Detecting the Complexity of a Functional Time Series

Enea G. Bongiorno, Lax Chan, Aldo Goia
Dipartimento di Studi per l'Economia e l'Impresa
Università del Piemonte Orientale

July 2023

The Version of Record of this manuscript has been published and is available in Journal of Nonparametric Statistics, July 2023:
http://www.tandfonline.com/10.1080/10485252.2023.2234507

Abstract

Consider a time series that takes values in a general topological space, and suppose that its Small-ball probability is factorized into two terms that play the role of a surrogate density and a volume term. The latter allows us to study the complexity of the underlying process. In some cases, the volume term can be analytically specified in a parametric form as a function of a complexity index. This work presents the study of an estimator for such an index whenever the volume term is monomial. Weak consistency and asymptotic Gaussianity are shown under an appropriate dependence structure, providing theoretical support for the construction of confidence intervals. A Monte Carlo simulation is performed to evaluate the performance of the approach under various conditions. Finally, the method is applied to identify the complexity of two real data sets.

Keywords: Small–ball Probability, β–mixing, Asymptotic Normality, Block–Jackknife
Subject classification: 62R10; 62G05

1 Introduction

The branch of statistics concerned with the development of methods for analyzing the so–called Functional data, i.e., discretized curves, surfaces, images, or other complex objects, is known as Functional Statistics. Functional data appear in various fields, such as medicine, engineering, economics, etc., and in the last two decades this discipline has experienced a strong growth thanks to the technological development of data collection devices and the increase in computing power. The interest of researchers in this field is evidenced by a conspicuous literature. For a general introduction to functional data analysis, interested readers can refer to the classic monographs by Ferraty and Vieu (2006); Horváth and Kokoszka (2012); Kokoszka and Reimherr (2017); Ramsay and Silverman (2005). For recent statistical methodologies, algorithms, and ad–hoc mathematical tools in (non)parametric functional setting, one can refer to the recent special issues Aneiros et al. (2019b), Aneiros et al. (2019a), Aneiros et al. (2022), book collections Aneiros et al. (2020), Aneiros et al. (2017) and references therein.

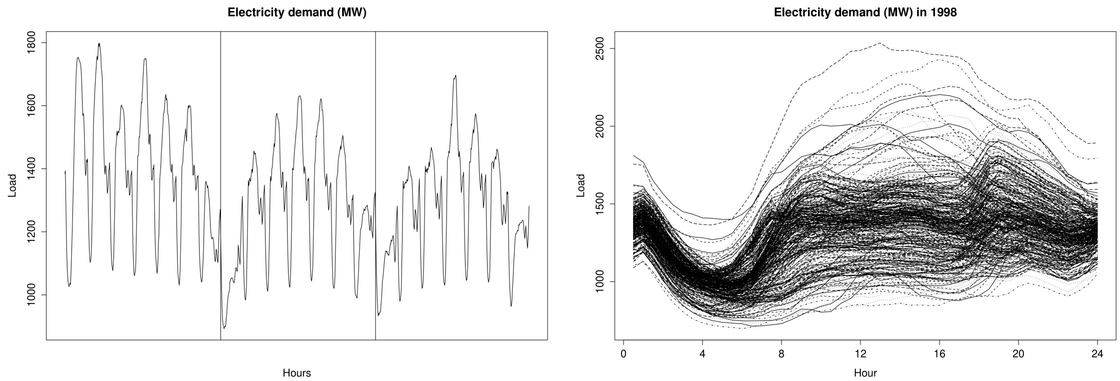

Figure 1: Left – Electricity demand in three randomly selected consecutive weeks in 1998. Right – Daily load curves obtained for 1998.

Among the many techniques that have been introduced in the field of Functional Statistics, those developed for the analysis of functional time series, i.e., sequences of functional random elements (such as, random curves) observed over time (e.g. Bosq, 2000 and Horváth and Kokoszka, 2012) play a very important role. These particular data can be obtained, for instance, by dividing a single real time series into successive segments, where each segment corresponds to a particular time interval, such as a day. In this way, some special recurring patterns (e.g., daily shapes) can be modelled as features of the space in which the random elements take values. Moreover, it is possible to compute an estimate of a probability density or a cumulative distribution, etc., for each segment, resulting in a functional time series of observed densities or related objects that can be considered as non–standard functional data.

As a first example, consider the time series of half-hourly electricity demand (in megawatts) recorded in Adelaide between January 1, 1998 and December 31, 2006 (see Magnano et al., 2008). An initial examination of the data reveals a periodic pattern within a day: This suggests that the original time series could be divided into functional observations, each coinciding with a particular daily load curve. The plots in Figure 1 represent the time series observed in three consecutive weeks, randomly selected from the original time series in 1998, and the resulting functional dataset for the same year.

A second example is the daily return densities for the S&P500 index over the period October 14, 2016 to May 6, 2017, with a frequency of 1 minute for a total of 140 market days: the resulting functional time series is the collection of daily densities of log–returns estimated using the standard Parzen–Rosenblatt approach. The original time series of S&P500 values and the resulting functional samples are shown in Figure 2.

Functional time series modelling was one of the first topics to be treated systematically both from a theoretical and a practical point of view, thanks to the monograph by Bosq (2000). This text focused on linear modelling (in particular, on autoregressive processes in Banach and Hilbert spaces) and laid a foundation for the subsequent literature. More recently, the book by Horváth and Kokoszka (2012) provided an update on the state of the art in this area. It is worth noting that in these references a second–order structure is always assumed for the underlying random element.

Beyond the linear world, one can study the functional time series in more general topological

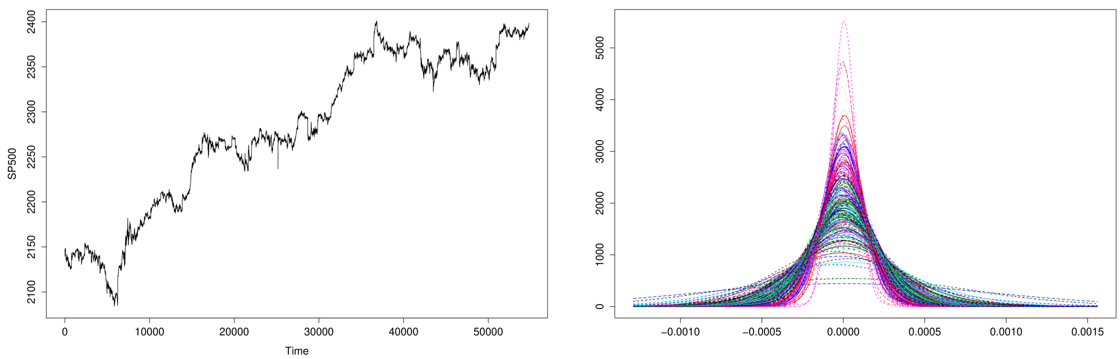

Figure 2: Left – S&P500 index from 14th October 2016, to 6th May 2017, with 1 minute frequency. Right – Daily density estimates of log–returns during the same period.

spaces using nonparametric approaches: in this framework, the notion of Small–Ball Probability (SmBP) emerges and plays a central role (see, e.g., Ferraty and Vieu, 2006). In general, given a random element X that takes values in $\mathcal{F}$, a set endowed with a (semi–)metric d, the SmBP is the asymptotic behaviour of the probability that X belongs to a ball centred at $\chi \in \mathcal{F}$ and whose radius h tends to zero:

$$\varphi_\chi(h) = \mathbb{P}\left(d\left(X, \chi\right) \leq h\right) \qquad h \longrightarrow 0.$$

A typical assumption is that $\varphi_\chi(h)$ factorizes into two terms:

$$\varphi_\chi(h) \sim \psi(\chi)\phi(h) \qquad h \longrightarrow 0 \tag{1}$$

with some identifiability constraints, such as $\mathbb{E}[\psi(X)] = 1$. Such a factorization allows to separate the contribution of the spatial term $\psi(\chi)$, which depends only on the centre of the sphere, and the volume term $\phi(h)$, which depends on the radius h. Interested readers can find some insights into the subject in Bogachev (1998); Bongiorno and Goia (2017); Li and Shao (2001) and the references therein.

As recently pointed out in Bongiorno et al. (2018), the behaviour of the function $\phi(h)$ when $h \longrightarrow 0$ carries information about the complexity of X without assuming a dominant measure on $\mathcal{F}$ or a second–order structure for X. For example, if $\phi(h)$ exhibits an exponential form $h^\gamma \exp\left(-c_{\gamma,\beta}/h^\beta\right)$ with $\beta, \gamma, c_{\gamma,\beta}$ positive constants, then X belongs to the so called *exponential family*, which includes many diffusion processes (see Lifshits, 2012; Li and Shao, 2001); while if $\phi(h)$ has a monomial form $c_\theta h^\theta$ with c_θ and θ positive constants, then X belongs to the so called *monomial family*, which includes the fractal processes of order θ (see Ferraty and Vieu, 2006, Definition 13.1) and the finite dimensional ones, with θ being a positive integer. In these cases, $\phi(h)$ has a parametric form identified by some real complexity indexes (namely γ, β, and θ). Thus, given the information contained in $\phi(h)$, to study the complexity represents a preliminary explorative step that could help in modelling the process.

This work deals with $\mathcal{F}$–valued stationary functional processes $(X_i)_{i\in\mathbb{N}}$, where it is assumed that each X_i belongs to the same family, and the goal is to evaluate the corresponding complexity indexes. Even though the methodology can be straightforwardly implemented to both families mentioned, only the monomial family is considered here because of its natural interpretation of the corresponding complexity index. For a better illustration and appreciation of this fact, consider the following examples. As a first example, let $\mathcal{F} = L^2_{[0,1]}$ be the space of square integrable real–valued functions defined over $[0, 1]$ with orthogonal basis $(\eta_1, \eta_2, \dots)$ and let $X_i = \mu + \sum_{j=1}^{q} Z_{ji}\eta_j$, where $\mu \in L^2_{[0,1]}$ and $\{Z_{1i}, \dots, Z_{qi}, i = 1, 2, \dots\}$ is a suitable q–dimensional real random process. In this case, taking d as the metric induced by the classical $L^2_{[0,1]}$ norm, the complexity index θ must be q. Another example considers a process X_i that takes values in $\mathcal{F}$, the subset of $L^2_{[0,1]}$ containing all the Gaussian densities. In this case, the parameters μ_i and σ_i for X_i are suitable real random processes and, taking d as in the previous example, $\theta = 2$. In both examples, the complexity index θ is the number of real random processes necessary to define X_i. Such an interpretation is similar to what happens in (regression) modelling where the concept of degrees of freedom is introduced to measure the complexity of specifications. Finally, note that the computation of $\varphi_\chi(h)$ might depend on the choice of d. In particular, if d is taken from a set of equivalent metrics, then the SmBP as well as θ do not change. On the other hand, if d is a semi–metric, then the value of the complexity index can be smaller than that obtained with a metric. This is the case in the first example, if there is a constant function among the elements of the basis η_j (which is assumed to be smooth) and d is the L^2 norm computed on the first derivatives X_i', then $\theta = q - 1$.

This paper discusses the evaluation of complexity from both theoretical and applied points of view at different levels. In particular, an estimate θ_n for the complexity index is derived by minimising a suitable dissimilarity measure between an empirical estimate ϕ_n (based on the U–statistic) for ϕ and the target complexity function $\phi_\theta(h) = c_\theta h^\theta$. The approach presented here does not require any second–order structure for the underlying process, unlike the usual techniques for determining dimensionality that, indeed rely on some covariance operators (see e.g., Bathia et al., 2010; Hall and Vial, 2006).

Some asymptotic properties for the estimator θ_n are analyzed: in particular, a weak law of large numbers and a central limit theorem are derived. For this purpose, among all possible dependence structures for the functional process $(X_i)_{i\in\mathbb{N}}$, the notion of $\mathcal{F}$–approximating functional on a stationary β–mixing sequence is assumed; see e.g., Doukhan (1994); Borovkova et al. (2001) for an overview of the various definitions of mixing and (Hörmann and Kokoszka, 2010, Section 2) for a comparison between the different notions of dependence for functional time series. Furthermore, the rates of convergence are derived in terms of the coefficients related to the dependence structure. This enriches the existing literature on the subject, which has so far focused only on the i.i.d. case (see Bongiorno et al., 2020). Finally, thanks to asymptotic normality, a confidence interval for θ can be written: to operationalise it, a suitable estimate of the standard error of θ_n is introduced based on a block–Jackknife approach, and then the coverage properties of this estimator are analysed using a Monte Carlo approach under different scenarios and experimental conditions. The procedure is then carried out for the functional time series of electricity demand presented above and for the log–returns densities of S&P500.

The article is structured as follows: in Section 2 the estimator θ_n is introduced, its asymptotic properties are given, and a confidence interval for θ is proposed; Section 3 shows numerical examples for functional time series of finite dimension and for parametric density families; Section 4 illustrates the methodology on two real datasets; finally, Section 6 and the supplementary materials collect all technical results and proofs.

2 Point and interval estimate for the complexity index

From now on, $\mathcal{F}$ is a set of real functions defined on a compact set $\mathcal{I}$, equipped with a semi–metric $d(\cdot,\cdot)$. Consider a stationary discrete–time random process $(X_i)_{i\in\mathbb{N}}$ where each X_i has the same distribution as the random curve X that takes values in $\mathcal{F}$, for which (1) holds with $\mathbb{E}[\psi(X)] = 1$, and $\phi(h) \sim c_\theta h^\theta$, as $h \longrightarrow 0$.

In order to estimate the complexity parameter θ from the observed sequence $(X_1, \ldots, X_n)$, one can minimise a dissimilarity measure between an estimate of the complexity function and the target $\phi_\theta(h) = c_\theta h^\theta$ with h in a suitable right neighbourhood of zero. To do this, consider the following empirical pointwise estimate of ϕ at the point $h \in \mathcal{H}$ (where $\mathcal{H} = [h_m, h_M]$, with $0 < h_m < h_M < +\infty$ are suitably chosen):

$$\phi_n(h) := \frac{1}{n(n-1)} \sum_{1\le i\ne j\le n} \mathbb{I}_{\{d(X_i,X_j)\le h\}}. \tag{2}$$

This is a second–order U–statistics that adapts the sample correlation integral (see Grassberger and Procaccia, 1983) to the functional setting.

For what concerns $\mathcal{H}$, its extremes must be chosen not too small to avoid that the indicator function in (2) does not vanish systematically. To do this, h_m is chosen so that

$$\mathbb{P}(D_{\min} < h_m) > 0, \tag{3}$$

where $D_{\min} = \min_{i\ne j} d(X_i, X_j)$ is the first–order statistics of the sample $D_{i,j} = d(X_i, X_j)$ ($1 \le i \ne j \le n$, with size $n(n-1) > 0$) of identically distributed random variables with common cdf F. Thanks to Markov inequality, it suffices to take $h_m > \mathbb{E}[D_{\min}]$ to obtain (3), and a direct application of (Rychlik, 1994, Corollary 3) leads to

$$h_m > n(n-1) \int_0^{b_n} y dF(y) \tag{4}$$

where $b_n = F^{-1}(1/(n(n-1)))$. Note that, when n diverges, b_n goes to zero and one can approximate the integral in (4) as $b_n F(b_n)$; thus, for n large enough it is sufficient to select h_m greater than the empirical quantile of order $1/(n(n-1))$ of $D_{i,j}$. Further practical details are provided in Section 3.

For what concerns the dissimilarity, a possible and effective option is the so-called cosine-dissimilarity

$$\Delta(\phi_n, \phi_\theta) = 1 - \frac{\langle \phi_n, \phi_\theta\rangle^2}{\|\phi_n\|^2 \|\phi_\theta\|^2},$$

where $\langle \phi_n, \phi_\theta\rangle = \int_{\mathcal{H}} \phi_n(h)\phi_\theta(h)\,dh$ and $\|\phi\|^2 = \langle\phi,\phi\rangle$. The resulting estimator θ_n of the complexity parameter θ is

$$\theta_n = \arg\min_{\theta\in\Theta} \Delta(\phi_n, \phi_\theta),$$

where Θ is a suitable compact subset of $(0, +\infty)$.

2.1 Some asymptotic results

Denote by θ_0 the true value of θ, and choose Θ such that $\theta_0 \in \Theta$. In this section, the weak consistency of θ_n to θ_0, and the asymptotic normality of $\sqrt{n}\,(\theta_n - \theta_0)$ are stated. To do this in a dependent framework, some special conditions on the nature of the stationary stochastic process $(X_i)_{i\in\mathbb{N}}$ are necessary, in particular this work deals with β–mixing sequences (see Definition 2.1) and with $\mathcal{F}$–approximating functional for $\mathcal{F}$–valued processes (see Definition 2.2) that generalizes 1–approximation functional for a real process (see Borovkova et al., 2001, Definition 1.4). The following definitions provide a feasible mathematical framework that allows one to derive an asymptotic theory by adapting existing techniques in the real–valued setting to the functional one.

Definition 2.1 *Let $(X_i)_{i\in\mathbb{N}}$ be a process taking values in $\mathcal{F}$, $\mathcal{M}_a^b$ be the σ-field generated by $\{X_i, a \leq i \leq b\}$ where $a \leq b \in \mathbb{N}$ and $(\beta_k)_{k\geq 0}$ be the sequence of β–mixing coefficients defined by*

$$\beta_k = \sup_{a\in\mathbb{N}} \mathbb{E}\left[\sup_{A\in\mathcal{M}_{a+k}^{\infty}} \mid \mathbb{P}\left(A|\mathcal{M}_1^a\right) - \mathbb{P}(A) \mid \right].$$

The process $(X_i)_{i\in\mathbb{N}}$ is called β–mixing if $\beta_k \longrightarrow 0$ as $k \longrightarrow +\infty$.

Definition 2.2 *The sequence $(X_i)_{i\in\mathbb{N}}$ is an $\mathcal{F}$-approximating functional on a process $(Z_i)_{i\in\mathbb{Z}}$ with approximation constants $(a_\ell)_{\ell\geq 0}$ if*

$$\mathbb{E}\left[d\left(X_1, \mathbb{E}\left[X_1|\mathcal{G}_{-\ell}^{\ell}\right]\right)\right] \leq a_\ell, \qquad \ell = 0, 1, 2, \ldots \tag{5}$$

where $\lim_{\ell\to\infty} a_\ell = 0$ and $\mathcal{G}_{-\ell}^{\ell}$ is a σ–algebra generated by $Z_{-\ell}, \ldots, Z_\ell$.

Roughly speaking, the $\mathcal{F}$-approximation functional condition states that the process $(X_i)_{i\in\mathbb{N}}$ approaches its mean given the dependency structure of finitely many Z_i (i.e. $\mathcal{G}_{-\ell}^{\ell}$), when ℓ diverges to infinity. In the following example, a $\mathcal{F}$-approximating functional on a real process is provided.

Example 2.3 *Let $X_i(t) = Z_i\eta(t)$, $t \in [0, 1]$, $i \in \mathbb{N}$, where $(Z_i)_{i\in\mathbb{N}}$ is a stationary real AR(1) and $\eta \in L^p$, $p \geq 1$, i.e. the space of p–integrable functions on $[0, 1]$. Then $X_1 = Z_1\eta$ is measurable with respect to $\mathcal{G}_0^{\ell}$ and hence, taking d as the L^p distance, Equation (5) holds for any positive sequence $(a_\ell)_{\ell\geq 0}$ decreasing to zero. The same holds whenever $X_i(t) = \sum_{j=1}^{q} Z_{ji}\eta_j(t)$, $t \in [0, 1]$ with $q \geq 1$ and Z_{ji} are real autoregressive processes independent on Z_{ki} with $k \neq j$ and $(\eta_j)_{j\in\mathbb{N}}$ is a basis of L^p.*

All the proposed asymptotic results are based on the following main assumptions. The first one models the dependency structure. In particular, among the different dependency structures that could be considered in the framework of functional time series (for a review see Hörmann and Kokoszka, 2010), the one chosen here provides a fairly general and convenient theoretical paradigm. The second assumption is related to the small–ball probability, which must be positive and factorisable, with a monomial volume term.

(A-1) $(X_i)_{i\in\mathbb{N}}$ is a $\mathcal{F}$–approximating functional on a stationary, β–mixing sequence with constants $(a_l)_{l\geq 0}$.

(A-2) X_i are distributed as X, a $\mathcal{F}$–valued random elements for which:

i. $\varphi_\chi(h) > 0$, for any $h > 0$ and χ;

ii. $\sup_\chi |\varphi_\chi(h)/\phi(h) - \psi(\chi)| \longrightarrow 0$, as $h \longrightarrow 0$;

iii. $\phi(h) = \phi_\theta(h) = c_\theta h^\theta$.

(A-3) Let $\delta_n(\theta) = 1 - \Delta(\phi_n, \phi_\theta)$ and $\delta(\theta) = 1 - \Delta(\phi_{\theta_0}, \phi_\theta)$. Assume that $\delta(\theta) \in C^2(\Theta)$ and strictly convex over the set Θ.

It is now possible to state the weak law of large numbers (WLLN) and the central limit theorem (CLT) for θ_n. To give an idea of the proofs, the first result is obtained as a consequence of the convergence in probability of $\delta_n(\theta)$ to its theoretical counterpart $\delta(\theta)$ uniformly over Θ; the second one is based on a delta–method approach.

Theorem 2.4 (WLLN) *Under assumptions (A-1) and (A-2), θ_n converges to θ_0 in probability, as $n \longrightarrow \infty$. If, in addition, (A-3) holds together with $\alpha_K = \left(2\sum_{k=K}^{\infty} a_k\right)^{1/2} < \infty$, then there exist two positive sequences of positive integers $L = L(n)$, $m = m(n)$ and a real valued function f such that*

$$|\theta_n - \theta_0| = O_p\left(\left(2\alpha_L + \beta_{m-2L} + f(2\alpha_L + \beta_{m-2L})\right)^{1/2}\right).$$

It is worth noting that the rate of convergence involves the coefficients a_k and β_k that characterize the dependency structure. For what concern L and m, only an implicit definition is available and, to avoid excessive technicalities at this stage, their natures and relationships are detailed in the proof of Lemma 6.1 in the supplementary materials.

Theorem 2.5 (CLT) *Under assumptions (A-1), (A-2) and (A-3) with $\beta_k = O(k^{-\beta})$ for $\beta \geq 8$ and $a_k = O(k^{-a})$ for $a = \max\{\beta + 3, 12\}$ and denoting by δ'' the second derivative of δ, then*

$$\sqrt{n}\frac{\|\phi_{\theta_0}\|^4 \delta''(\theta_0)}{2}(\theta_n - \theta_0)$$

is asymptotically distributed as a centered Gaussian random variable.

The proofs of theorems 2.4 and 2.5 are inspired by the theory of U–processes as done in Bongiorno et al. (2020) in the independent case, and to handle the dependency structure the proofs take advantage of results by (Borovkova et al., 2001, Theorem 6) and (Wendler, 2012, Theorem 1 and Corollary 1) respectively. Theorem 2.5 could be proved alternatively with arguments similar to those in (Borovkova et al., 2001, Theorem 9), but at the cost of a more elaborate set of assumptions such as the summability conditions on the β–mixing coefficients, the approximation constants a_k and additional constraints on the kernel of the U–statistic (2). By using the indicator function in the latter, the assumptions such as 1–continuity condition Borovkova et al. (2001) and the uniform variation Wendler (2012) could be straightforwardly verified.

To improve the readability of the paper and for the sake of completeness, all the proofs are deferred to Section 6 and the supplementary materials. Note that theorems 2.4 and 2.5 can be extended beyond the monomial family to include processes whose volume term exhibits an exponential form. In such a case, assumption (A-2) iii. can be replaced by $\phi(h) = \phi_\theta(h) = c_1 h^\alpha \exp\{-C_2/h^\beta\}$, where $\theta = (\alpha, \beta) \in [0, \infty) \times [0, \infty)$ and positive constants C_1, C_2 while the cosine dissimilarity measure involves the logarithm of the volume terms, see Bongiorno et al. (2020) for more details in the i.i.d. case.

2.2 Confidence intervals for the complexity index

This Section provides a way to operationalise the above theoretical results. In particular, thanks to Theorem 2.5, it is possible to write the following asymptotic confidence interval for θ at the level $1-\alpha$:

$$\theta_n \pm z_{1-\alpha/2}\ se\,(\theta_n)\,,$$

where $z_{1-\alpha/2}$ is the quantile of order $1-\alpha/2$ of a standard normal distribution and $se\,(\theta_n)$ is the standard error of θ_n.

Since an explicit form of the standard error of the estimator θ_n is not available, an estimate must be provided in order to implement the procedure. A possible strategy is to exploit resampling methods: among the existing alternatives for dependent data, the overlapping block–Jackknife method introduced by Liu and Singh (1992) is used here: the procedure is briefly recalled below and interested readers can find more details in the cited article.

Consider the blocks $B_j = (X_j, \dots, X_{j+\ell-1})$ of length ℓ, extracted from $\mathcal{X} = (X_1, \dots, X_n)$, where $j = 1, \dots, N$, $N = n - l + 1$ and $1 \leq \ell \leq n$ is a suitably selected parameter. For $j = 1, \dots, N$, define the j–th Jackknife block–deleted point value $\theta_n^{(j)}$, as the estimate of θ evaluated by removing the block B_j from $\mathcal{X}$. The block–Jackknife estimate of the variance θ_n then has the following form:

$$\widehat{Var}_{\mathrm{BJ}}(\theta_n) = \frac{\ell}{(n-\ell)N}\sum_{j=1}^{N}\left(\widetilde{\theta}_n^{(j)} - \theta_n\right)^2,$$

where $\widetilde{\theta}_n^{(j)} = (n\theta_n - (n-\ell)\theta_n^{(j)})/\ell$ is the j–th Jackknife pseudovalue of θ_n.

To assess the practical performances of the proposed confidence interval and to show how it can be used in practice, a numerical illustration of the method, through simulation studies, is provided in the next section.

3 Simulations

The proposed study shows how the introduced method works on simulated standard and non–standard functional time series. For different experimental settings (nature of the involved curves, sample size and complexity index θ), the mean and the estimated standard error for pointwise estimates θ_n over 1000 MC simulations are evaluated. For each replication, the confidence interval for θ at the nominal level 95% is computed, with the standard error estimated by the Jackknife approach presented above. Finally, the coverage of the latter is evaluated as the percentage of times in which the true dimension is covered by the CI. To complete the analysis, the coverage for dimensions around the true one are also computed. The examples are accompanied by a discussion of the tuning of the involved parameters, the influence of the choice of d, and the role of a possible presence of a noise term. The remainder of this section is divided into two parts, each devoted to a different simulation scenario for the functional time series: the finite dimensional and the parametric densities cases.

3.1 Finite dimensional process

The first experiment aims to illustrate and discuss the application of the methodology to functional time series that take values in finite dimensional linear spaces.

Settings. The functional data are generated according to the following model (see Example 2.3):

$$X_i(t) = \sum_{j=1}^{\theta_0} Z_{ji}\eta_j(t), \qquad t \in [0,1], i = 1,\ldots,n \tag{6}$$

where the basis functions are $\eta_j(u) = \sqrt{2}\cos(\pi j t)$ and, for each $j = 1,\ldots,\theta_0$, $(Z_{ji})_{i\in\mathbb{N}}$ is an autoregressive process of order 1 with coefficient $(-1)^j(0.9 - b/\theta_0)$ and $0 < b < 0.9$ independent on $(Z_{ki})_{i\in\mathbb{N}}$ with $k \neq j$. The value b controls the decay of variability with j of coefficients Z_{ji} in the θ_0–dimensional expansion (6): the smaller b is, the slower the decay is. Each curve is generated over a finite discrete mesh of 100 equispaced points in $[0,1]$.

The experiment is based on the combination of these settings:

- sample sizes $n = 100, 200, 300$ and 500;
- $b = 0.2, 0.5$ and 0.8
- $\theta_0 = 1, 2, 3, 4, 5$ and 6.

Choice of d, $\mathcal{H}$ and ℓ. In this simulation the distance involved in the SmBP is the classical L^2 metric:

$$d(f,g) = \sqrt{\int_0^1 (f(t) - g(t))^2\, dt},$$

where the integral is suitably approximated by summations.

To select the interval bounds of $\mathcal{H}$, one has to balance a negative bias for θ_n that arises when $\mathcal{H}$ is too close to zero, and its variability that increases when the length of $\mathcal{H}$ decreases. Some numerical experiments are performed and a good compromise seems to be achieved when the interval bounds of $\mathcal{H}$ are the empirical quantiles of order $40/\left(n^2 - n\right)$ and $440/\left(n^2 - n\right)$ of the dissimilarities $d(X_i, X_j)$ with $i \neq j$. In this way, there are at least 40 and at most 440 computed dissimilarities that contribute positively to the computation of (2).

Preliminary simulations suggest that the choice of the length ℓ of the blocks does not significantly affect the estimates, and it is chosen here in a heuristic way by $\ell = \lceil \sqrt[3]{n} \rceil$ (where $\lceil a \rceil$ denotes the smallest integer greater than a). In this way, ℓ ranges from 5 to 8 according to the used sample size.

Results and discussions. With the above settings, the results of the numerical experiments are computed and collected in tables 1, 2, and 3. Each table is divided into four parts, each one containing results for the different used sample sizes. For each dimension θ_0, the mean $\overline{\theta}_n$ and the standard deviation $std(\theta_n)$ of θ_n, computed over the 1000 MC replications, are reported. The last three columns of each table collect the estimated coverages of the confidence intervals for the dimensions $\theta_0 - 1$, θ_0 and $\theta_0 + 1$ ("x"denotes an unavailable data).

Reading the tables, it emerges that in all cases there is a negative bias that increases with θ_0 for a fixed sample size. The standard deviation of the estimates also behaves similarly. On the other hand, these behaviours are less pronounced as the sample size increases: the method performs very well for rather small dimensions and relatively small sample sizes, whereas it does not seem to produce excellent results when $\theta_0 \geq 5$. Regarding the coverage of the true complexity parameter θ_0, it can be seen that in general all confidence intervals are rather conservative. The presence of a

				Coverage		
n	θ_0	$\overline{\theta}_n$	$std\,(\theta_n)$	$\theta_0 - 1$	θ_0	$\theta_0 + 1$
100	1	1	0.075	x	0.99	0
	2	1.985	0.147	0	0.987	0.002
	3	2.88	0.212	0.131	0.974	0.043
	4	3.674	0.269	0.688	0.913	0.083
	5	4.37	0.328	0.967	0.786	0.079
	6	4.965	0.381	0.992	0.52	x
200	1	1.003	0.078	x	0.993	0
	2	1.999	0.149	0	0.995	0.003
	3	2.957	0.231	0.078	0.987	0.1
	4	3.853	0.288	0.521	0.975	0.231
	5	4.665	0.356	0.872	0.955	0.257
	6	5.414	0.414	0.973	0.883	x
300	1	1.006	0.077	x	0.997	0
	2	2.001	0.148	0.001	0.995	0.006
	3	2.988	0.222	0.056	0.997	0.126
	4	3.915	0.299	0.45	0.986	0.282
	5	4.773	0.349	0.851	0.978	0.353
	6	5.558	0.418	0.96	0.949	x
500	1	1.003	0.078	x	0.997	0
	2	2.005	0.151	0	0.994	0.004
	3	2.987	0.226	0.065	0.991	0.127
	4	3.951	0.295	0.454	0.995	0.342
	5	4.85	0.366	0.789	0.987	0.451
	6	5.715	0.431	0.944	0.972	x

Table 1: Case $b = 0.2$. Estimated mean, standard deviation of θ_n over 1000 MC replications and coverage of confidence intervals at level 95% for θ_0 and $\theta_0 \pm 1$.

negative bias, the high variability of the estimates for $\theta_0 \geq 5$ and n not being large enough produce intervals that systematically also cover the dimension $\theta_0 - 1$, but not the dimension $\theta_0 + 1$, leading to a parsimonious modelling of the process. Finally, note that the value of b has no significant effect on the pointwise and interval estimates.

In order to evaluate the impact of choice of d, the above experiment is performed again with a fixed setting ($n = 200$ and $b = 0.5$) and varying d among the following:

- L^2: see above;
- L^1: $d\,(f, g) = \int_{[0,1]} |f\,(t) - g\,(t)|\,dt$;
- L^∞: $d\,(f, g) = \sup_{t \in [0,1]} |f\,(t) - g\,(t)|$;
- H: $d\,(f, g) = \sqrt{\int_{[0,1]} \left(f'\,(t) - g'\,(t)\right)^2 dt}$, where f' denotes the first derivative of f.

				Coverage		
n	θ_0	$\overline{\theta}_n$	$std\,(\theta_n)$	$\theta_0 - 1$	θ_0	$\theta_0 + 1$
100	1	1.002	0.076	x	0.994	0
	2	1.980	0.153	0	0.991	0.004
	3	2.872	0.216	0.114	0.978	0.04
	4	3.684	0.27	0.697	0.93	0.079
	5	4.43	0.334	0.959	0.83	0.089
	6	5.111	0.38	0.997	0.672	x
200	1	1.004	0.075	x	0.996	0
	2	1.997	0.151	0	0.993	0.002
	3	2.96	0.221	0.064	0.989	0.094
	4	3.859	0.285	0.501	0.983	0.207
	5	4.682	0.356	0.884	0.949	0.273
	6	5.471	0.398	0.981	0.911	x
300	1	1.005	0.076	x	0.997	0
	2	1.996	0.154	0	0.992	0.003
	3	2.978	0.227	0.053	0.993	0.115
	4	3.905	0.284	0.459	0.993	0.265
	5	4.765	0.349	0.824	0.975	0.354
	6	5.604	0.407	0.967	0.953	x
500	1	1.005	0.073	x	0.996	0
	2	2.009	0.148	0	0.997	0.006
	3	2.989	0.230	0.06	0.996	0.128
	4	3.95	0.287	0.42	0.994	0.337
	5	4.85	0.361	0.789	0.992	0.443
	6	5.742	0.428	0.942	0.98	x

Table 2: Case $b = 0.5$. Estimated mean, standard deviation of θ_n over 1000 MC replications and coverage of confidence intervals at level 95% for θ_0 and $\theta_0 \pm 1$.

				Coverage		
n	θ_0	$\overline{\theta}_n$	$std\,(\theta_n)$	$\theta_0 - 1$	θ_0	$\theta_0 + 1$
100	1	1.006	0.074	x	0.992	0
	2	1.981	0.142	0	0.992	0.001
	3	2.892	0.221	0.099	0.977	0.052
	4	3.689	0.263	0.683	0.933	0.066
	5	4.456	0.339	0.934	0.828	0.1
	6	5.103	0.382	0.993	0.653	x
200	1	1.004	0.076	x	0.994	0
	2	2.006	0.15	0	0.995	0.005
	3	2.96	0.223	0.064	0.993	0.109
	4	3.857	0.287	0.523	0.981	0.21
	5	4.683	0.345	0.891	0.956	0.247
	6	5.489	0.401	0.987	0.928	x
300	1	1	0.077	x	0.996	0
	2	2	0.152	0	0.993	0.004
	3	2.989	0.236	0.061	0.995	0.128
	4	3.905	0.29	0.443	0.986	0.269
	5	4.772	0.356	0.837	0.978	0.346
	6	5.618	0.42	0.961	0.956	x
500	1	1.009	0.075	x	0.996	0
	2	2.006	0.154	0	0.996	0.01
	3	2.986	0.223	0.059	0.995	0.128
	4	3.939	0.289	0.44	0.989	0.318
	5	4.858	0.367	0.789	0.985	0.456
	6	5.722	0.436	0.945	0.979	x

Table 3: Case $b = 0.8$. Estimated mean, standard deviation of θ_n over 1000 MC replications and coverage of confidence intervals at level 95% for θ_0 and $\theta_0 \pm 1$.

				Coverage		
n	θ_0	$\bar{\theta}_n$	$std\,(\theta_n)$	$\theta_0 - 1$	θ_0	$\theta_0 + 1$
L^2	1	1.004	0.075	x	0.996	0
	2	1.997	0.151	0	0.993	0.002
	3	2.96	0.221	0.064	0.989	0.094
	4	3.859	0.285	0.501	0.983	0.207
	5	4.682	0.356	0.884	0.949	0.273
	6	5.471	0.398	0.981	0.911	x
L^1	1	1.004	0.075	x	0.996	0
	2	1.996	0.15	0	0.995	0.003
	3	2.957	0.223	0.071	0.99	0.086
	4	3.855	0.287	0.506	0.984	0.214
	5	4.672	0.352	0.881	0.949	0.254
	6	5.471	0.4	0.981	0.911	x
L^∞	1	1.004	0.075	x	0.996	0
	2	2	0.151	0	0.992	0.007
	3	2.958	0.22	0.053	0.993	0.099
	4	3.844	0.272	0.53	0.984	0.185
	5	4.646	0.348	0.905	0.947	0.238
	6	5.419	0.397	0.984	0.887	x
H	1	1.004	0.075	x	0.996	0
	2	2.002	0.151	0	0.99	0.003
	3	2.948	0.217	0.058	0.992	0.092
	4	3.823	0.274	0.55	0.98	0.178
	5	4.572	0.345	0.928	0.108	0.170
	6	5.268	0.385	0.993	0.812	x

Table 4: Case $n = 200$, $b = 0.5$. Estimated mean, standard deviation of θ_n over 1000 MC replications and coverage of confidence intervals at level 95% for θ_0 and $\theta_0 \pm 1$.

Table 4 summarises the obtained results. It arises that the choice of semi–metric has no effect on the estimates of the complexity index. This is because when the process is in a finite dimensional space, the L^1, L^2 and L^∞ norms are equivalent. Moreover, since in this controlled simulation the used basis (η_j) does not include a constant, the semi–metric H also yields similar results for the estimation of θ_0.

			Coverage		
θ_0	$\bar{\theta}_n$	$std\,(\theta_n)$	$\theta_0 - 1$	θ_0	$\theta_0 + 1$
1	1	0.072	x	0.99	0
2	2	0.152	0.001	0.994	0.006
3	2.93	0.216	0.084	0.99	0.076
4	3.76	0.287	0.646	0.957	0.151
5	4.51	0.327	0.952	0.892	0.148
6	5.17	0.393	0.997	0.721	x

Table 5: Case: trajectories without second–orderstructure, $n = 200$, $b = 0.5$. Estimated mean, standard deviation of θ_n over 1000 MC replications and coverage of confidence intervals at level 95% for θ_0 and $\theta_0 \pm 1$.

As mentioned in the Introduction, the proposed procedure is free of assumptions about the second–order structure of the process. To illustrate this via numerical experiment, 1000 MC replications of the previous setting are performed, where here the innovation of the autoregressive process Z_{ki} is a t–Student with two degrees of freedom and the following parameters are adopted $\theta = 1, \ldots, 6$, $b = 0.5$, $n = 200$ and $d = L^1$. The chosen innovation and the L^1–metric guarantee that process in (6) does not possess a second–order structure, but still satisfies the assumptions introduced in Section 2.1; in particular, the combination of the $t(2)$ and the L^1–metric guarantees the existence of the conditional mean in the $\mathcal{F}$–approximating functional (see Definition 2.2). The results are collected in Table 5, which shows similar behaviour to the previous tables in terms of negative bias, its relation to θ_0 and coverage.

Another important problem is to evaluate the sensitivity of the method when the trajectories are affected by noise. To do this, consider the model

$$X_i\,(t) = Y_i\,(t) + \sigma \mathcal{E}_i\,(t) \qquad t \in [0, 1]\,, i = 1, \ldots, n$$

where $Y_i\,(t)$ are genereted according to (6) with $b = 0.5$, $\theta_0 = 1, 2, 3, 4, 5$ and 6, $n = 200$, and $\mathcal{E}_i\,(t)$ is a noise process independent from $Y_i\,(t)$. In this study, a Standard Brownian Motion, a process belonging to the exponential family, is used and $\sigma = 0.01, 0.05, 0.1$ is chosen to determine its impact. This produces functional data which are elements of a more complex family than the monomial one; this leads to estimates larger than θ_0, and this effect is particularly emphasized when the complexity of Ys is low (θ_0 is small) and σ rather large. This trade–off is corroborated by the results reported in Table 6.

3.2 Parametric densities

Aside from the standard functional time series in linear spaces, one can also consider non–standard objects: here, sequences of probability densities are used. In particular, let $\phi\,(t, \mu, \sigma)$, $t \in \mathbb{R}$, be the density of a Gaussian distribution with parameters μ and σ, the data are generated according to the following mixture model:

$$X_i(t) = w_i \phi\,(t, \mu_{1,i}, \sigma_{1,i}) + (1 - w_i)\,\phi\,(t, \mu_{2,i}, \sigma_{2,i})\,, \qquad i = 1, \ldots .n, \tag{7}$$

discretized over a suitable mesh of 100 equispaced points (t_j), selected so that $\int_{t_1}^{t_{100}} X_i\,(t)\,dt \geq 0.99$ for any i: roughly speaking, all the densities can be represented on the same interval without

				Coverage		
σ	θ_0	$\overline{\theta}_n$	$std\,(\theta_n)$	$\theta_0 - 1$	θ_0	$\theta_0 + 1$
0.01	1	1.532	0.113	x	0.029	0.201
	2	2.001	0.152	0	0.996	0.003
	3	2.961	0.219	0.071	0.993	0.094
	4	3.860	0.285	0.494	0.981	0.213
	5	4.682	0.356	0.875	0.95	0.256
	6	5.471	0.398	0.979	0.916	x
0.05	1	2.626	0.196	x	0	0.339
	2	2.125	0.161	0	0.989	0.025
	3	2.985	0.219	0.047	0.996	0.114
	4	3.870	0.286	0.483	0.983	0.208
	5	4.687	0.357	0.872	0.95	0.264
	6	5.475	0.396	0.979	0.917	x
0.1	1	3.306	0.242	x	0	0
	2	2.411	0.186	0	0.761	0.379
	3	3.054	0.222	0.024	0.994	0.172
	4	3.899	0.289	0.454	0.988	0.237
	5	4.704	0.357	0.877	0.958	0.292
	6	5.487	0.396	0.977	0.926	x

Table 6: Case: Trajectories affected by noise, $n = 200$, $b = 0.5$. Estimated mean, standard deviation of θ_n over 1000 MC replications and coverage of confidence intervals at level 95% for θ_0 and $\theta_0 \pm 1$.

substantially truncating the tails. In such a case, the complexity is driven by the number of (non–constant) random parameters involved in the expression (7).

In particular, the following scenarios are investigated:

A. only the first addend in (7) is considered ($w_i = 1$ for any i), μ_{1i} is generated according to an $AR\,(1)$ model with coefficient 0.8 on $[0, 1]$, whereas $\sigma_{1i} = 1$, for any i;

B. $w_i = 1$ and $\mu_{1i} = 0$ for any i, whereas σ_{1i} is generated according to a Geometric Brownian Motion (GBM), that is $\sigma_{1i} = \exp\{B\,(i)\}$, where $B\,(i)$ is a standard Brownian motion on $[0, 1]$;

C. $w_i = 1$ and $\mu_{1i} \sim AR(1)$, $\sigma_{1i} \sim GBM$;

D. w_i is uniformly distributed over the interval $[0.4, 0.6]$, $\mu_{1i} = 2$ and $\mu_{2i} = -2$ for any i, whereas σ_{1i} and σ_{2i} are two independent GBMs.

It is worth noticing that for scenarios A. and B. the complexity parameter is $\theta_0 = 1$, for the scenario C. it equals 2, while $\theta_0 = 3$ for the scenario D.

Taking d as the classical L^2–norm and $n = 100, 200, 300$ and 500, each scenario is replicated 1000 times and, as described at the beginning of Section 3, the means, the estimated standard errors for pointwise estimates θ_n and the coverages of the confidence intervals are reported. The results are collected in Table 7.

					Coverage		
n	Scenarios	θ_0	$\overline{\theta}_n$	$std\,(\theta_n)$	$\theta_0 - 1$	θ_0	$\theta_0 + 1$
100	A	1	1.003	0.078	x	0.994	0
	B	1	0.999	0.075	x	0.989	0
	C	2	1.831	0.159	0.035	0.941	0
	D	3	2.529	0.223	0.576	0.633	0.001
200	A	1	1.006	0.081	x	0.997	0
	B	1	1.005	0.078	x	0.996	0
	C	2	1.922	0.16	0.009	0.972	0.008
	D	3	2.704	0.219	0.29	0.858	0.017
300	A	1	1.001	0.076	x	0.997	0
	B	1	1.006	0.076	x	0.997	0
	C	2	1.954	0.15	0.005	0.987	0.002
	D	3	2.777	0.225	0.191	0.927	0.029
500	A	1	1.002	0.076	x	0.997	0
	B	1	1.004	0.078	x	0.998	0
	C	2	1.977	0.152	0	0.992	0.007
	D	3	2.824	0.217	0.137	0.948	0.03

Table 7: Estimated mean, standard deviation of θ_n over 1000 MC replications and coverage of confidence intervals at level 95% of different scenarios.

From the table, it emerges that when the complexity is low, the method produces good results even with small sample sizes. As the complexity increases, the method tends to underestimate for small sample sizes (similar to the finite dimensional setting, see Section 3.1): as sample sizes become larger, the negative bias becomes less pronounced.

To conclude this section, it is useful to remark that when data are densities, it is common to consider d as the L^2–norm between the centered log–ratio transformations (see van den Boogaart et al., 2014). From this point of view, a simulation study has been carried out using this particular distance in the same experimental scenarios as before: the results obtained are very similar to those in Table 7 and are therefore omitted.

4 Applications to real data

This section illustrates an application of the proposed methodology to the real datasets presented in the Introduction, that is the time series of load curves and log–returns densities.

4.1 Load Curves

The original dataset, available in the R package *fds*, consists of time series of half-hourly electricity demand (in megawatts). The dataset is split into functional observations, each corresponding to a particular daily load curve; the resulting functional dataset consists of 3287 observed daily curves discretized over a mesh of 48 equispaced points

$$\mathbb{X} = \{X_i(t_j),\ i = 1, \dots, 3287,\ j = 1, \dots, 48\}.$$

For practical purposes, data are divided into 9–year datasets due to a long–term trend and variation between years:

$$\mathbb{X}_y = \{X_d^y(t_j),\ d = 1, \dots, 365 \text{ (or 366)},\ j = 1, \dots, 48\}, \qquad y = 1998, \dots, 2006.$$

As an instance, the daily load curves for the year 1998 are plotted in the right panel of Figure 1.

The illustrated technique is implemented on each annual dataset using the classical L^2–norm as the metric. First, the empirical estimates $\phi_n(h)$ are computed over $\mathcal{H}$ (selected according to the rule provided in Section 3). The corresponding *Log–volugrams*, i.e., the plots of $\log h$ against $\log \phi_n(h)$ over $\mathcal{H}$ (see Bongiorno et al., 2018), are visualized in the left panel of Figure 3: the exhibited linearity suggests that the observed processes belong to the monomial family. In this view, the cosine dissimilarities Δ between ϕ_n and the target $\phi_\theta(h) = c_\theta h^\theta$ are calculated and depicted for each year in the right panel of Figure 3. These functions, approximated over a fine equispaced grid discretizing $\Theta = [4, 8]$, appear convex, and their minima provide the estimates of the unknown dimension θ_0, whose values are collected in the second column of Table 8. To complete the analysis, the corresponding block–Jackknife standard error estimates with block length $\ell = 8$ are computed and used to build the 95% confidence interval for θ_0, see the last columns of Table 8.

Interpreting this result in the context of finite dimensional representations of functions in linear spaces, and taking into account the fact that the confidence intervals tend to systematically cover dimensions lower than θ_0 (see Section 3.1), it follows that it may be possible to model the observed data for each year using representations based on 6 or 7 random components. Nevertheless, this information does not indicate what kind of basis can be reasonably used or what type of random processes are actually involved: given the exploratory nature of the approach, these latter aspects require further investigation.

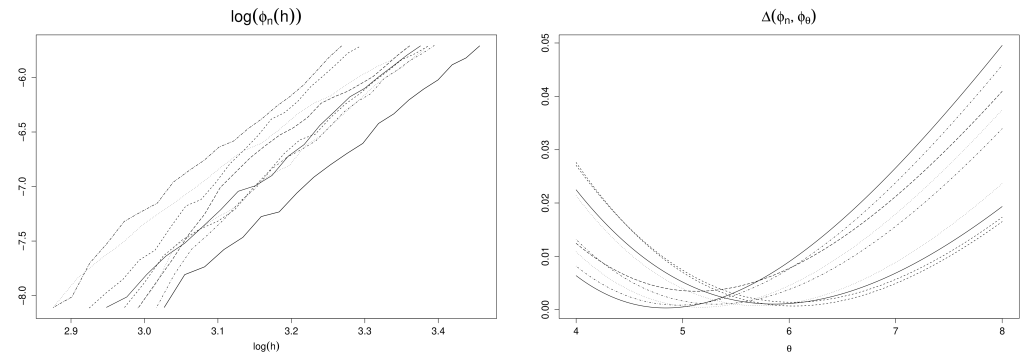

Figure 3: Left – Log–Volugrams for each year. Right – the behavior of dissimilarities $\Delta(\phi_n, \phi_\theta)$ over Θ for each year.

Year	θ_n	$se_J(\theta_n)$	CI 95%	
1998	4.84	0.52	3.83	5.86
1999	6.03	0.83	4.40	7.66
2000	5.17	0.50	4.19	6.16
2001	5.31	0.64	4.04	6.57
2002	5.13	0.49	4.17	6.09
2003	5.86	0.89	4.11	7.60
2004	6.04	0.69	4.69	7.40
2005	5.72	0.58	4.57	6.86
2006	4.96	0.57	3.84	6.07

Table 8: Estimates for θ_0, their standard errors and 95% confidence intervals for each year.

4.2 Log–returns densities

The second example deals with the detection of complexity of probability densities of intradaily log–returns. In particular, the *S&P500* index is considered during the period from October 14 2016 to May 6 2017: data are collected at a frequency of 1 minute for a total of 54810 observations during 140 market days.

The time series of log–returns are calculated from the original data and then the probability densities of log–returns are estimated using the standard kernel approach. Thanks to the large amount of available data, it is possible to split the set of data referring to a given day into two disjoint parts to obtain a functional time series of 280 densities:

$$X_i(r) = \frac{1}{n_i h_i} \sum_{t=1}^{n_i} K\left(\frac{R_{i,t} - r}{h_i}\right) \qquad i = 1, \ldots, 280, \quad r \in \mathcal{R}$$

where $R_{i,t}$ is the log–return at minute t–th during the i–th half day, n_i is the number of observations during that half day, K is the Gaussian kernel, h_i is a bandwidth and $\mathcal{R}$ is a common support for all the densities that, due to the presence of outliers, it has been fixed as $\mathcal{R} = [q_{0.001}, q_{0.999}]$, where q_β stands for the quantile of order β of the observed log–returns, and discretized in 200 equispaced points. A delicate issue is the selection of the bandwidths, on which depends the degree of smoothness of each functional observation, that affects the estimation of complexity; indeed, the smoother the curves are, the less the influence of noise is. The solution adopted is to apply a scale factor to Silverman's classical rule–of–thumb, i.e., $\widetilde{h}_i = 1.06\widehat{\sigma}_i n_i^{-1/5}$, where $\widehat{\sigma}_i$ is the estimated standard deviation for the log–returns in the i–th half day. Here, h_i is set as 1, 1.5 and 2 times $\widetilde{h}_i$.

The application of the methodology performed in a similar manner as in Section 4.1, using the classical L^2–norm, $\Theta = [3, 8]$ discretized over a fine mesh of equispaced points, and $\ell = 7$ for the block–Jackknife step, produces the graphs in Figure 4, where the *Log–volugrams* and the cosine dissimilarities varying the scale factor are plotted, and the numerical results in Table 9.

As a general comment, all the results confirm the intuition regarding the role of the bandwidth; on the other hand, the variability of the estimate seems to decrease as the scale factor increases. It also emerges that a simple specification with 2 parameters (position and variability) is not sufficient to describe the complexity of the distributions of log–returns and that at least 4 random coefficients must be introduced. This is coherent with the financial literature, where the Gaussian distribution alone is considered inadequate to model the phenomenon preferring some finite mixture of densities, which are able to reproduce the leptokurtic nature of the data and some heteroskedasticity effects over time (see Fusai and Roncoroni, 2007, Chapter 12).

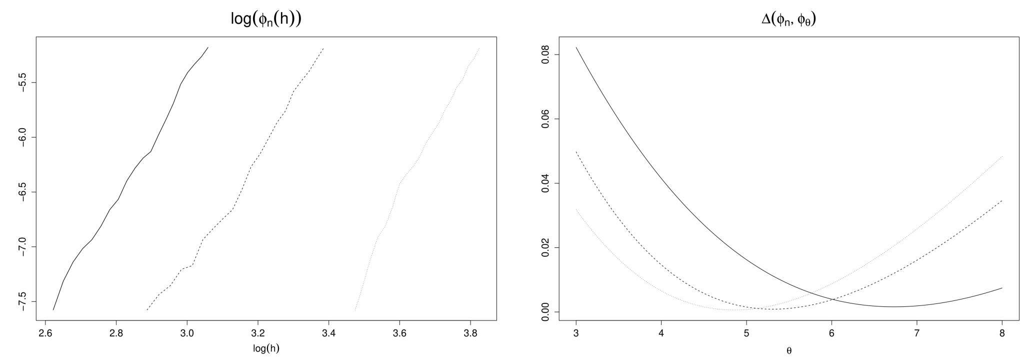

Figure 4: Left – Log–Volugrams for each scale factor (solid: 1, dashed: 1.5, dotted: 2). Right – the behavior of dissimilarities $\Delta(\phi_n, \phi_\theta)$ over Θ for scale factor (solid: 1, dashed: 1.5, dotted: 2).

Scale	θ_n	$se_J(\theta_n)$	CI 95%	
1	6.73	0.93	4.92	8.54
1.5	5.31	0.68	3.97	6.65
2	4.86	0.47	3.94	5.77

Table 9: Estimates for θ_0, their standard errors and 95% confidence intervals for each selected scale factor.

5 Concludings

The methodology presented in this work provides a nonparametric exploratory technique that leads to measure the complexity index of functional time series. The study was carried out from both practical and theoretical points of view with the aim of enriching the existing literature on the topic, which has so far was limited to the i.i.d. case (see Bongiorno et al., 2018, 2019, 2020). In particular, to frame the study in a convenient mathematical framework, the dependency concept of the $\mathcal{F}$–approximating functional on a β–mixing sequence is adopted. The present work focuses on the monomial family and shows that the complexity index could be interpreted as the number of independent univariate random components underlying the process, and that this number could also coincide with the dimension of the subspace spanned by the process. From this perspective, complexity plays a role analogous to the concept of degrees of freedom in probabilistic and statistical modeling. The considered procedure is free of assumptions about an underlying dominant measure and about the existence of a second–order structure for the process. In addition, consistency is derived for the estimator of the complexity index, and although it is not made explicit, the rate of convergence yields a direct relationship with the dependency structure in terms of beta–mixing coefficients and approximation constants.

The present work leads to some interesting practical and theoretical issues that require further extensive studies. First, since a systematic bias appears from the simulations, it is interesting to quantify it theoretically in terms of h and to study the mean squared error. In this context, a second possible research direction is to establish a theoretical justification for the selection of the interval $\mathcal{H}$ in which ϕ_n is computed. Finally, note that in the monomial case the complexity index can theoretically assume any real positive value. To the best of the authors' knowledge, only processes with integer complexity index are available. Thus, one may wonder whether it is possible to theoretically define a process in the monomial family whose complexity index is not integer, analogous to what is done in the literature of the degrees of freedom that can take non–integer values.

6 Technical results

This section contains all the justifications for the theoretical results stated in Section 2.1. Many of the techniques used here are inspired by results presented in previous works of Bongiorno et al. (2020) but are appropriately adapted to the specific used setting.

6.1 Proof of Theorem 2.4

The proof of the consistency is divided into two steps: firstly the convergence in probability of δ_n to δ uniformly over Θ is proven; secondly, the thesis is provided as a consequence of the convergence of $\delta(\theta_n)$ to $\delta(\theta_0)$.

Consider $\widetilde{\phi}_\gamma = \phi_\gamma / \|\phi_\gamma\|$, the normalised version of ϕ_γ with γ being one among θ, θ_0, n, that is wellposed since $\phi_\theta, \phi_{\theta_0} > 0$ over $(0, +\infty)$ and $\phi_n > 0$ over $\mathcal{H}$ by construction (see Section 2). Note that by definition $\delta(\theta) = \langle \widetilde{\phi}_{\theta_0}, \widetilde{\phi}_\theta \rangle^2$ and $\delta_n(\theta) = \langle \widetilde{\phi}_n, \widetilde{\phi}_\theta \rangle^2$ which are continuous functions over Θ. Consider

$$|\delta_n(\theta) - \delta(\theta)| = \left| \langle \widetilde{\phi}_\theta, \widetilde{\phi}_{\theta_0} \rangle^2 - \langle \widetilde{\phi}_\theta, \widetilde{\phi}_n \rangle^2 \right|. \tag{8}$$

Thanks to algebraic arguments and invoking the Cauchy–Schwarz inequality twice together with the fact that $\widetilde{\phi}$'s are normalized, one obtains for all $\theta \in \Theta$:

$$\begin{aligned} \left|\langle\widetilde{\phi}_\theta, \widetilde{\phi}_{\theta_0}\rangle^2 - \langle\widetilde{\phi}_\theta, \widetilde{\phi}_n\rangle^2\right| &\le 2\left|\langle\widetilde{\phi}_\theta, \widetilde{\phi}_{\theta_0}\rangle - \langle\widetilde{\phi}_\theta, \widetilde{\phi}_n\rangle\right| \\ &= 2\left|\langle\widetilde{\phi}_\theta, \widetilde{\phi}_{\theta_0} - \widetilde{\phi}_n\rangle\right| \le 2\left\|\widetilde{\phi}_{\theta_0} - \widetilde{\phi}_n\right\|. \end{aligned} \tag{9}$$

Denoting by C a general positive constant, note that

$$\begin{aligned} \left\|\widetilde{\phi}_{\theta_0} - \widetilde{\phi}_n\right\| &\le \left\|\frac{\phi_{\theta_0}}{\|\phi_{\theta_0}\|} - \frac{\phi_n}{\|\phi_n\|}\right\| = \left\|\frac{\phi_{\theta_0}\|\phi_n\| - \phi_n\|\phi_{\theta_0}\|}{\|\phi_{\theta_0}\|\,\|\phi_n\|}\right\| \\ &\le C\,\|\ \phi_{\theta_0}\|\phi_n\| - \phi_n\|\phi_{\theta_0}\|\| \\ &= C\,\|\phi_{\theta_0}\|\phi_n\| - \phi_n\|\phi_n\| + \phi_n\|\phi_n\| - \phi_n\|\phi_{\theta_0}\|\| \\ &= C\,\|(\phi_{\theta_0} - \phi_n)\|\phi_n\| + \phi_n(\|\phi_n\| - \|\phi_{\theta_0}\|)\| \\ &\le C\,\|\phi_n\|\,\|\phi_{\theta_0} - \phi_n\| + C\,\|\phi_n\|\ \ |\ \|\phi_n\| - \|\phi_{\theta_0}\|\ \ | \end{aligned}$$

and, by reverse triangle inequality $|\ \ \|\phi_n\| - \|\phi_{\theta_0}\|\ \ | \le \|\phi_{\theta_0} - \phi_n\|$, it holds

$$\left\|\widetilde{\phi}_{\theta_0} - \widetilde{\phi}_n\right\| \le 2C\,\|\phi_n\|\,\|\phi_{\theta_0} - \phi_n\| \le C\,\|\phi_{\theta_0} - \phi_n\|.$$

Then

$$\left\|\widetilde{\phi}_{\theta_0} - \widetilde{\phi}_n\right\| \le C\,|\mathcal{H}| \sup_{h\in\mathcal{H}} |\ \phi_n(h) - \phi_{\theta_0}(h)|,$$

where $|\mathcal{H}|$ is the length of $\mathcal{H}$. Combining (8), (9) and using the fact that ϕ_n converges in probability to ϕ_{θ_0} uniformly on $\mathcal{H}$ (see Lemma 6.1), it follows that

$$\sup_{\theta\in\Theta} |\ \delta_n(\theta) - \delta(\theta)| \longrightarrow 0,$$

in probability as in $n \longrightarrow \infty$ that concludes the first step.

For the second step, note that, thanks to the uniqueness of the minimum θ_0, δ satisfies the following property: for any $\epsilon > 0$, there exists a $\zeta > 0$ such that for any θ, $|\theta_0 - \theta| > \epsilon$ implies that $|\delta(\theta_0) - \delta(\theta)| \ge \zeta$. In fact, suppose that this is false, then there exists an $\epsilon > 0$ and a sequence $\{t_n\}$ such that as $n \to \infty$, $\delta(t_n) \to \delta(\theta_0)$ and $\|\theta_0 - t_n\| \ge \epsilon$. The fact that Θ is compact and the last statement imply that $t_n \to \theta \in \Theta\backslash\{\theta_0\}$ with $\delta(\theta) = \delta(\theta_0)$ but this contradicts with the assumption of the uniqueness of the minimum θ_0. Hence to prove that $\theta_n \longrightarrow \theta_0$ in probability, it is enough to show that $\delta(\theta_n) \longrightarrow \delta(\theta_0)$ in probability.

Consider then

$$|\delta(\theta_n) - \delta(\theta_0)| \le |\delta(\theta_n) - \delta_n(\theta_n)| + |\delta_n(\theta_n) - \delta(\theta_0)|,$$

where

$$|\delta(\theta_n) - \delta_n(\theta_n)| \le \sup_{\theta\in\Theta} |\delta(\theta) - \delta_n(\theta)|$$

and, because θ_n and θ_0 are the maximizer of δ_n and δ respectively,

$$|\delta_n(\theta_n) - \delta(\theta_0)| = \left|\sup_{\theta\in\Theta} \delta_n(\theta) - \sup_{\theta\in\Theta} \delta(\theta)\right| \le \sup_{\theta\in\Theta} |\delta(\theta) - \delta_n(\theta)|$$

one gets

$$|\delta(\theta_n) - \delta(\theta_0)| \leq 2 \sup_{\theta \in \Theta} |\delta(\theta) - \delta_n(\theta)|,$$

which goes to zero in probability as $n \longrightarrow \infty$ and this provides the consistency.

For what concerns the rate of convergence, since $\delta(\theta) \in C^2(\Theta)$, consider the Taylor expansion with Lagrange remainder

$$\delta(\theta_n) = \delta(\theta_0) + \delta'(\theta_0)(\theta_n - \theta_0) + \frac{\delta''(\theta^\star)}{2}(\theta_n - \theta_0)^2$$

where $\theta^\star$ is a suitable element of the interval whose extremes are θ_n, θ_0. Since θ_0 minimize δ, $\delta'(\theta_0) = 0$, $\delta''(\theta^\star) > 0$ and, thanks to above results one obtains

$$|\theta_n - \theta_0| = \sqrt{\left|\frac{2(\delta(\theta_n) - \delta(\theta_0))}{\delta''(\theta^\star)}\right|} \leq C \left(\sup_{h \in \mathcal{H}} |\phi_n(h) - \phi_{\theta_0}(h)| \right)^{1/2}.$$

Finally, thanks to Lemma 6.1, there exist two positive sequences of positive integers $L = L(n)$, $m = m(n)$ and a real valued function f such that:

$$|\theta_n - \theta_0| = O_p\left((2\alpha_L + \beta_{m-2L} + f(2\alpha_L + \beta_{m-2L}))^{1/2} \right)$$

and this concludes the proof.

6.2 Proof of Theorem 2.5

Compute the first and second derivatives of $\delta(\theta)$ with respect to θ:

$$\delta'(\theta) = -2\langle \widetilde{\phi}_{\theta_0}, \widetilde{\phi}_\theta \rangle \left\langle \widetilde{\phi}_{\theta_0}, \widetilde{\phi}'_\theta \right\rangle,$$

$$\delta''(\theta) = -2 \left(\left\langle \widetilde{\phi}_{\theta_0}, \widetilde{\phi}'_\theta \right\rangle^2 + \langle \widetilde{\phi}_{\theta_0}, \widetilde{\phi}_\theta \rangle \left\langle \widetilde{\phi}_{\theta_0}, \widetilde{\phi}''_\theta \right\rangle \right).$$

Evaluating the above expressions at θ_0, one obtains:

$$\delta'(\theta_0) = 0 \quad \text{and} \quad \delta''(\theta_0) = -2 \left\langle \widetilde{\phi}_{\theta_0}, \widetilde{\phi}''_{\theta_0} \right\rangle. \tag{10}$$

The Taylor expansion of $\delta'_n(\theta)$ around θ_0 with Lagrange remainder is

$$\delta'_n(\theta) = \delta'_n(\theta_0) + \delta''_n(\theta^\star)(\theta - \theta_0), \tag{11}$$

where $\theta^\star$ is between θ and θ_0.

Since $\delta'_n(\theta_n) = 0$ and $\delta''_n(\theta^\star) > 0$, Equation (11) can be re–expressed as:

$$\theta_n - \theta_0 = -\delta'_n(\theta_0)/\delta''_n(\theta^\star),$$

where

$$\delta'_n(\theta_0) = -2\langle \widetilde{\phi}_n, \widetilde{\phi}_{\theta_0} \rangle \left\langle \widetilde{\phi}_n, \widetilde{\phi}'_{\theta_0} \right\rangle,$$

and

$$\delta_n''(\theta^*) = -2\left(\left\langle \widetilde{\phi}_n, \widetilde{\phi}'_{\theta^*}\right\rangle^2 + \langle \widetilde{\phi}_n, \widetilde{\phi}_{\theta^*}\rangle \left\langle \widetilde{\phi}_n, \widetilde{\phi}''_{\theta^*}\right\rangle\right).$$

Using Lemma 6.1 and (10), $\langle \widetilde{\phi}_n, \widetilde{\phi}_{\theta_0}\rangle$ and $\left\langle \widetilde{\phi}_n, \widetilde{\phi}'_{\theta_0}\right\rangle$ converge in probability to 1 and 0 respectively, as $n \to \infty$. By the definition of θ^*, Lemma 6.1 and Theorem 2.4, $\left\langle \widetilde{\phi}_n, \widetilde{\phi}'_{\theta^*}\right\rangle$ converges in probability to 0 as $n \longrightarrow \infty$ and so $\delta_n''(\theta^*)$ to $\delta''(\theta_0)$. Finally, by combining the previous results, one gets:

$$\theta_n - \theta_0 \longrightarrow 2\left\langle \widetilde{\phi}_n, \widetilde{\phi}'_{\theta_0}\right\rangle / \delta''(\theta_0),$$

in probability as $n \longrightarrow \infty$.

Hence, to derive the asymptotic distribution of $\theta_n - \theta_0$, it is sufficient to study the law of $\left\langle \widetilde{\phi}_n, \widetilde{\phi}'_{\theta_0}\right\rangle = \langle \phi_n, \eta_{\theta_0}\rangle / \|\phi_n\| \, \|\phi_{\theta_0}\|^3$ where $\eta_{\theta_0} = \phi'_{\theta_0} \|\phi_{\theta_0}\|^2 - \phi_{\theta_0} \langle \phi_{\theta_0}, \phi'_{\theta_0}\rangle$.

Noticing that $\|\phi_n\|$ converges to $\|\phi_{\theta_0}\|$ as $n \longrightarrow \infty$ (thanks to Lemma 6.1), and $\langle \phi_{\theta_0}, \eta_{\theta_0}\rangle = 0$, it remains to study the asymptotic distribution of

$$\sqrt{n}\,\langle \phi_n, \eta_{\theta_0}\rangle = \sqrt{n}\,\langle \phi_n - \phi_{\theta_0}, \eta_{\theta_0}\rangle.$$

Since ϕ_n is a U–statistic, then the finite dimensional distributions of $\sqrt{n}(\phi_n(h) - \phi_{\theta_0}(h))$ converge to those of a centered Gaussian process (see Lemma 6.2), and hence $\sqrt{n}\,\langle \phi_n, \eta_{\theta_0}\rangle$ is asymptotically distributed as a Gaussian centered random variable.

The proof is achieved by combining the previous results and applying Slutsky Theorem.

6.3 Auxiliary lemmas

The following two lemmas provide the weak consistency of the U–statistic ϕ_n towards to $\phi_{\theta_0}(h)$ and the asymptotic normality of the U–process $\sqrt{n}\,(\phi_n(h) - \phi_{\theta_0}(h))$. The first lemma exploits similar arguments as in (Borovkova et al., 2001, Theorem 6) and Bongiorno et al. (2020), whereas the second one is a direct consequence of (Wendler, 2012, Theorem 1 and Corollary 1). Nevertheless, for the sake of completeness, the proofs are provided in the supplementary materials.

Lemma 6.1 *Under assumptions (A-1), (A-2) and* $\alpha_L = \left(2\sum_{l=L}^{\infty} a_l\right)^{1/2} < \infty$ *it holds:*

$$|\phi_n(h) - \phi(h)| = O_p\left(2\alpha_L + \beta_{m-2L} + f(2\alpha_L + \beta_{m-2L})\right), \qquad \forall h \in \mathcal{H},$$
$$\sup_{h\in\mathcal{H}} |\phi_n(h) - \phi(h)| = O_p\left(2\alpha_L + \beta_{m-2L} + f(2\alpha_L + \beta_{m-2L})\right),$$

where $L = L(n)$ *and* $m = m(n)$ *are two positive sequences of positive integers and* f *a suitable real valued function.*

Lemma 6.2 *Under assumptions (A-1) and (A-2) with* $\beta_k = O(k^{-\beta})$ *for* $\beta \geq 8$ *and* $a_k = O(k^{-a})$ *for* $a = \max\{\beta + 3, 12\}$, $\sqrt{n}\,(\phi_n(h) - \phi_{\theta_0}(h))$ *converges to a centered Gaussian process over* $\mathcal{H}$ *as* $n \to \infty$*.*

Acknowledgement

The authors would like to thank the Associate Editor and the three anonymous referees for their helpful comments which greatly improve the quality of the paper. Lax Chan holds a career grant supported by the European Commission – FSE REACT-EU, PON Ricerca e Innovazione 2014-2020. The authors are members of the Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni (GNAMPA) of the Istituto Nazionale di Alta Matematica (INdAM). This work is based upon work from COST Action HiTEc, CA21163, supported by COST (European Cooperation in Science and Technology).

References

Aneiros, G., Bongiorno, E. G., Cao, R., and Philippe, V. (2017). *Functional Statistics and Related Fields*. Contributions to Statistics. Springer, 1st ed. 2017 edition.

Aneiros, G., Cao, R., Fraiman, R., Genest, C., and Vieu, P. (2019a). Recent advances in functional data analysis and high-dimensional statistics. *Journal of Multivariate Analysis*, 170:3–9.

Aneiros, G., Cao, R., and Vieu, P. (2019b). Editorial on the special issue on functional data analysis and related topics. *Computational Statistics*, 34(2):447–450.

Aneiros, G., Horová, I., Hušková, M., and Vieu, P. (2022). Special issue on functional data analysis and related fields. *Journal of Multivariate Analysis*.

Aneiros, G., Horová, I., Hušková, M., and Vieu, P., editors (2020). *Functional and high–dimensional statistics and related fields*. Contributions to Statistics. Springer, Cham. Papers associated with the 5th International Workshop on Function and Operational Statistics (IWFOS 2020) held at Masaryk University, Brno, June 2020 (postponed).

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

Bogachev, V. I. (1998). *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI.

Bongiorno, E., Goia, A., and Vieu, P. (2020). Estimating the complexity index of functional data: Some asymptotics. *Statistics & Probability Letters*, 161:108731.

Bongiorno, E. G. and Goia, A. (2017). Some insights about the small ball probability factorization for Hilbert random elements. *Statist. Sinica*, 27(4):1949–1965.

Bongiorno, E. G., Goia, A., and Vieu, P. (2018). Evaluating the complexity of some families of functional data. *SORT*, 42(1):27–44.

Bongiorno, E. G., Goia, A., and Vieu, P. (2019). Modeling functional data: a test procedure. *Comput. Statist.*, 34(2):451–468.

Borovkova, S., Burton, R., and Dehling, H. (2001). Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Transactions of the American Mathematical Society*, 353(11):4261–4318.

Bosq, D. (2000). *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Doukhan, P. (1994). Mixing, volume 85 of lecture notes in statistics.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice (Springer Series in Statistics)*. Springer Series in Statistics. Springer, 1 edition.

Fusai, G. and Roncoroni, A. (2007). *Implementing models in quantitative finance: methods and cases*. Springer Science & Business Media.

Grassberger, P. and Procaccia, I. (1983). Characterization of strange attractors. *Phys. Rev. Lett.*, 50:346–349.

Hall, P. and Vial, C. (2006). Assessing the finite dimensionality of functional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(4):689–705.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics 200. Springer-Verlag New York, 1 edition.

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

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. Chapman & Hall CRC Texts in Statistical Science. Chapman and Hall CRC, 1 edition.

Li, W. and Shao, Q.-M. (2001). Gaussian processes: Inequalities, small ball probabilities and applications. In *Stochastic Processes: Theory and Methods*, volume 19 of *Handbook of Statistics*, pages 533–597. Elsevier.

Lifshits, M. (2012). *Lectures on Gaussian Processes*. Springer Briefs in Mathematics. Springer-Verlag Berlin Heidelberg, 1 edition.

Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the limits of bootstrap (East Lansing, MI, 1990)*, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., pages 225–248. Wiley, New York.

Magnano, L., Boland, J. W., and Hyndman, R. J. (2008). Generation of synthetic sequences of half-hourly temperature. *Environmetrics: The official journal of the International Environmetrics Society*, 19(8):818–835.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis (Springer Series in Statistics)*. Springer, 2nd edition.

Rychlik, T. (1994). Distributions and expectations of order statistics for possibly dependent random variables. *Journal of Multivariate Analysis*, 48(1):31–42.

van den Boogaart, K. G., Egozcue, J. J., and Pawlowsky-Glahn, V. (2014). Bayes Hilbert spaces. *Aust. N. Z. J. Stat.*, 56(2):171–194.

Wendler, M. (2012). U-processes, u-quantile processes and generalized linear statistics of dependent data. *Stochastic Processes and their Applications*, 122(3):787–807.

Supplementary materials

Proof of Lemma 6.1

Consider the U-distribution function defined as $\mathbb{P}\left(d\left(X,Y\right)\leq h\right)=\mathbb{E}\left[\mathbb{I}_{\{d(X,Y)\leq h\}}\right]$, where Y is an independent copy of X and whose empirical counterpart is $\phi_n(h)$. Note that, thanks to the law of total expectation, Assumption (A-2), the identification constraint $\mathbb{E}\left[\psi\left(X\right)\right]=1$, when h is small enough it holds:

$$\begin{aligned}\mathbb{E}\left[\mathbb{I}_{\{d(X,Y)\leq h\}}\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{I}_{\{d(X,Y)\leq h\}}|Y\right]\right]=\mathbb{E}\left[\mathbb{P}\left(X\in B\left(Y,h\right)\right)\right]\\ &= \mathbb{E}\left[\psi\left(Y\right)\phi_{\theta_0}\left(h\right)\right]+o\left(\phi_{\theta_0}\left(h\right)\right)\\ &= \phi_{\theta_0}\left(h\right)\left(1+o\left(1\right)\right).\end{aligned}$$

Thus, to prove the asymptotic behavior of $\phi_n\left(h\right)$ towards $\phi_{\theta_0}\left(h\right)$ is equivalent to show that $\phi_n\left(h\right)$ converges to the U-distribution function $\mathbb{E}\left[\mathbb{I}_{\{d(X,Y)\leq h\}}\right]$. Moreover, it is useful to note that, since the monomial behaviour is assumed, $\phi_{\theta_0}\left(h\right)$ is Lipschitz continuous on $\mathcal{H}$ and the U-distribution as well. In fact, for $h,h'\in\mathcal{H}$, there exists a positive constant L that

$$\begin{aligned}\left|\mathbb{P}\left(d\left(X,Y\right)\leq h\right)-\mathbb{P}\left(d\left(X,Y\right)\leq h'\right)\right| &= \left|\phi_{\theta_0}\left(h\right)-\phi_{\theta_0}\left(h'\right)\right|\left(1+o\left(1\right)\right)\\ &\leq L\left|h-h'\right|\left(1+o\left(1\right)\right).\end{aligned}\tag{12}$$

Make now the following observations.

Observation 1. $\phi_n(h)$ is a second–order$U$-statistic with bounded symmetric kernel $\mathbb{I}_{\{d(\cdot,\cdot)\leq h\}}$.

Observation 2. Each random variable $\mathbb{I}_{\{d(X_i,X_j)\leq h\}}$ is bounded for every $i\neq j$ and then the family $\left\{\mathbb{I}_{\{d(X_i,X_j)\leq h\}},1\leq i,j\leq n\right\}$ is uniformly integrable.

Observation 3. Consider X_1,X_1',X_k, $\mathcal{F}$-valued random elements that are independently and identically distributed. If $d(X_1,X_k)\leq h-\eta$ and $d(X_1,X_1')\leq\eta$, as a consequence of the triangle inequality, $d(X_1',X_k)\leq h$, and similarly if $d(X_1,X_k)>h+\eta$ and $d(X_1,X_1')\leq\eta$ the reverse triangle inequality implies that $d(X_1',X_k)>h$. Hence

$$\left|\mathbb{I}_{\{d(X_1,X_k)\leq h\}}-\mathbb{I}_{\{d(X_1',X_k)\leq h\}}\right|\mathbb{I}_{\{d(X_1,X_1')\leq\eta\}}\leq\mathbb{I}_{\{h-\eta\leq d(X_1,X_k)\leq h+\eta\}}$$

and thus when h and η are small enough:

$$\begin{aligned}\mathbb{E}\left[\left|\mathbb{I}_{\{d(X_1,X_k)\leq h\}}-\mathbb{I}_{\{d(X_1',X_k)\leq h\}}\right|\mathbb{I}_{\{d(X_1,X_1')\leq\eta\}}\right] &\leq \mathbb{P}(h-\eta\leq d(X_1,X_k)\leq h+\eta)\\ &= \mathbb{P}(d(X_1,X_k)\leq h+\eta)-\mathbb{P}(d(X_1,X_k)\leq h-\eta)\\ &\leq f(\eta)\end{aligned}$$

where $f(\eta)=\max\left(\sup_k\mathbb{P}\left(h-\eta\leq d(X_1,X_k)\leq h+\eta\right),\mathbb{P}\left(h-\eta\leq d(Y_1,Y_2)\leq h+\eta\right)\right)$ and Y_1,Y_2 being independent with the same distribution as X_1. Moreover, because of (12), the kernel $\mathbb{I}_{\{d(\cdot,\cdot)\leq h\}}$ satisfies the 1-Lipschitz condition (see Borovkova et al., 2001, Equation (2.10)).

Thanks to above observations together with the assumptions (A-1) and (A-2), it emerges that the proof of the first part of this lemma is an adaptation of (Borovkova et al., 2001, Theorem 6) and it is reported for the benefit of the reader. The proof proceeds by considering the two preliminary technical results and, in what follows, C denotes a universal positive constant.

First, the uniform integrability guarantees that, for a fixed $\varepsilon>0$ there exists $0<\delta_0\leq\varepsilon$ for which for any measurable sets B such that $\mathbb{P}(B)<\delta<\delta_0$ then

$$\mathbb{E}\left[\mathbb{I}_{\{d(X_i,X_j)\leq h\}}\mathbb{I}_B\right]=\mathbb{E}\left[\mathbb{I}_{\{d(X_i,X_j)\leq h\}\cap B}\right]=\mathbb{P}(\{d\left(X_i,X_j\right)\leq h\}\cap B)\leq\mathbb{P}(B)<\varepsilon.\tag{13}$$

Second, choose three sequences of positive integers $m = m(n), L = L(n), N = N(n)$ (in general increasing with n) such that

$$\begin{cases} L < m/2, \\ m/N < \delta \leq \varepsilon \\ 2\alpha_L + \beta_{m-2L} < 1 \\ f(\delta) < \varepsilon \end{cases}$$

where $\delta = \delta_0(2\alpha_L + \beta_{m-2L})$. Define the integer $n_k = (k-1)(m+N)$ and consider the blocks of length N

$$\xi_k = (X_{n_k+1}, \ldots, X_{n_k+N}) .$$

Given $(X_1, \ldots, X_n)$, $p = \lfloor n/(N+m) \rfloor$ is the index of the last block ξ_k to be fully contained in $(X_1, \ldots, X_n)$ (where $\lfloor a \rfloor$ denotes the integer part of a). By adapting the notion of *nearly regular process* as in the Definition 2.6. of Borovkova et al. (2001) and as a consequence of (Borovkova et al., 2001, Theorem 3), it is possible to find a sequence of i.i.d. N–dimensional vectors $\xi_1', \xi_2', \ldots$ with the same distribution as $(\xi_k)_k$ such that for all $k = 1, 2, \ldots,$

$$\mathbb{P}\left(\rho\left(\xi_k, \xi_k^{'}\right) \geq \delta\right) \leq \delta/2,$$

with $\rho(\xi_k, \xi_k^{'}) = \max_j d\left(X_j, X_j'\right)$, where X_j' in an independent copy of X_j. Here $\rho(\xi_k, \xi_k^{'})$ replaces the L^1-norm on $\mathbb{R}^N$ and $(\xi_k^{'})_k$ are independent $\mathcal{F}$-valued random elements.

The idea of the proof is to decompose the global sum defining ϕ_n in partial sums of suitable blocks and controlling their asymptotic behavior. To do this define the kernel $K_h : \mathcal{F}^N \times \mathcal{F}^N \to \mathbb{R}$ by

$$K_h(\mathbf{x}, \mathbf{y}) := \frac{1}{N^2} \sum_{1 \leq i,j \leq N} \mathbb{I}_{\{d(x_i, y_j) \leq h\}},$$

where $\mathbf{x} = (x_1, \ldots, x_N)$ and $\mathbf{y} = (y_1, \ldots, y_N)$. From (13), one can infer that, for $k \neq l$

$$\mathbb{E}\left[\, K_h(\xi_k, \xi_l)\ \mathbb{I}_B\right] \leq \delta \tag{14}$$

for all measurable sets B with $\mathbb{P}(B) < \delta$. Given the independence of ξ_k' and ξ_l', by Hoeffding's classical U-statistics law of large numbers for independent observations, it follows

$$\frac{1}{p(p-1)} \sum_{1 \leq k \neq l \leq p} K_h(\xi_k', \xi_l') \longrightarrow \mathbb{E}\left[\mathbb{I}_{\{d(X,Y) \leq h\}}\right] \tag{15}$$

almost surely and in L^1, where Y is an independent copy of X.

As a consequence of 1-Lipschitz condition, the error introduced by replacing ξ_i by ξ_i' is negligible:

$$\mathbb{E}\left[\left|\frac{1}{p(p-1)} \sum_{1 \leq k \neq l \leq p} K_h(\xi_k, \xi_l) - \frac{1}{p(p-1)} \sum_{1 \leq k \neq l \leq p} K_h(\xi_k', \xi_l')\right|\right] \leq C(\delta + f(\delta)). \tag{16}$$

Moreover, since

$$\left|\frac{1}{p(p-1)} - \frac{N^2}{n(n-1)}\right| \leq \frac{2\delta}{p(p-1)}$$

for p large enough,

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{p(p-1)}\sum_{1\le k\neq l\le p}K_h(\xi_k,\xi_l)-\frac{N^2}{n(n-1)}\sum_{1\le k\neq l\le p}K_h(\xi_k,\xi_l)\right|\right]\\
&=\mathbb{E}\left[\left|\left(\frac{1}{p(p-1)}-\frac{N^2}{n(n-1)}\right)\sum_{1\le k\neq l\le p}K_h(\xi_k,\xi_l)\right|\right]\\
&\le\left|\left(\frac{1}{p(p-1)}-\frac{N^2}{n(n-1)}\right)\right|\sum_{1\le k\neq l\le p}\mathbb{E}\left[K_h(\xi_k,\xi_l)\right]\\
&\le 2\delta\mathbb{E}\left[K_h(\xi_1,\xi_2)\right]\le C\delta.
\end{aligned}$$

The last estimate, together with (15) and (16), shows that for n large enough

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{N^2}{n(n-1)}\sum_{1\le k\neq l\le p}K_h(\xi_k,\xi_l)-\mathbb{E}[\mathbb{I}_{\{d(X,X')\le h\}}]\right|\right]\\
&=\mathbb{E}\left[\left|\left(\frac{1}{p(p-1)}\sum_{1\le k\neq l\le p}K_h(\xi_k,\xi_l)-\frac{1}{p(p-1)}\sum_{1\le k\neq l\le p}K_h(\xi'_k,\xi'_l)\right)+\right.\right.\\
&\left.\left.-\left(\frac{1}{p(p-1)}\sum_{1\le k\neq l\le p}K_h(\xi_k,\xi_l)-\frac{N^2}{n(n-1)}\sum_{1\le k\neq l\le p}K_h(\xi_k,\xi_l)\right)\right|\right]\le C(\delta+f(\delta)). \qquad (17)
\end{aligned}$$

Introduce the following decomposition involving the original kernel $\mathbb{I}_{\{d(X_i,X_j)\le h\}}$:

$$\begin{aligned}
n(n-1)\phi_n(h)&=\sum_{1\le i\neq j\le n}\mathbb{I}_{\{d(X_i,X_j)\le h\}}=\sum_{1\le k\neq l\le p}\sum_{i=n_k+1}^{n_k+N}\sum_{j=n_l+1}^{n_l+N}\mathbb{I}_{\{d(X_i,X_j)\le h\}}\\
&+\sum_{k=1}^{p}\sum_{n_k+1\le i\neq j\le n_k+N}\mathbb{I}_{\{d(X_i,X_j)\le h\}}+2\sum_{1\le k,l\le p}\sum_{i=n_k+N+1}^{n_k+1}\sum_{j=n_l+1}^{n_l+N}\mathbb{I}_{\{d(X_i,X_j)\le h\}}\\
&+\sum_{1\le k\neq l\le p}\sum_{i=n_k+N+1}^{n_{k+1}}\sum_{j=n_l+N+1}^{n_l+1}\mathbb{I}_{\{d(X_i,X_j)\le h\}}+\sum_{k=1}^{p}\sum_{n_k+N+1\le i\neq j\le n_k+1}\mathbb{I}_{\{d(X_i,X_j)\le h\}}\\
&+\sum_{i=n_p+N+1}^{n}\sum_{j=1}^{n}\mathbb{I}_{\{d(X_i,X_j)\le h\}}+\sum_{i=1}^{n_p+N}\sum_{j=n_p+N+1}\mathbb{I}_{\{d(X_i,X_j)\le h\}}
\end{aligned}$$

By studying the indexes, one gets

$$\begin{aligned}
&\mathbb{E}\left[\left|n(n-1)\phi_n(h)-\sum_{1\le k\neq l\le p}\sum_{i=n_k+1}^{n_k+N}\sum_{j=n_l+1}^{n_l+N}\mathbb{I}_{\{d(X_i,X_j)\le h\}}\right|\right]\\
&\le pN^2+2p^2mN+p^2mN+p^2m^2+2n(m+N).
\end{aligned}$$

Provided $p \leq n/N + m$, one has $p \leq n/N$ and $m \leq \delta N$ (this shows that the random variables in the small separating blocks of length m can be neglected). Using these facts and some algebraic manipulation, the right hand side of the inequality obtained above is bounded by $C(\delta + N/n)n^2$ and hence,

$$\mathbb{E}\left[\left|\phi_n(h) - \frac{N^2}{n(n-1)}\sum_{1\leq k\neq l\leq p} K_h(\xi_k, \xi_l)\right|\right] \leq C\delta,$$

for n large enough.

The above inequality combines with (17), gives

$$\begin{aligned}\mathbb{E}\Bigg[\Bigg|\phi_n(h) - \frac{N^2}{n(n-1)}\sum_{1\leq k\neq l\leq p} K_h(\xi_k, \xi_l) + \\ + \frac{N^2}{n(n-1)}\sum_{1\leq k\neq l\leq p} K_h(\xi_k, \xi_l) - \mathbb{E}[\mathbb{I}_{\{d(X,X')\leq h\}}]\Bigg|\Bigg] \\ \leq \mathbb{E}\left[\left|\phi_n(h) - \frac{N^2}{n(n-1)}\sum_{1\leq k\neq l\leq p} K_h(\xi_k, \xi_l)\right|\right] + \\ + \mathbb{E}\left[\left|\frac{N^2}{n(n-1)}\sum_{1\leq k\neq l\leq p} K_h(\xi_k, \xi_l) - \mathbb{E}[\mathbb{I}_{\{d(X,X')\leq h\}}]\right|\right] \leq C(\delta + f(\delta)) \leq C\varepsilon\end{aligned}$$

and, thanks to Markov inequality and for the arbitrariness of ε, ϕ_n converges in probability to ϕ_{θ_0} at the rate $\delta + f(\delta) = 2\alpha_L + \beta_{m-2L} + f(2\alpha_L + \beta_{m-2L})$.

The second part of the lemma, that is the uniform convergence in probability, is guaranteed by Proposition 1 in Bongiorno et al. (2020) from which one can deduce that the rate of convergence does not change and this concludes the proof.

Proof of Lemma 6.2

Given the fact that, under Assumption (A-2), the U-distribution function is Lipschitz-continuous as shown in Equation (12), the proof is a direct consequence of (Wendler, 2012, Theorem 1 and Corollary 1).

References

Aneiros, G., Bongiorno, E. G., Cao, R., and Philippe, V. (2017). *Functional Statistics and Related Fields*. Contributions to Statistics. Springer, 1st ed. 2017 edition.

Aneiros, G., Cao, R., Fraiman, R., Genest, C., and Vieu, P. (2019a). Recent advances in functional data analysis and high-dimensional statistics. *Journal of Multivariate Analysis*, 170:3–9.

Aneiros, G., Cao, R., and Vieu, P. (2019b). Editorial on the special issue on functional data analysis and related topics. *Computational Statistics*, 34(2):447–450.

Aneiros, G., Horová, I., Hušková, M., and Vieu, P. (2022). Special issue on functional data analysis and related fields. *Journal of Multivariate Analysis*.

Aneiros, G., Horová, I., Hušková, M., and Vieu, P., editors (2020). *Functional and high–dimensional statistics and related fields*. Contributions to Statistics. Springer, Cham. Papers associated with the 5th International Workshop on Function and Operational Statistics (IWFOS 2020) held at Masaryk University, Brno, June 2020 (postponed).

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

Bogachev, V. I. (1998). *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI.

Bongiorno, E., Goia, A., and Vieu, P. (2020). Estimating the complexity index of functional data: Some asymptotics. *Statistics & Probability Letters*, 161:108731.

Bongiorno, E. G. and Goia, A. (2017). Some insights about the small ball probability factorization for Hilbert random elements. *Statist. Sinica*, 27(4):1949–1965.

Bongiorno, E. G., Goia, A., and Vieu, P. (2018). Evaluating the complexity of some families of functional data. *SORT*, 42(1):27–44.

Bongiorno, E. G., Goia, A., and Vieu, P. (2019). Modeling functional data: a test procedure. *Comput. Statist.*, 34(2):451–468.

Borovkova, S., Burton, R., and Dehling, H. (2001). Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation. *Transactions of the American Mathematical Society*, 353(11):4261–4318.

Bosq, D. (2000). *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

Doukhan, P. (1994). Mixing, volume 85 of lecture notes in statistics.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice (Springer Series in Statistics)*. Springer Series in Statistics. Springer, 1 edition.

Fusai, G. and Roncoroni, A. (2007). *Implementing models in quantitative finance: methods and cases*. Springer Science & Business Media.

Grassberger, P. and Procaccia, I. (1983). Characterization of strange attractors. *Phys. Rev. Lett.*, 50:346–349.

Hall, P. and Vial, C. (2006). Assessing the finite dimensionality of functional data. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(4):689–705.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics 200. Springer-Verlag New York, 1 edition.

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

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis.* Chapman & Hall CRC Texts in Statistical Science. Chapman and Hall CRC, 1 edition.

Li, W. and Shao, Q.-M. (2001). Gaussian processes: Inequalities, small ball probabilities and applications. In *Stochastic Processes: Theory and Methods*, volume 19 of *Handbook of Statistics*, pages 533–597. Elsevier.

Lifshits, M. (2012). *Lectures on Gaussian Processes.* Springer Briefs in Mathematics. Springer-Verlag Berlin Heidelberg, 1 edition.

Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the limits of bootstrap (East Lansing, MI, 1990)*, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., pages 225–248. Wiley, New York.

Magnano, L., Boland, J. W., and Hyndman, R. J. (2008). Generation of synthetic sequences of half-hourly temperature. *Environmetrics: The official journal of the International Environmetrics Society*, 19(8):818–835.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis (Springer Series in Statistics).* Springer, 2nd edition.

Rychlik, T. (1994). Distributions and expectations of order statistics for possibly dependent random variables. *Journal of Multivariate Analysis*, 48(1):31–42.

van den Boogaart, K. G., Egozcue, J. J., and Pawlowsky-Glahn, V. (2014). Bayes Hilbert spaces. *Aust. N. Z. J. Stat.*, 56(2):171–194.

Wendler, M. (2012). U-processes, u-quantile processes and generalized linear statistics of dependent data. *Stochastic Processes and their Applications*, 122(3):787–807.