiles X_t^W as exogenous functional variables. However, the authors relied on FPCA reduction techniques and testing different configurations in order to identify the structure of the model. Besides, no white noise identification test was applied on the residuals. The purpose of this section is to illustrate the application of the autocorrelation functions in the identification and diagnosis of a model, by testing the white noise hypothesis on the residuals.

The top panels of Figure 8 show the FACF and FPACF for the first 30 lags of the price functional time series. This diagnostic tool provides useful information about the correlation structure of the series: firstly, all the coefficients of the FACF lie above the 99% critical bound, so the white noise assumption is strongly rejected. The HHR test provides the same conclusion. Secondly, the diagnostic plot exhibits a seasonal dependency with period 7, which can be explained by the daily nature of the series. Moreover, a slow decrease in the values of the FACF is observed, which might imply that the series should be differenced. The bottom panels of Figure 8 show the FACF and FPACF of the price series after being seasonally differenced at lag 7. As can be seen the slow decrease of the coefficients in the FACF is no longer observed. Hence, the differenced series are

used, which are be denoted as

$$Z_t^P = Y_t^P - Y_{t-7}^P \tag{24}$$

$$U_t^D = X_t^D - X_{t-7}^D \tag{25}$$

$$U_t^W = X_t^W - X_{t-7}^W, \tag{26}$$

where Z_t^P denotes the series of seasonally differenced daily electricity price profiles and U_t^D and U_t^W denote the demand and wind production series after the seasonal differentiation.

In order to model the price time series, an initial SARMAHX regression model is estimated using the differenced demand and wind production series as inputs. The expression of this regression model is as follows:

$$Z_t^P = \Gamma_W(U_t^W) + \Gamma_D(U_t^D) + \varepsilon_t, \tag{27}$$

where Γ_W and Γ_D are the integral operators related to the differenced demand and wind exogenous variables. The top panels of Figure 9 show the FACF and FPACF of the regression model residuals, which can be used to extract information about the correlation structure of the underlying process that generated the series. Most of the values of the FACF fall below the critical limit, however there seems to be a strong correlation on the first lags, and the white noise assumption is rejected at the 99% confidence level using the HHR test. Consequently, using the exogenous variables is not enough for modeling the series. When compared with Figures 4 and 5, the FACF and FPACF plots seems to have two components: a regular autoregressive behavior (identified by the first coefficient of the FPACF) and a moving average effect on lag 7.

The next step in this analysis is to fit a SARMAHX$(0,0) \times (0,1)_7$ model to the differenced series, in order to capture this seasonal moving average effect. The expression for this models is:

$$Z_t^P = \Gamma_W(U_t^W) + \Gamma_D(U_t^D) - \Theta(\varepsilon_{t-7}) + \varepsilon_t. \tag{28}$$

The correlation structure of the residuals of that model is shown in the bottom panels of Figure 9. As can be seen, the moving average effect has been removed. However, the first lags of the FACF function lie above the critical limit and exhibit a certain exponential decay pattern in the FACF, whereas the FPACF still has one significant value at lag 1. This causes both the proposed test and the HHR test to reject the white noise hypothesis at 99% confidence level.

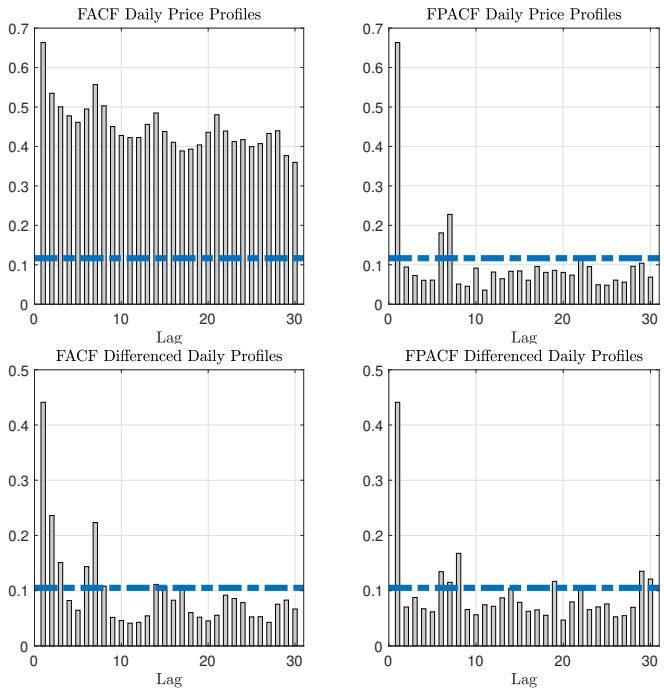

Figure 8: Top: FACF and FPACF of the daily price profiles series Y_t^P. Bottom: FACF and FPACF of the seasonally differenced series Z_t^P. The dotted line represents the 99% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

Finally, a SARMAHX$(1,0)\times(0,1)_7$ model has been fitted to the differenced series of daily price profiles. The expression for this model is:

$$Z_t^P = \Gamma_W(U_t^W) + \Gamma_D(U_t^D) + \Psi(Z_{t-1}^P) - \Theta(\varepsilon_{t-7}) + \varepsilon_t. \tag{29}$$

The autocorrelation plot of the residuals of the model is shown in Figure 10. All the values of both the FACF and FPACF fall below the critical limit, hence the white noise assumption cannot be rejected. This is validated by the HHR test.

This process illustrates the proposed identification and diagnosis procedure that has been developed for FTS linear models. For models within the SARMAHX family of models, this tool is useful to select the optimum configuration of its parameters. This enables the practitioner to capture the dynamics of the underlying process, obtaining more accurate forecasts of the FTS.

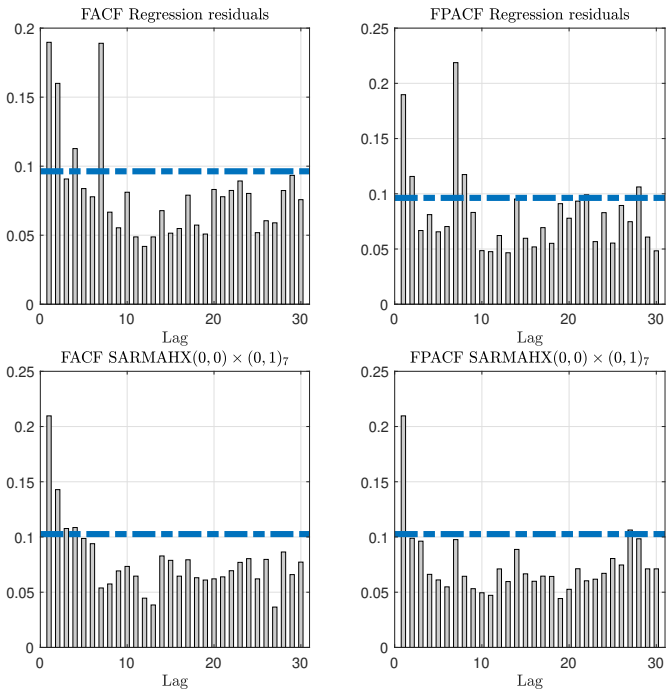

Figure 9: Top: FACF and FPACF of the residuals of the regression model. Bottom: FACF and FPACF of the residuals of the fitted SARMAHX$(0,0)\times(0,1)_7$ model. The dotted line represents the 99% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

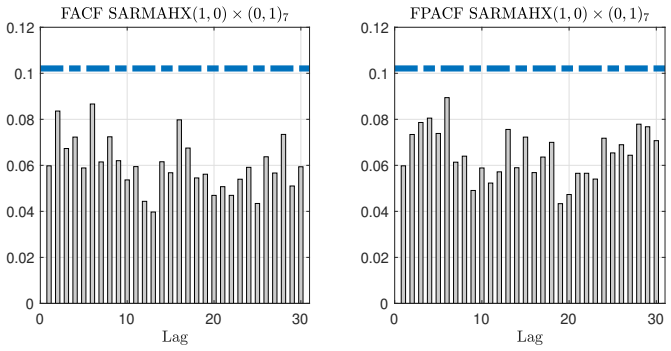

Figure 10: FACF and FPACF of the residuals of the SARMAHX$(1,0)\times(0,1)_7$ model. The dotted line represents the 99% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

5. Conclusions

This paper presents statistical tools that can be used to identify the serial dependence structure of a given FTS and identify suitable FTS models. The motivation for these tools was to define

versions of the autocorrelation and partial autocorrelation functions for FTS, with similar properties as their scalar counterpart. The statistical tool proposed in this paper differs from other white noise testing procedures found in the literature in several aspects: where most testing methods rely on FPCA to reduce the dimension of the curves of the series, the proposed methods use the correlation structure of the FTS to capture the structure of the data. Hence, no dimensionality reduction is applied, avoiding the need of choosing functional principal components and approximating the functional series with a finite number of terms, thus losing some information in this process. In addition, while existing functional white noise tests can be used to evaluate whether a given series or sequence of residuals is plausibly a white noise, the use of FACF and FPACF plots are helpful in identifying the specific serial correlation structure, providing some insight when fitting a linear model to that series. This is particularly useful in identifying the order of functional autoregressive and moving average processes.

The performance of the method has been validated both numerically and empirically, identifying the correlation structure of both Eurodollar future contracts and spanish electricity price profiles FTS and characterizing the residuals of the fitted models as functional white noise. It has been shown that the white noise identification has a similar performance to other white noise tests. The proposed methods have been implemented and are currently available in the `R` package `fdaACF`.

Although the proposed FACF and FPACF can be used in a similar way as their scalar counterparts to identify and diagnose FTS models, the fact that these functions are based on the L^2 norm of the lagged covariance operators makes it difficult to identify whether the correlation exhibited by the series is "positive" or "negative" in nature; only the magnitude is measured. The development of a signed version of the proposed autocorrelation functions could overcome these problems, providing the practitioner with a useful tool to model FTS. We leave this topic as an area for future research.

References

Aue, A., Norinho, D.D., Hörmann, S., 2015. On the Prediction of Stationary Functional Time Series. Journal of the American Statistical Association 110, 378–392. doi:10.1080/01621459.2014.909317.

Bagchi, P., Characiejus, V., Dette, H., 2018. A Simple Test for White Noise in Functional Time Series. Journal of Time Series Analysis 39, 54–74. doi:10.1111/jtsa.12264.

Bosq, D., 2000. Linear Processes in Function Spaces - Theory and Applications. Number 149 in Lecture Notes in Statistics, Springer-Verlag New York.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 2008. Time series analysis: forecasting and control. Wiley series in probability and statistics. 4th ed ed., John Wiley, Hoboken, N.J.

Canale, A., Vantini, S., 2016. Constrained functional time series: Applications to the Italian gas market. International Journal of Forecasting 32, 1340–1351. doi:10.1016/j.ijforecast.2016.05.002.

Characiejus, V., Rice, G., 2020. A general white noise test based on kernel lag-window estimates of the spectral density operator. Econometrics and Statistics 13, 175–196. doi:10.1016/j.ecosta.2019.01.003.

Damon, J., Guillas, S., 2002. The inclusion of exogenous variables in functional autoregressive ozone forecasting. Environmetrics 13, 759–774. doi:10.1002/env.527.

Damon, J., Guillas, S., 2005. Estimation and Simulation of Autoregressive Hilbertian Processes with Exogenous Variables. Statistical Inference for Stochastic Processes 8, 185–204. doi:10.1007/s11203-004-1031-6.

Didericksen, D., Kokoszka, P., Zhang, X., 2012. Empirical properties of forecasts with the functional autoregressive model. Computational Statistics 27, 285–298. doi:10.1007/s00180-011-0256-2.

Gabrys, R., Horváth, L., Kokoszka, P., 2010. Tests for Error Correlation in the Functional Linear Model. Journal of the American Statistical Association 105, 1113–1125. doi:10.1198/jasa.2010.tm09794.

Gabrys, R., Kokoszka, P., 2007. Portmanteau Test of Independence for Functional Observations. Journal of the American Statistical Association 102, 1338–1348. doi:10.1198/016214507000001111.

Hörmann, S., Horváth, L., Reeder, R., 2013. A functional version of the ARCH model. Econometric Theory 29, 267–288. doi:10.1017/S0266466612000345.

Horváth, L., Hušková, M., Rice, G., 2013. Test of independence for functional data. Journal of Multivariate Analysis 117, 100–119. doi:10.1016/j.jmva.2013.02.005.

Horváth, L., Kokoszka, P., 2012. Inference for Functional Data with Applications. Springer Series in Statistics, Springer-Verlag New York.

Hyndman, R.J., Shang, H.L., 2010. Rainbow Plots, Bagplots, and Boxplots for Functional Data. Journal of Computational and Graphical Statistics 19, 29–45. doi:10.1198/jcgs.2009.08158.

Imhof, J.P., 1961. Computing the Distribution of Quadratic Forms in Normal Variables. Biometrika 48, 419–426. doi:10.2307/2332763.

Kargin, V., Onatski, A., 2008. Curve forecasting by functional autoregression. Journal of Multivariate Analysis 99, 2508–2526. doi:10.1016/j.jmva.2008.03.001.

Klepsch, J., Klüppelberg, C., Wei, T., 2017. Prediction of functional ARMA processes with an application to traffic data. Econometrics and Statistics 1, 128–149. doi:10.1016/j.ecosta.2016.10.009.

Kokoszka, P., Reimherr, M., 2013. Determining the order of the functional autoregressive model. Journal of Time Series Analysis 34, 116–129. doi:10.1111/j.1467-9892.2012.00816.x.

Kokoszka, P., Reimherr, M., 2017. Introduction to Funcional Data Analysis. Statistical Theory & Methods, CRC press.

Kokoszka, P., Rice, G., Shang, H.L., 2017. Inference for the autocovariance of a functional time series under conditional heteroscedasticity. Journal of Multivariate Analysis 162, 32–50. doi:10.1016/j.jmva.2017.08.004.

Martínez-Hernández, I., Genton, M.G., González-Farías, G., 2019. Robust depth-based estimation of the functional autoregressive model. Computational Statistics & Data Analysis 131, 66–79. doi:10.1016/j.csda.2018.06.003.

Mas, A., Pumo, B., 2010. Linear Processes for Functional Data, in: Ferraty, F., Romain, Y. (Eds.), The Oxford Handbook of Functional Data. Oxford University Press. volume 1, pp. 47–51.

Mestre, G., Portela, J., Muñoz, A., Alonso, E., 2020. Forecasting hourly supply curves in the Italian Day-Ahead electricity market with a double-seasonal SARMAHX model. International Journal of Electrical Power & Energy Systems 121, 106083. doi:10.1016/j.ijepes.2020.106083.

Portela, J., Muñoz, A., Alonso, E., 2018. Forecasting Functional Time Series with a New Hilbertian ARMAX Model: Application to Electricity Price Forecasting. IEEE Transactions on Power Systems 33, 545–556. doi:10.1109/TPWRS.2017.2700287.

Ramsay, J., Hooker, G., Graves, S., 2009. Functional Data Analysis with R and MATLAB. Springer New York, New York, NY.

Ramsay, J., Silverman, B.W., 2005. Functional Data Analysis. Springer Series in Statistics, Springer-Verlag New York.

Turbillon, C., Marion, J.M., Pumo, B., 2007. Estimation of the moving-average operator in a Hilbert space, in: Recent Advances in Stochastic Modeling and Data Analysis, World Scientific Publications, Chania, Greece. pp. 597–604.

Zhang, X., 2016. White noise testing and model diagnostic checking for functional time series. Journal of Econometrics 194, 76–95. doi:10.1016/j.jeconom.2016.04.004.

Appendix A. Proofs of the results of Section 2

Proof of Theorem 2. We begin by showing that

$$\hat{Q}_{h,h} = T\|\hat{C}_{h,h}\|^2 \xrightarrow{D} Q = \sum_{j=1}^{\infty}\sum_{l=1}^{\infty} \lambda_j \lambda_l \chi^2_{j,l}(1). \tag{A.1}$$

This follows from the continuous mapping theorem upon establishing that $\sqrt{T}\hat{C}_{h,h}$ converges weakly in $L^2([0,1]^2)$ to a Gaussian element in $L^2([0,1]^2)$ with mean zero and covariance kernel

$$E[\Gamma(u,v)\Gamma(u',v')] = C_0(u,u')C_0(v,v'). \tag{A.2}$$

It follows from elementary arguments that

$$\sqrt{T}\|\hat{C}_{h,h} - \tilde{C}_{h,h}\| = o_P(1), \tag{A.3}$$

where

$$\tilde{C}_{h,h}(t,s) = \frac{1}{T}\sum_{i=1}^{T}[Y_i(t) - \hat{\mathbf{\Psi}}_{h,1}(\boldsymbol{Y}_{i,h})(t)][Y_{i+h}(s) - \hat{\mathbf{\Psi}}_{h,2}(\boldsymbol{Y}_{i,h})(s)].$$

Expanding the product in $\tilde{C}_{h,h}$ we obtain that

$$\sqrt{T}\tilde{C}_{h,h}(t,s) = \frac{1}{\sqrt{T}}\sum_{i=1}^{T} Y_i(t)Y_{i+h}(s) - \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\hat{\mathbf{\Psi}}_{h,1}(\boldsymbol{Y}_{i,h})(t)Y_{i+h}(s) \tag{A.4}$$

$$- \frac{1}{\sqrt{T}}\sum_{i=1}^{T} Y_i(t)\hat{\mathbf{\Psi}}_{h,2}(\boldsymbol{Y}_{i,h})(s) + \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\hat{\mathbf{\Psi}}_{h,1}(\boldsymbol{Y}_{i,h})(t)\hat{\mathbf{\Psi}}_{h,2}(\boldsymbol{Y}_{i,h})(s) \tag{A.5}$$

$$=: \sum_{i=1}^{4} A_{i,T}(t,s). \tag{A.6}$$

It follows from Theorem 2.1 of Kokoszka et al. (2017) that $A_{1,T}$ converges weakly in $L^2([0,1]^2)$ to a mean zero Gaussian element with covariance determined by (A.2), and hence the Theorem follows upon showing that $\|A_{i,T}\| = o_P(1)$ for $i = 2, 3$ and 4. We have according to the definition of $\hat{\mathbf{\Psi}}_{1,h}$ that

$$\begin{aligned}
-A_{2,T}(t,s) &= \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\hat{\mathbf{\Psi}}_{h,1}(\boldsymbol{Y}_{i,h})(t)Y_{i+h}(s) \\
&= \frac{1}{\sqrt{T}}\sum_{i=1}^{T}\sum_{j=1}^{h-1}\int \hat{\psi}_{h,1}^{(j)}(v,t)Y_{i+j}(v)Y_{i+h}(s)dv \\
&= \sum_{j=1}^{h-1}\int \hat{\psi}_{h,1}^{(j)}(v,t)\frac{1}{\sqrt{T}}\sum_{i=1}^{T} Y_{i+j}(v)Y_{i+h}(s)dv.
\end{aligned}$$

By the triangle and Cauchy-Schwarz inequalities it follows that

$$\|A_{2,T}\| \leq \sum_{j=1}^{h-1}\|\hat{\psi}_{h,1}^{(j)}\|\|\sqrt{T}\hat{C}_{j,h}^{*}\|, \tag{A.7}$$

where

$$\hat{C}^*_{j,h}(t,s) = \frac{1}{T}\sum_{i=1}^{T} Y_{i+j}(t)Y_{i+h}(s).$$

Since $\|\sqrt{T}\hat{C}^*_{j,h}\| = O_P(1)$, and $\|\hat{\psi}^{(j)}_{h,1}\| = o_P(1)$, it follows that $\|A_{2,T}\| = o_P(1)$. A similar argument may be used to establish that $\|A_{3,T}\| = o_P(1)$. Regarding $A_{4,T}$, we have again by applying the definitions of the operators $\hat{\mathbf{\Psi}}_{h,1}$ and $\hat{\mathbf{\Psi}}_{h,2}$ that

$$A_{4,T}(t,s) = \sum_{j_1=1}^{h-1}\sum_{j_2=1}^{h-1}\iint \hat{\psi}^{(j_1)}_{h,1}(v_1,t)\hat{\psi}^{(j_2)}_{h,2}(v_2,s)\sqrt{T}\hat{C}^*_{j_1,j_2}(v_1,v_2)dv_1dv_2.$$

Using the Cauchy-Schwarz inequality it follows that

$$\left(\iint\left|\iint \hat{\psi}^{(j_1)}_{h,1}(v_1,t)\hat{\psi}^{(j_2)}_{h,2}(v_2,s)\sqrt{T}\hat{C}^*_{j_1,j_2}(v_1,v_2)dv_1dv_2\right|^2 dtds\right)^{1/2} \leq \sqrt{T}\|\hat{\psi}^{(j_1)}_{h,1}\|\|\hat{\psi}^{(j_2)}_{h,2}\|\|\hat{C}^*_{j_1,j_2}\|.$$

By Theorem 4.7 of Bosq (2000), $\|\hat{C}^*_{j_1,j_2}\| = O_P(1)$ for all $1 \leq j_1, j_2 \leq h-1$. Since $\|\hat{\psi}^{(j_1)}_{h,1}\| = o_P(T^{-1/4})$ and $\|\hat{\psi}^{(j_2)}_{h,2}\| = o_P(T^{-1/4})$ by assumption, we have by the triangle inequality that

$$\|A_{4,T}\| \leq \sqrt{T}\sum_{j_1=1}^{h-1}\sum_{j_2=1}^{h-1}\sqrt{T}\|\hat{\psi}^{(j_1)}_{h,1}\|\|\hat{\psi}^{(j_2)}_{h,2}\|\|\hat{C}^*_{j_1,j_2}\| = \sqrt{T}o_P(T^{-1/2}) = o_P(1).$$

These bounds along with (A.3) and the weak convergence of $A_{1,T}$ establish (A.1). We now show that

$$\hat{\gamma}_{1,h} \xrightarrow{P} \left[\int C_0(u,u)du\right]^{1/2},$$

and the result for $\hat{\gamma}_{2,h}$ follows similarly. It follows from elementary arguments that

$$|\hat{\gamma}_{1,h} - \tilde{\gamma}_{1,h}| = o_P(1), \tag{A.8}$$

where

$$\tilde{\gamma}_{1,h} = \left(\int \frac{1}{T}\sum_{i=1}^{T}[Y_i(t) - \hat{\Psi}_{h,1}(\mathbf{Y}_{i,h})(t)]^2 dt\right)^{1/2}.$$

Expanding the square inside the integrand gives that

$$\tilde{\gamma}^2_{1,h} = \int \frac{1}{T}\sum_{i=1}^{T} Y_i^2(t)dt - 2\int \frac{1}{T}\sum_{i=1}^{T} Y_i(t)\hat{\Psi}_{h,1}(\boldsymbol{Y}_{i,h})(t)dt + \int \frac{1}{T}\sum_{i=1}^{T}[\hat{\Psi}_{h,1}(\boldsymbol{Y}_{i,h})(t)]^2 dt =: \sum_{j=1}^{3} B_{j,T}.$$

According to Corollary 4.1 of Bosq (2000), $B_{1,T} \xrightarrow{P} \int C_0(u,u)du$. It follows similarly to the proof that $\|A_{2,T}\| = o_P(1)$ and $\|A_{4,T}\| = o_P(1)$ that $B_{2,T} = o_P(1)$ and $B_{3,T} = o_P(1)$. Therefore $\tilde{\gamma}^2_{1,h} \xrightarrow{P} \int C_0(u,u)du$, which proves the result with (A.8).

□

Proof of Theorem 3. It follows again from elementary arguments that $\sqrt{T}\|\hat{C}_{h,h} - \tilde{C}_{h,h}\| = o_P(1)$.. In this case, since we suppose that $\{Y_i : i \in \mathbb{Z}\}$ is a causal solution to (17), we write

$$\begin{aligned}
\tilde{C}_{h,h}(t,s) &= \frac{1}{T}\sum_{i=1}^{T}[Y_i(t) - \hat{\boldsymbol{\Psi}}_{h,1}(\boldsymbol{Y}_{i,h})(t)][\epsilon_{i+h}(s) + \{\boldsymbol{\Psi}^{ARH}_{h,p} - \hat{\boldsymbol{\Psi}}_{h,2}\}(\boldsymbol{Y}_{i,h})(s)] \qquad \text{(A.9)}\\
&= \frac{1}{T}\sum_{i=1}^{T} Y_i(t)\epsilon_{i+h}(s) - \frac{1}{T}\sum_{i=1}^{T}\hat{\boldsymbol{\Psi}}_{h,1}(\boldsymbol{Y}_{i,h})(t)\epsilon_{i+h}(s)\\
&\quad - \frac{1}{T}\sum_{i=1}^{T} Y_i(t)\{\boldsymbol{\Psi}^{ARH}_{h,p} - \hat{\boldsymbol{\Psi}}_{h,2}\}(\boldsymbol{Y}_{i,h})(s) + \frac{1}{T}\sum_{i=1}^{T}\hat{\boldsymbol{\Psi}}_{h,1}(\boldsymbol{Y}_{i,h})(t)\{\boldsymbol{\Psi}^{ARH}_{h,p} - \hat{\boldsymbol{\Psi}}_{h,2}\}(\boldsymbol{Y}_{i,h})(s)\\
&=: \sum_{i=1}^{4} D_{i,T}(t,s).
\end{aligned}$$

It follows readily by the causality of $\{Y_i : i \in \mathbb{Z}\}$ that $EY_i(t)\epsilon_{i+h}(s) = 0$ for almost every $t,s \in [0,1]$, and that $Y_i \otimes \epsilon_{i+h}$ is an uncorrelated sequence in $L^2([0,1]^2)$. Hence by Chebyshev's inequality, $\|D_{1,T}\| = o_P(1)$. It follows similarly as the argument to establish the bound in (A.7) that

$$\|D_{2,T}\| \leq \|\hat{\boldsymbol{\Psi}}_{h,1}\|_{HS} \max_{1\leq j\leq h-1}\left(\int\int\left[\frac{1}{T}\sum_{i=1}^{T} Y_{i+j}(t)\epsilon_{i+h}(s)\right]^2 dtds\right)^{1/2} = o_P(1),$$

since once again for all $1 \leq j \leq h$, $\left\|\frac{1}{T}\sum_{i=1}^{T} Y_{i+j}\otimes\epsilon_{i+h}\right\| = o_P(1)$ and $\|\hat{\boldsymbol{\Psi}}_{h,1}\| = O_P(1)$ by assumption. In order to handle $D_{3,T}$, it follows from similar arguments as those used to establish (A.7) that

$$\|D_{3,T}\| \leq \|\boldsymbol{\Psi}^{ARH}_{h,p} - \hat{\boldsymbol{\Psi}}_{h,2}\|_{HS} \max_{1\leq j\leq h-1}\left(\int\int\left[\frac{1}{T}\sum_{i=1}^{T} Y_i(t)Y_{i+j}(s)\right]^2 dtds\right)^{1/2} = o_P(1),$$

since for a stationary and causal ARH(p) process with $E\|\epsilon_0\|^4 < \infty$, for each $1 \leq j \leq h-1$

$$\left(\int\int \left[\frac{1}{T} \sum_{i=1}^{T} Y_i(t) Y_{i+j}(s) \right]^2 dtds \right)^{1/2} = O_P(1),$$

see Theorem 4.8 of Bosq (2000). Similarly we have that

$$\|D_{4,T}\| \leq \|\hat{\mathbf{\Psi}}_{h,1}\|_{HS} \|\mathbf{\Psi}_{h,p}^{ARH} - \hat{\mathbf{\Psi}}_{h,2}\|_{HS} \max_{1 \leq j_1, j_2 \leq h-1} \left(\int\int \left[\frac{1}{T} \sum_{i=1}^{T} Y_{i+j_1}(t) Y_{i+j_2}(s) \right]^2 dtds \right)^{1/2} = o_P(1).$$

Therefore,

$$\|\tilde{C}_{h,h}\| = o_P(1).$$

It follows similarly that

$$\hat{\gamma}_{1,h}^2 \xrightarrow{P} \int E\{Y_0(t) - \mathbf{\Psi}_{h,1}(\mathbf{Y}_{0,h})(t)\}^2 dt, \;\; \hat{\gamma}_{2,h}^2 \xrightarrow{P} \int E\epsilon_0^2(t) dt.$$

By the assumption that $E\|\epsilon_0\|^4 > 0$, $\int E\epsilon_0^2(t)dt > 0$. The existence of $\mathbf{\Psi}_{h,1}$ so that $\int E\{Y_0(t) - \mathbf{\Psi}_{h,1}(\mathbf{Y}_{0,h})(t)\}^2 dt = 0$ contradicts $E\|\epsilon_0\|^4 > 0$, and so $\int E\{Y_0(t) - \mathbf{\Psi}_{h,1}(\mathbf{Y}_{0,h})(t)\}^2 dt > 0$ as well. It follows then that

$$\hat{\rho}_{h,h} = \frac{\hat{C}_{h,h}}{\hat{\gamma}_{1,h}\hat{\gamma}_{2,h}} \xrightarrow{P} 0,$$

completing the proof. □