Learning the Smoothness of Weakly Dependent Functional Times Series

Hassan Maissoro[1], Valentin Patilea[2] & Myriam Vimond[3]

[1] *CREST, Ensai, and Datastorm ; France, hassan.maissoro@datastorm.fr*
[2] *CREST, Ensai; France, valentin.patilea@ensai.fr*
[3] *CREST, Ensai; France, myriam.vimond@ensai.fr*

Résumé. Nous considérons des séries temporelles fonctionnelles stationnaires où chaque observation est une trajectoire, mesurée avec erreur à des instants discrets, éventuellement aléatoires, du domaine d'observation. Nous considérons l'estimateur des paramètres de régularité locale des trajectoires introduit par Golovkine et al. (2022) sous l'hypothèse de faible dépendance, la $L^p - m-$approximabilité. Dans ce contexte, des bornes de concentration non-asymptotiques de l'estimateur de la régularité locale sont établies. Par la suite cette procédure permettra de diagnostiquer les changement de régularité le long de la trajectoire, construire une estimation optimale de la fonction d'autocovariance, *etc.* Au travers d'une étude de simulation, nous illustrons les bonnes capacités de la méthode proposée, ce qui nous permet d'esquisser quelques conseils pratiques sur la façon de sélectionner les hyper-paramètres de la procédure.

Mots-clés. Bornes de concentration, Lissage à noyau, Inégalité de Nagaev, Processus stochastique.

Abstract. We consider functional time series where the sample paths are observed with error at possibly random discrete in the domain. We reconsider the local regularity estimator proposed by Golovkine et al. (2022) in the context of weakly dependent curves, under the assumption of $L^p - m-$approximability. In this new framework, we derive non asymptotic exponential bounds for the concentration of the regularity estimators. This will further allow to diagnose a change of regularity along the sample paths, to build optimal estimator of mean and (auto)covariance functions, *etc.* An extensive simulation study illustrate the good performance of our estimators with finite time series. The simulation experiments also provide guidance for the choice of the hyperparameters involved in our estimation method.

Keywords. Concentration bounds, Kernel smoothing, Nagaev inequality, Stochastic processes.

1 Introduction

In functional data analysis (FDA), the unit of observation is an entire curve (also called trajectory or sample path), or a vector of curves. Thus, under reasonable assumptions, it is possible to learn the distribution of the underlying stochastic process, even if the number of measurements on each trajectory remains small (see Yao et al., 2005). However, the observations are in general noisy measurements of the curves, at discrete points in their domain, not necessarily regular or not necessarily the same from one curve to another. It is therefore necessary to reconstruct the trajectories, for instance using a nonparametric smoother. This is a very important step in FDA. With the reconstructed trajectories, quantities of interest, such as for instance the mean function and the covariance functions, can be easily estimated.

The quality of the estimates of these objects depends strongly on the nonparametric smoothing, and thus on the optimal choice of the smoothing bandwidth, which itself depends on the regularity of the underlying process that generated the trajectories. Golovkine et al. (2022) have introduced a method for estimating the local regularity parameters of the underlying process when the trajectories are independent and identically distributed. Next, they derive adaptive optimal estimates of irregular mean and covariance function (see Golovkine et al., 2021).

In the context of functional time series (FTS), estimators of the mean and autocovariance functions have been already proposed by Rubín and Panaretos (2020) under the assumption that these functions admit at least one derivative. However, in some cases, for example in the energy domain, the mean and autocovariance functions can be very irregular, of unknown irregularity. Several phenomena are naturally described by this type of data. This is the case for photovoltaic electricity production, which depends on the clouds. Thus, if the production of a photovoltaic park is observed for a sufficiently long period of time, it naturally generates data under the form of a set of irregular daily curves that are dependent on each other. In this work, we extend the results of Golovkine et al. (2022) to FTS in order to estimate the regularity of an underlying process which generated a set of dependent random curves. A final goal will be to have optimal estimates of objects such as the autocovariance function or the parameters of a functional autoregressive model when the trajectories are irregular.

Section 2 describes the statistical model associated with the observation of the FTS at discrete points in the domain, in the presence of additive noise. Section 3 and Section 4 describe respectively the local regularity assumption and the weak dependence assumption considered here. Section 5 presents the estimators and their concentration bounds.

2 The functional time series model

Functional time series consist of a collection of random functions observed over time, for which there is a temporal dependence. Let $I = [0, 1]$ be the domain where the functions are defined. The functional time series is then denoted $\{X_n\} = \{X_n(u),\ u \in I\}_{n\in\mathbb{Z}}$. We assume that for each $n \in \mathbb{Z}$, X_n is a random element of the Hilbert space $\mathcal{H} = \mathbb{L}^2(I)$ of square integrable functions equipped with the inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ (and $\|\cdot\|_{\mathcal{H}}$ is the associated norm). Moreover we assume that the paths of X_n are almost surely continuous functions, *i.e.*, X_n belongs to the Banach space $\mathcal{C} = \mathcal{C}(I)$ of continuous functions equipped with the sup-norm $\|\cdot\|_\infty$. We will see in Section 4 why we need X_n to belong to $\mathcal{C}$. For $p \geq 1$, we denote by $\mathbb{L}^p$ the space of real valued random variable Z such that $\nu_p(Z) = (\mathbb{E}[|Z|^p])^{1/p} < \infty$. The spaces $\mathbb{L}^p_{\mathcal{H}}$ (resp. $\mathbb{L}^p_{\mathcal{C}}$) denote the space of $\mathcal{H}$-valued (resp. $\mathcal{C}$-valued) random function X

such that $\nu_p(\|X\|_{\mathcal{H}}) < \infty$ (resp. $\nu_p(\|X\|_\infty) < \infty$).

Observation scheme. For each $1 \leq n \leq N$, we observe the trajectory of X_n at M_n points $\{T_{n,i}, 1 \leq i \leq M_n\}$ of the interval I in the presence of an additive noise. The observations associated with a curve, or trajectory, X_n consist of the pairs $(Y_{n,i}, T_{n,i}) \in \mathbb{R} \times I$ where $Y_{n,i}$ is defined as

$$Y_{n,i} = X_n(T_{n,i}) + \epsilon_{n,i}, \qquad 1 \leq n \leq N, \;\; 1 \leq i \leq M_n. \tag{1}$$

We assume that:

(H1) $\{X_n\}$ is a stationary process of $\mathbb{L}^2_{\mathcal{H}}$;

(H2) the number of observation points in the domain for each curve, denoted $M_1, \ldots, M_N$, are independent copies of an integer-valued random variable M, with expectation μ ;

(H3) the observation points $\{T_{n,i}\}_{n,i}$ are independent copies of a random variable T taking values in I ;

(H4) the additive noises $\{\epsilon_{n,i}\}_{n,i}$ are independent copies of a centered error variable ε with a finite variance ;

(H5) the sequences $\{X_n\}_n$, $\{M_n\}_n$, $\{T_{n,i}\}_{n,i}$ and $\{\epsilon_{n,i}\}_{n,i}$ are mutually independent.

In some applications, Assumption (H4) of homoscedasticity could be unrealistic. Nevertheless this assumption is convenient in order that $\{X_n\}$ transmits its stationarity (Assumption (H1)) to the pre-smoothing process $\{\widetilde{X}_n\}$, see (5).

3 The local regularity parameters

We are interested in studying the *local regularity* at $t \in (0,1)$ of the stationary distribution of $\{X_n\}$ as it is introduced by Golovkine et al. (2022). First, to simplify the presentation, we assume that the process is not differentiable almost surely in a neighborhood J of t such that the process is *locally Hölder in quadratic mean* :

$$\mathbb{E}\left[(X(u) - X(v))^2\right] \approx L_t^2 |u - v|^{2H_0}, \qquad u, v \in J. \tag{2}$$

Here $\approx$ means the left-hand side is equal to the right-hand side times a quantity which tends to 1 when $|u - v| \to 0$. More precisely, for some $H_0 \in (0,1)$ and for an open sub-interval J of I such that $t \in J$, we assume that

(H6) the stationary distribution of $\{X_n\}$ belongs to the class $\mathcal{X}(H_0, J)$ describing the set of stochastic processes $X = \{X(t), t \in I\}$ verifying the following assertions,

(a)

$$0 < \underline{a}_0 = \inf_{u \in J} \mathbb{E}\left[X(u)^2\right] \leq \sup_{u \in J} \mathbb{E}\left[X(u)^2\right] = \overline{a}_0 < \infty.$$

(b) for all $t \in J$, there exist $\Delta \in (0,1]$ with $[t-\Delta/2, t+\Delta/2] \subset J$ such that for all u, v satisfying $t - \Delta/2 \leq u \leq t \leq v \leq t + \Delta/2$, we have

$$\left|\theta(u,v) - L_t^2|u-v|^{2H_0}\right| \leq S_t^2|u-v|^{2H_0}\Delta^{2\beta_0},$$

where $\theta(u,v) = \mathbb{E}\left[(X(u)-X(v))^2\right]$, $\beta_0 > 0$ is a fixed positive constant, and $L_t > 0$, $S_t \geq 0$ are constants which could vary with t.

If the process $\{X_n\}$ is δ times differentiable, then conditions (a) and (b) of (H6) are considered for the $\delta-$th order derivative of the sample path. Here we focus on the non-differentiable case and we aim to estimate the local regularity parameter H_0 and the Hölder constant L_t^2 at $t \in J$.

Example 1 (Locally Hölder regularity in quadratic mean of a FAR(1) process)**.** Let $\{X_n\}_{n\in\mathbb{Z}}$ be a mean zero process following a FAR(1),

$$X_n(t) = \Psi(X_{n-1})(t) + \xi_n(t), \qquad t \in I, \quad n \in \mathbb{Z}, \tag{3}$$

where $\{\xi_n\}_{n\in\mathbb{Z}}$ are i.i.d. fractional Brownian motion (fBm) of Hurst exponent $H_\xi \in (0,1)$ with

$$\mathbb{E}\left[(\xi_n(u)-\xi_n(v))^2\right] = L_{H_\xi}^2|u-v|^{2H_\xi}.$$

The autoregressive operator Ψ is assumed to be an integral operator defined by,

$$\Psi(x)(t) = \int_I \psi(s,t)x(s)ds, \qquad x \in \mathcal{C}.$$

A sufficient condition of the existence of a stationary solution for (3), under the form (4) below, is $\int_I\int_I \psi^2(s,t)dsdt < 1$, see for instance Kokoszka and Reimherr (2017, Section 8.8). Assume further that the kernel of the operator is such that,

$$|\psi(s,u) - \psi(s,v)|^2 \leq C|u-v|^{2H_\psi}, \qquad \text{for every } s \text{ and } |u-v| < \Delta \leq 1,$$

where $H_\psi \in (0,1)$ is greater than H_ε, and C is a positive constant not depending on s. Then, FAR(1) process belongs to the class $\mathcal{X}(H_\xi, J)$.

4 Weak Dependency

Since we are studying the process $\{X_n\}$ locally at $t \in J$, we need to consider a dependence type of $\{X_n\}$ that can be inherited by $\{X_n(t)\}$. Our choice is to borrow the concept of $\mathbb{L}^p_{\mathcal{H}} - m-$*approximability* introduced by Hörmann and Kokoszka (2010) in context of $\mathcal{H}$-valued random elements. The general idea is to approximate $\{X_n\}_{n\in\mathbb{Z}}$ by a m-dependent process $\{X_n^{(m)}\}_{n\in\mathbb{Z}}$ such that, for every $n \in \mathbb{Z}$, the sequence $\{X_n^{(m)}\}_{m\geq 1}$ converges in some sens to X_n as $m \to \infty$. If the convergence is fast enough, then the limiting behaviour of the original process may be similar to the m-dependence sequences. Here, in order to get the $\mathbb{L}^p - m$-approximability for the sequences $\{X_n(t)\}_{n\in\mathbb{Z}}$ for all $t \in I$, instead of considering the norm of $\mathcal{H}$ to define this convergence, we use the norm $\|\cdot\|_\infty$ of $\mathcal{C}$:

(H7) the stationary process $\{X\}$ is $\mathbb{L}^p_{\mathcal{C}} - m - \textit{approximable}$ with $p \geq 4$,

(a) The paths of $\{X_n\}$ are almost surely continuous such that $\{X_n\} \subset \mathbb{L}^p_{\mathcal{C}}$.

(b) The process $\{X_n\}$ admits a moving average representation,

$$X_n = f(\xi_n, \xi_{n-1}, \ldots) \tag{4}$$

where the $\{\xi_n\}$ are i.i.d. elements in a measurable space S, and $f : S^\infty \to \mathcal{H}$ is measurable.

(c) For every $n \in \mathbb{Z}$, let $\{\xi_k^{(n)}\}_k$ be an independent copy of $\{\xi_n\}_n$ defined on the same probability space. The coupled version of X_n is defined by

$$X_n^{(m)} = f(\xi_n, \xi_{n-1}, \ldots, \xi_{n-m+1}, \xi_{n-m}^{(n)}, \xi_{n-m-1}^{(n)}, \ldots),$$

(d) The sequence $\{X_n^{(m)}\}_{m\geq 1}$ converges to X_n as $m \to \infty$ in the following way

$$\sum_{m\geq 1} \nu_p \left(\left\| X_m - X_m^{(m)} \right\|_\infty \right) < \infty.$$

Notice that $\mathbb{L}^p_{\mathcal{C}} - m$-approximability implies $\mathbb{L}^p_{\mathcal{H}} - m$-approximability since the $\| \cdot \|_{\mathcal{H}}$ norm is bounded by the infinity norm. Moreover the basic algebra properties of $\mathbb{L}^p_{\mathcal{H}} - m$-approximability established by Hörmann and Kokoszka (2010, Lemma 2.1) hold true with this definition, and the $\mathbb{L}^p_{\mathcal{C}} - m$-approximability of $\{X_n\}$ entails the $\mathbb{L}^p - m$-approximability of $\{X_n(t)\}_{n\in\mathbb{Z}}$ for all $t \in I$. The FAR(1) defined in Exemple 1 is $\mathbb{L}^p_{\mathcal{C}} - m$-approximable.

5 Estimation of the local regularity parameters

First we consider the case where the process is not differentiable almost surely in $t \in (0, 1)$. Let t_1, t_2 and t_3 be in $[t - \Delta/2, t + \Delta/2] \subset J$ such that $t_3 - t_1 = \Delta$ and $t_2 = t$ is the middle point of $[t_1, t_3]$. Using the definition of local regularity (Assumption (H6)), the following proxy values of H_0 and L_t^2 are considered,

$$\widetilde{H}_0(\Delta) = \frac{\log(\theta(t_1, t_3)) - \log(\theta(t_1, t_2))}{2\log(2)} \qquad \text{and} \qquad \widetilde{L}_t^2(\Delta) = \frac{\theta\left(t_1, t_3\right)}{\Delta^{2H_0}}.$$

It can be shown that $\widetilde{H}_0(\Delta)$ and $\widetilde{L}_t^2(\Delta)$ converge respectively to H_0 and L_t^2 as Δ goes to 0. Moreover we remark that $\widetilde{L}_t^2(\Delta) = L_t^2$ and $\widetilde{H}_0(\Delta) = H_0$ if $S_t^2 = 0$. Our estimators of (H_0, L_t^2) is obtained by plugging in a suitable estimator of $\theta(u, v)$ for $(u, v) \in \{(t_1, t_3), (t_1, t_2)\}$ in the definition of these proxy values.

Presmoothing step The estimation of $\theta(u, v)$ implies the reconstruction of the observed curves $X_1, \ldots, X_N$ at point u and v since these curves are discretely observed in presence of an additive noise, see (1). To this aim of recovering the continuity property of trajectories and preserving the stationarity, we use the same linear non-parametric procedure for every X_n using its associated sampled points $(Y_{n,k}, T_{n,k})$, $1 \leq k \leq M_n$:

$$\widetilde{X}_n(u) = \sum_{i=1}^{M_n} W_{n,i}(u) Y_{n,i}, \qquad u \in J,\ n = 1 \ldots, N, \tag{5}$$

where the weights $\{W_{n,i}\}_{i=1 \ldots M_n}$ are function of $(M_n, T_{n,1}, \ldots, T_{n,M_n})$ and a suitably selected smoothing parameter. This pre-smoothing is such that,

(H8) the sum of absolute values of weights are uniformly bounded,

$$\sup_{n=1\ldots N} \sup_{u \in J} \sum_{i=1}^{M_n} |W_{n,i}(u)| \leq 1 \qquad \text{almost surely.}$$

(H9) for all curves X_n, the quadratic risk of $\widehat{X}_n$ is uniformly contrrolled,

$$R_2 = \sup_{u \in J} \mathbb{E}\left[(\widetilde{X}_n(u) - X_n(u))^2\right] \leq B\mu^{-\tau},$$

where $B > 0$ and $\tau > 0$ are fixed constants.

Assumption (H8) is satisfied by the Nadaraya-Watson procedure, while Assumption (H9) is true under some smoothness assumptions on the density of T and the functions $\{X_n\}$.

Local regularity estimators Given a nonparametric estimator $\widetilde{X}_n$ of X_n, for $t \in J$, this suggests to define a natural estimator of $\theta(u, v)$ as,

$$\widehat{\theta}(u, v) = \frac{1}{N} \sum_{n=1}^{N} \left(\widetilde{X}_n(v) - \widetilde{X}_n(u)\right)^2, \quad u, v \in J.$$

Then our estimators of H_0 and L_t^2 are defined as,

$$\widehat{H}_0 = \frac{\log(\widehat{\theta}(t_1, t_3)) - \log(\widehat{\theta}_0(t_1, t_2))}{2\log(2)} \qquad \text{and} \qquad \widehat{L}_t^2 = \frac{\widehat{\theta}\,(t_1, t_3)}{\Delta^{2\widehat{H}_0}}.$$

Theorem 1. *Assume the assumptions (H1) — (H9) hold such that*

$$\nu_4(X_m - X_m^{(m)}) = \mathcal{O}(1/m^\alpha), \quad \alpha > 3/2,$$
$$\Delta^{2\beta_0} S_t^2 < \frac{L_t^2 \log(2)}{4} \varphi,$$
$$\mu^{-\tau/2} < C L_t^2 \varphi \Delta^{2H_0},$$

for some $\varphi \in (0, 1)$ and for some positive constant C depending on B, $\overline{a}_0$ and τ. Then for any enough large μ, we have

$$\mathbb{P}\left(|\widehat{H}_0 - H_0| > \varphi\right) \leq \frac{\mathfrak{f}_1}{N\varphi^2\Delta^{4H_0}} + 4\mathfrak{b}\exp\left(-\mathfrak{f}_2 N\varphi^2\Delta^{4H_0}\right),$$

where $\mathfrak{b}$ is a positive constant and $\mathfrak{f}_1$ and $\mathfrak{f}_2$ are also positive constants depending on the dependence measure.

Theorem 2. *Assume the assumptions of Theorem 1 are fullfilled. Moreover we assume,*

$$3\Delta^{-2\varphi}\Delta^{2\beta_0}S_t^2 < \psi,$$
$$6L_t^2\Delta^{-2\varphi}\varphi|\log\Delta| < \psi,$$
$$\mu^{-\tau/2} < \widetilde{C}\Delta^{2\varphi}\psi\Delta^{2H_0},$$

for some $\varphi, \psi \in (0,1)$ and for some positive constant $\widetilde{C}$ depending on B, $\overline{a}_0$ and τ. Then for any enough large μ, we have

$$\mathbb{P}\left(\left|\widehat{L_t^2} - L_t^2\right| > \psi\right) \leq \frac{\mathfrak{g}_1}{N\psi^2\Delta^{4H_0+4\varphi}} + \frac{\mathfrak{f}_1}{N\varphi^2\Delta^{4H_0}} + 4\mathfrak{b}\exp\left(-\mathfrak{f}_2 N\varphi^2\Delta^{4H_0}\right) + 2\mathfrak{b}\exp\left(-\mathfrak{g}_2 N\psi^2\Delta^{4H_0+4\varphi}\right),$$

where $\mathfrak{b}$ is a positive constant and $\mathfrak{f}_1$, $\mathfrak{f}_2$, $\mathfrak{l}_1$ and $\mathfrak{l}_2$ are also positive constants depending on the dependence measure.

Following Golovkine et al. (2022) we choose Δ, φ and ψ as function of the mean number μ of observation points per curve :

$$\Delta(\mu) = \exp\left(-(\log\mu)^\gamma\right), \quad \text{where } \gamma \in (0,1] \quad \text{is fixed}, \tag{6}$$
$$\varphi(\mu) = C_\varphi\left(\log\mu\right)^{-2}, \tag{7}$$
$$\psi(\mu) = C_\psi\left(\log\mu\right)^{-1}, \tag{8}$$

where C_φ and C_ψ are some positive constants.

6 Illustration and discussion

We simulate a FAR(1) (see (3)) where $\{\xi_n\}$ are i.i.d. 'tied-down' *multifractional Brownian motion* (see Stoev and Taqqu (2006)) paths with a logistic function for H_t, a constant function L_t^2, (see Figure 1), and the kernel function $\psi(s,t) = (9/4)st$. The distribution of the additive error noise is $\varepsilon \sim \mathcal{N}(0, 0.04)$.

Figure 2 illustrates that our estimates concentrate arround the true values as the sample size N and the mean number μ of observation points per curve increase. The concentration bounds established in Theorem 1 and Theorem 2 provide guidance for calibrating hyper-parameters such as Δ, φ and ψ. Equations (6) to (8) propose choices of these quantities to obtain a concentration of local regularity estimators. For each N, the pre-smoothing bandwidth of the trajectories has been chosen as the median of the bandwidths obtained by cross-validation on a sample of 20 curves. To choose the parameter γ of the Δ expression, a grid of Δ values was tested. It appears that Δ decreases well when μ is large and that a $\gamma \approx 1/2$ gives satisfying results. However, this choice would depend on the data and further investigations are necessary, especially as this choice affects the quality of the estimates.

With good estimates of H_t and L_t^2, the next step is to determine optimal bandwidths to estimate the mean and autocovariance functions using the "smooth first, estimate later"

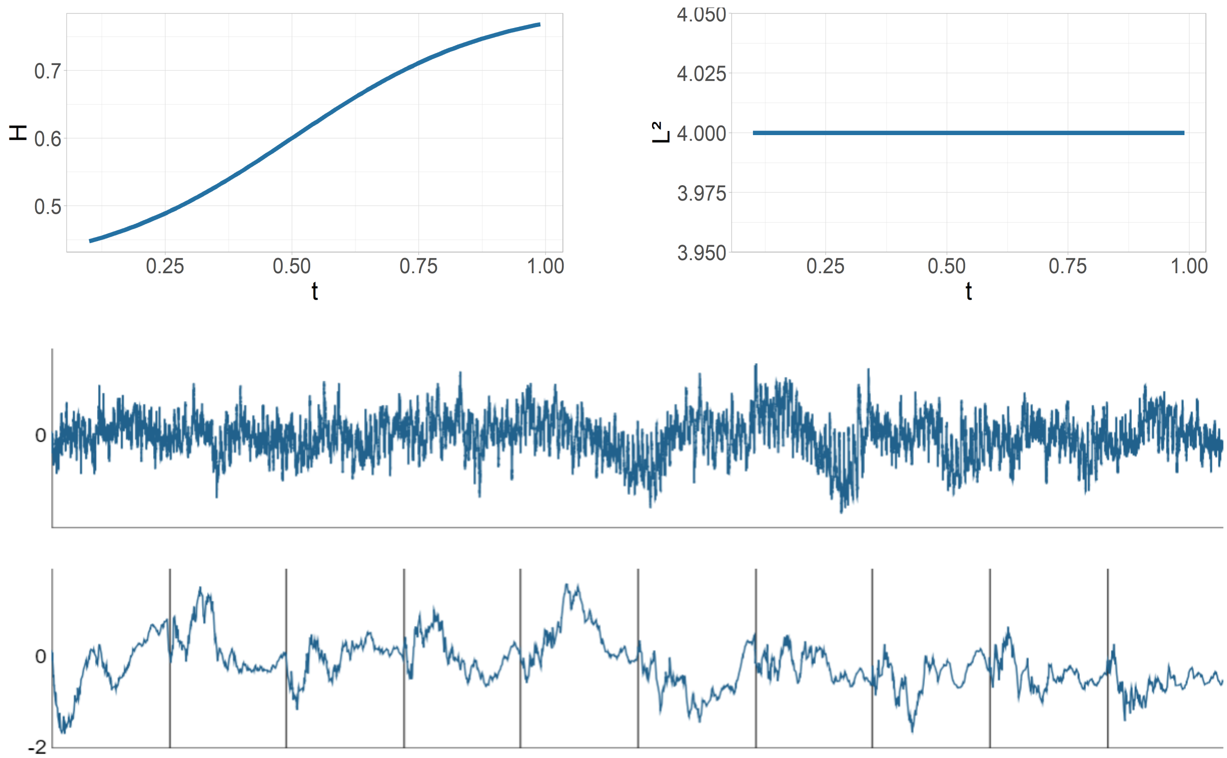

Figure 1: Overview of simulated data. **Top left:** the true H_t function. **Top right:** The true L_t^2 function, here $L_t^2 = 4$. **Middle:** Sample of $N = 250$ functions from a simulated FAR(1) without error. **Bottom:** The last ten functions of the middle graph.

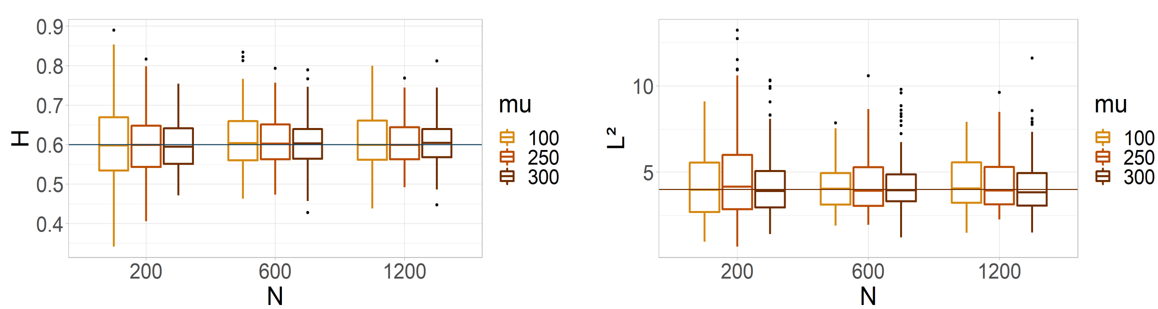

Figure 2: Empirical distribution of the regularity parameters the estimatorsx at $t = 1/2$ according to the sample size N and the mean number μ of observation times per curve over 50 replications. **Left:** Boxplots of $\widehat{H}_t$, the true value is $H_t = 0.6$. **Right:** Boxplots of $\widehat{L}_t^2$, the true value is $L_t^2 = 4$.

method or the “two-step procedure”. These approaches have been considered by, among others, Golovkine et al. (2021), Hall et al. (2006) and Zhang and Chen (2007). The final goal is to make inferences on electricity production curves of wind or photovoltaic farms which are very irregular.

References

Golovkine, S., Klutchnikoff, N., and Patilea, V. (2021). Adaptive optimal estimation of irregular mean and covariance functions. *arXiv preprint arXiv:2108.06507*.

Golovkine, S., Klutchnikoff, N., and Patilea, V. (2022). Learning the smoothness of noisy curves with application to online curve estimation. *Electronic Journal of Statistics*, 16(1):1485–1560.

Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.*, 34(3):1493–1517.

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*. Texts in Statistical Science Series. CRC Press, Boca Raton, FL.

Rubín, T. and Panaretos, V. M. (2020). Sparsely observed functional time series: estimation and prediction. *Electronic Journal of Statistics*, 14(1):1137 – 1210.

Stoev, S. A. and Taqqu, M. S. (2006). How rich is the class of multifractional brownian motions? *Stochastic Processes and their Applications*, 116(2):200–221.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American statistical association*, 100(470):577–590.

Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *Ann. Statist.*, 35(3).