Fast and Asymptotically-efficient estimation in a Fractional autoregressive process

Samir Ben Hariz, Alexandre Brouste, Chunhao Cai, Marius Soltane

FAST AND ASYMPTOTICALLY-EFFICIENT ESTIMATION IN A FRACTIONAL AUTOREGRESSIVE PROCESS

SAMIR BEN HARIZ, ALEXANDRE BROUSTE, CHUNHAO CAI, AND MARIUS SOLTANE

Abstract. This paper considers the joint estimation of the parameters of a first-order fractional autoregressive model by constructing an initial estimator with convergence speed lower than $\sqrt{n}$ and singular asymptotic joint distribution. The one-step procedure is then used in order to obtain an asymptotically-efficient estimator. This estimator is computed faster than the maximum likelihood or Whittle estimator and therefore allows for faster inference on large samples. The paper illustrates the performance of this method on finite-size samples via Monte Carlo simulations.

1. Introduction

Parametric estimation in fractional Gaussian processes has been widely studied over recent decades. The asymptotic properties of the maximum likelihood estimator (MLE) have been derived under conditions of regularity on the spectral density of the process [Fox and Taqqu, 1986, Dahlhaus, 1989, Dahlhaus, 2006, Lieberman et al., 2012]. Moreover, for a stationary Gaussian process observed in the large sample scheme, the local asymptotic normality (LAN) of the likelihood ratio was derived by [Cohen et al., 2013]. In this setting, the MLE is asymptotically efficient but is not in a closed form, such that numerical optimization of the likelihood is then necessary to compute the estimate for an observation sample. This step is particularly time consuming and numerically unstable. It is therefore worth seeking an alternative estimator which can be computed faster and keeps efficient asymptotic properties.

To achieve this, we extend the method presented in [Le Cam, 1956] for independent and identically-distributed random variables. Under this procedure, a single Fisher scoring step on the loglikelihood is applied, starting from an initial $\sqrt{n}$-consistent estimator. This is performed in order to obtain a new estimator (called the one-step estimator) whose asymptotic variance is optimal in the Fisher sense. This procedure has already been successfully extended for diffusion processes [Kamatani and Masayuki, 2015, Gloter and Yoshida, 2021], ergodic Markov chains [Kutoyants and Motrunich, 2016], fractional Gaussian noise observed at high frequency [Brouste et al., 2020b] and stable noise [Brouste and Masuda, 2018], for instance. Generally, we derive the asymptotic properties of the one-step estimator under the Sweeting conditions [Sweeting, 1980] and the properties of the initial guess estimator depend on the statistical experiment which is considered. It should also be noted that the one-step procedure has only recently been extended to $n^{\frac{\delta}{2}}$-consistent initial guess estimators by [Kamatani and Masayuki, 2015].

In this paper, we consider a special case of fractional Gaussian process (X_t), satisfying for any $t \in \mathbb{Z}$ the recursive relation

$$X_t = \alpha X_{t-1} + \sqrt{\sigma_2}\varepsilon_t^H \tag{1.1}$$

where $|\alpha| < 1$, $\sigma_2 > 0$ and (ε_t^H) is the fractional Gaussian noise of the Hurst exponent, $H \in]0;1[$. The process (ε_t^H) is a stationary Gaussian sequence with autocovariance function

$$\rho(k) = \frac{1}{2}\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right). \tag{1.2}$$

Key words and phrases. Geweke Porter-Hudak estimator, LAN property, Log periodogram, Maximum Likelihood estimator, One-step estimator, Quadratic form of Gaussian process, Semi-parametric estimation, Toeplitz matrix, Whittle estimator.

The estimation of the parameter α in the autoregressive process directed by a fractional Gaussian noise (1.1) has been studied by [Brouste et al., 2014, Soltane, 2018]. In these papers, the Hurst exponent, H, is assumed to be known and the asymptotic properties of the MLE are studied. Generalized least square estimators (GLSEs) of α were studied in [Esstafa, 2019] and the Hurst exponent was again assumed to be known. This type of process has therefore been widely used to model realized volatility in stochastic volatility models (see [Gatheral et al., 2018] and the reference therein). In this application, large datasets are treated conducting to the impossibility of carrying out numerical optimization for the MLE. This study therefore extends the one-step procedure in this context.

As is demonstrated in Section 2, the covariance function of the process X is not in a closed form, such that the computation of the MLE is time consuming and the MLE emerges numerically unstable. In this paper, an initial guess semi-parametric estimator of H is considered, as inspired by the work of [Hurvich et al., 1998] for fractional integrated Gaussian processes. The paper then considers an estimator of α by plugging-in the estimate of H into the GLSE presented in [Esstafa, 2019]. It then estimates σ_2 via the residual process in the same spirit as a standard AR model by the same plug-in. The unexpected result of this method is that the asymptotic joint law of the initial estimators is a singular Gaussian vector. It is then evidenced that this derives from the statistical error of the last two estimators (of α and σ_2), being a function of the statistical error of H modulo a negligible remainder term.

As the aforementioned initial estimator has a slow convergence speed and as it is difficult to construct a joint asymptotic confidence region for this initial estimate because of the singularity, the one-step procedure is applied on the initial estimate in order to build asymptotically-efficient estimator. To obtain the asymptotic properties of an the one-step estimator with rate improvement, we need stronger conditions than those of Sweeting [Sweeting, 1980], which in our case are estimates of the convergence rate of Toeplitz's matrix trace, as described in [Lieberman et al., 2012]. We formulate a generic condition on the spectral density to apply the results of [Cohen et al., 2013] and [Lieberman et al., 2012] in the proof of the asymptotic properties of the one-step estimator.

The paper is organized as follows. Section 2 presents the main results, divided into two subsections: the first presents the initial estimator and states its asymptotic properties, and the second presents the one-step procedure and states its asymptotic properties. Section 3 displays the performance of the estimators on finite-size samples via the Monte Carlo simulations presented. Section 4 concludes and considers the perspectives raised by this research. 5 provides the numerous technical lemmas necessary to prove the main results. Section 6 lists proofs of the results presented in Section 2.

2. Main results

2.1. **Initial estimator.** The spectral density of the fractional Gaussian noise $\left(\varepsilon_t^H\right)$ is given by

$$f_H(\lambda) = C_{H,\sigma_2}\, 2\,(1-\cos\lambda) \sum_{k\in\mathbb{Z}} \frac{1}{|\lambda + 2k\pi|^{2H+1}}, \tag{2.1}$$

where $C_{H,\sigma_2} = \frac{\sigma_2}{2\pi}\Gamma(2H+1)\sin(\pi H)$ and $\lambda \in [-\pi;\pi]$. Thus, we consider the following parametric space

$$\vartheta = (H,\alpha,\sigma_2) \in \Theta = [a;b]\times]-1;1[\times]0;+\infty[, \tag{2.2}$$

where $[a,b]$ is any compact set in $]0;1[$. Therefore, the spectral density of the process (X_t) is given by the following proposition.

Proposition 2.1. *Letting g_{H,α,σ_2} be the spectral density of (X_t), then*

$$g_\vartheta(\lambda) = g_{H,\alpha,\sigma_2}(\lambda) = \left(1 - 2\alpha\cos(\lambda) + \alpha^2\right)^{-1} f_H(\lambda) \tag{2.3}$$

Proof. See Section 6.1. □

We estimate H via the log-periodogram method presented in [Geweke and Porter-Hudak, 1983, Hurvich et al., 1998] for general integrated Gaussian series (GPH estimator). The choice of this method is motivated by Theorem 1, as provided in [Geweke and Porter-Hudak, 1983], which states that a general integrated Gaussian series of memory parameter $d = H - \frac{1}{2}$ can be represented (via its spectral density) by a general fractional Gaussian noise of Hurst exponent H and vice versa. The addition of the autoregressive part will not disturb this method, as the spectral density of a standard autoregressive process is regular and bounded. The spectral density of (X_t) will be the product of that of $\left(\varepsilon_t^H\right)$ by the autoregressive part.

Letting $(X_1, X_2, \ldots, X_n)$ be an observation sample generated via the relation (1.1) and considering an integer m satisfying $m < n.$, we define

$$I(\lambda) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} X_t \exp(it\lambda)\right|^2 \tag{2.4}$$

$$\lambda_j = \frac{2\pi j}{n} \text{ where } j \in \{1, \ldots, m\} \tag{2.5}$$

$$a_j = \log\left(2\sin\left(\frac{\lambda_j}{2}\right)\right),\ \overline{a}_m = \frac{1}{m}\sum_{j=1}^{m} a_j,\ S_m = \sum_{j=1}^{m}(a_j - \overline{a}_m)^2.$$

We estimate d by regressing $\log I(\lambda_j)$ with respect to a_j, such that

$$\widehat{d}_n = -\frac{1}{2S_m}\sum_{j=1}^{m}(a_j - \overline{a}_m)\log I(\lambda_j). \tag{2.6}$$

The estimator $\widehat{H}_n$ of H is defined by

$$\widehat{H}_n = \widehat{d}_n + \frac{1}{2}. \tag{2.7}$$

***Remark* 2.1.** *Another semi-parametric method based on the log-periodogram regression is proposed by* [Robinson, 1995b]*. This method does not take into account Fourier frequencies close to* 0*, which induces a stronger bias in the estimation of H than the GPH estimator.*

***Remark* 2.2.** *For the discretely observed fractional Ornstein-Uhlenbeck process (which is similar to our model), an estimator of H was studied using the variogram of* (X_t) *(see* [Brouste and Iacus, 2013]*). However, we cannot use this type of approach as we are working on a large-sample statistical experiment.*

We now return to estimate α and σ_2 by first considering a GLSE of α and then building the residual process to estimate σ_2 as in a classical autoregressive model. We let

$$\Phi_i^j = (X_i, \ldots, X_j)^* \text{ for } i \leqslant j,$$

and

$$\Gamma_n(H) = \rho(|i-j|)_{\{1\leqslant i,j\leqslant n\}}. \tag{2.8}$$

The estimators $\widehat{\alpha}_n$ and $\widehat{\sigma}_{2,n}$ are defined by

$$\widehat{\alpha}_n = \frac{\Phi_2^{n*}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{n-1}}{\Phi_1^{n-1*}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{n-1}}, \tag{2.9}$$

and

$$\widehat{\sigma}_{2,n} = \frac{1}{n-1}\left(\Phi_2^n - \widehat{\alpha}_n\Phi_1^{n-1}\right)^*\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\left(\Phi_2^n - \widehat{\alpha}_n\Phi_1^{n-1}\right), \tag{2.10}$$

respectively.

From this point we use the generic notations $\vartheta = (H, \alpha, \sigma_2)$ and its estimator $\widehat{\vartheta}_n = \left(\widehat{H}_n, \widehat{\alpha}_n, \widehat{\sigma}_{2,n}\right)$ to present our first result:

Theorem 2.1. *Letting $m = \left[n^{\delta}\right]$ for some $\frac{1}{2} < \delta < \frac{2}{3}$, the initial estimator $\widehat{\vartheta}_n$ is weakly consistent. Moreover, when $n \longrightarrow \infty$, we have,*

$$\sqrt{m}\begin{pmatrix}\widehat{H}_n - H\\ \widehat{\alpha}_n - \alpha\\ \widehat{\sigma}_{2,n} - \sigma_2\end{pmatrix} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0; \Sigma_{\vartheta}\right). \tag{2.11}$$

The covariance matrix Σ_{ϑ} is of the form

$$\Sigma_{\vartheta} = V_H \widetilde{\Sigma}_{\vartheta}, \tag{2.12}$$

where V_H is the asymptotic variance of $\sqrt{m}(\widehat{H}_n - H)$, $\frac{\pi^2}{24}$ and $\widetilde{\Sigma}_{\vartheta}$ is a built-in singular matrix (6.29).

Proof. See Section 6.2 □

***Remark* 2.3.** *It is possible to estimate H via the local Whittle method presented in* [Robinson, 1995a, Shimotsu and Phillips, 2005]. *These methods lead us to an estimator which does not adopt an explicit form but is less sensitive to the parameters and more suitable for small samples.*

***Remark* 2.4.** *Theorem 2.1 remains valid for any asymptotic Gaussian estimator of H whose rate satisfies the conditions of this Theorem. The limiting covariance matrix is then obtained by replacing V_H with the asymptotic variance of the estimator of H considered here.*

***Remark* 2.5.** *These results can be extended to the fractional $AR(p)$ case using the causal representation of an $AR(p)$ process in the vectorial case.*

***Remark* 2.6.** *Theorem 2.1 remains valid if the initial estimate of α and σ_2 is carried out on a subsample of size n^{β} with $\beta > \delta$. The convergence in law towards the singular Gaussian vector will be in this case slower than when the estimation is carried out on the whole sample.*

***Remark* 2.7.** *We can also use the method used in* [Brouste et al., 2014, Soltane, 2018] *to estimate α.*

2.2. **One-step estimator.** We assume in this subsection that (Y_n) is a stationary centred Gaussian process with spectral density f_{ϑ} for some unknown parameter $\vartheta \in \mathbb{R}^d$. We consider f_{ϑ} to satisfy regularity conditions if the following conditions are satisfied.

We let Θ be an open subset of $\mathbb{R}^d$,

- For any $\vartheta \in \Theta$, $f_{\vartheta}(\lambda)$ is three-times continuously differentiable on Θ. In addition, for any $0 \leqslant \ell \leqslant 3$ and $j_1, \ldots, j_m$, the partial derivate
$$\frac{\partial^{\ell}}{\partial\vartheta_{j_1}\ldots\partial\vartheta_{j_\ell}} f_{\vartheta}(\lambda),$$
is continuous on $\Theta \times [-\pi;\pi]\backslash\{0\}$, is continuously differentiable with respect to λ and its partial derivate
$$\frac{\partial^{\ell+1}}{\partial\lambda\partial\vartheta_{j_1}\ldots\partial\vartheta_{j_\ell}} f_{\vartheta}(\lambda),$$
and is continuous on $\Theta \times [-\pi;\pi]\backslash\{0\}$.
- There also exists a continuous function $\alpha : \Theta \longrightarrow]-1,1[$, such that for any compact set $\Theta^* \subset \Theta$ and $\delta > 0$, the following conditions hold for every $(\vartheta,\lambda) \in \Theta^* \times [-\pi;\pi]\backslash\{0\}$.
$$c_{1,\delta,\Theta^*}\left|\lambda\right|^{-\alpha(\vartheta)+\delta} \leqslant f_{\vartheta}(\lambda) \leqslant c_{2,\delta,\Theta^*}\left|\lambda\right|^{-\alpha(\vartheta)-\delta}$$
and
$$\left|\frac{\partial}{\partial\lambda} f_{\vartheta}(\lambda)\right| \leqslant c_{2,\delta,\Theta^*}\left|\lambda\right|^{-\alpha(\vartheta)-1-\delta}.$$
For any $\ell \in \{1,2,3\}$ and any $j \in (1,\ldots,m)^{\ell}$,
$$\left|\frac{\partial^{\ell}}{\partial\vartheta_{j_1}\ldots\partial_{j_\ell}} f_{\vartheta}(\lambda)\right| \leqslant c_{2,\delta,\Theta^*}\left|\lambda\right|^{-\alpha(\vartheta)-\delta},$$
where c_{i,δ,Θ^*} is some positive finite constant which only depends upon δ and Θ^*.

We now consider the local asymptotic normality property of the likelihood ratio.

Proposition 2.2. *We let ℓ_n be the log-likelihood function of a stationary Gaussian process (which can be expressed using spectral density, like in Proposition 2.1 from* [Cohen et al., 2013]*). We assume that f_ϑ satisfies the regularity conditions and let $B(\vartheta, R)$ (open ball of centre ϑ and radius R) for some $R > 0$. For any $t \in B(\vartheta, R)$,*

$$\ell_n\left(\vartheta + \frac{t}{\sqrt{n}}\right) - \ell_n(\vartheta) = t\frac{\nabla \ell_n(\vartheta)}{\sqrt{n}} - \frac{t\mathcal{I}(\vartheta)\,t^*}{2} + r_{n,\vartheta}(t),$$

where, under $\mathbb{P}_\vartheta^{(n)}$, the score function $\nabla(\cdots)$ satisfies

$$\frac{\nabla \ell_n(\vartheta)}{\sqrt{n}} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0;\mathcal{I}(\vartheta)), \tag{2.13}$$

and

$$r_{n,\vartheta}(t) \xrightarrow[n\to\infty]{a.s.} 0 \tag{2.14}$$

uniformly on each compact set. The Fisher information matrix is given in our case by

$$\mathcal{I}(\vartheta) = \frac{1}{4\pi}\left(\int_{-\pi}^{\pi} \frac{\partial \log g_\vartheta(\lambda)}{\partial \vartheta_k}\frac{\partial \log g_\vartheta(\lambda)}{\partial \vartheta_j}\,d\lambda\right)_{1\leqslant k,j\leqslant d}. \tag{2.15}$$

Proof. See Section 6.3. □

***Remark* 2.8.** *When the LAN property is verified, it is possible to define a notion of asymptotic efficiency for the estimators (see* [Ibragimov and Has'minskii, 1981]*). It is also possible in a such a statistical experiment to determine the estimators maximum speed of convergence ($\sqrt{n}$ in our case).*

***Remark* 2.9.** *Proposition 2.2 is derived from Theorem 3.4, as proposed by* [Cohen et al., 2013]*, in which weaker conditions are given for the application α. In our case, we employ this Theorem in order to derive the asymptotic properties of the one-step estimator.*

We now build an estimator whose asymptotic properties are similar to the maximum likelihood (in terms of convergence rate and asymptotic variance).

Proposition 2.3. *Fist, we assume*

- *$\widehat{\vartheta}_n^{(1)}$ is an initial estimator of ϑ, such that, for some $\frac{1}{2} < \delta \leqslant 1$,*

$$n^{\frac{\delta}{2}}\left(\widehat{\vartheta}_n^{(1)} - \vartheta\right) = O_{\mathbb{P}}(1),$$

- *f_ϑ satisfies the regularity conditions.*

We then consider the estimator

$$\widetilde{\vartheta}_n^{(1)} = \widehat{\vartheta}_n^{(1)} + \mathcal{I}\left(\widehat{\vartheta}_n^{(1)}\right)^{-1}\frac{1}{n}\nabla \ell_n\left(\widehat{\vartheta}_n^{(1)}\right) \tag{2.16}$$

where $\nabla \ell_n(\cdot)$ and $\mathcal{I}(\cdot)$ are defined in Proposition 2.2. Then,

$$\sqrt{n}\left(\widetilde{\vartheta}_n^{(1)} - \vartheta\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\left(0, \mathcal{I}(\vartheta)^{-1}\right). \tag{2.17}$$

Proof. See Section 6.4. □

***Remark* 2.10.** *To apply Proposition 2.3, it is not necessary to know the asymptotic distribution of the initial estimator but only a convergence rate of statistical errors. When the convergence rate of the initial estimator is too slow, it is possible to apply a multi-step procedure to obtain an asymptotically-efficient estimator at optimal speed.*

***Remark* 2.11.** *The one-step estimator can be easily extended to other classes of stationary Gaussian processes, satisfying the assumptions of Proposition 2.3.*

***Remark* 2.12.** *In the Gaussian setting, we can use a similar one-step procedure, replacing the classical score by the Whittle score. This procedure is faster to compute and does not require the inversion of the covariance matrix. However, we leave analysis of this method for future studies.*

Theorem 2.2. *The initial estimator $\widehat{\vartheta}_n$ satisfies the conditions of Theorem 2.3 and the spectral density g_ϑ meets the regularity conditions. Proposition 2.3 allows us to deduce that the one-step estimator of $\widehat{\vartheta}_n$ is asymptotically efficient with speed $\sqrt{n}$.*

3. Simulations study

The log-likelihood function produced by the sample $\mathbf{X}^{(n)} = (X_1, X_2, \ldots, X_n)^*$ which satisfies the recursive relation (1.1) is given by

$$\ell_n(\vartheta) = -\frac{1}{2}\log\left(det\left(\Gamma_n(\vartheta)\right)\right) - \frac{1}{2}\mathbf{X}^{(n)^*}\Gamma_n^{-1}(\vartheta)\,\mathbf{X}^{(n)},$$

where $\Gamma_n(\vartheta)$ is the covariance matrix of $\mathbf{X}^{(n)}$. For any $k \in \mathbb{N}$,

$$\mathbb{E}\left(X_0 X_k\right) = \int_{-\pi}^{\pi} \exp\left(ik\lambda\right) g_\vartheta(\lambda)\, d\lambda$$

The score function with respect to ϑ for $i \in \{1, 2, 3\}$ is given by

$$\frac{\partial \ell_n(\vartheta)}{\partial \vartheta_i} = -\frac{1}{2} Tr\left(\Gamma_n^{-1}(\vartheta)\frac{\partial}{\vartheta_i}\Gamma_n(\vartheta)\right) + \frac{1}{2}\mathbf{X}^{(n)^*}\Gamma_n^{-1}(\vartheta)\frac{\partial}{\partial\vartheta_i}\Gamma_n(\vartheta)\,\Gamma_n^{-1}(\vartheta)\,\mathbf{X}^{(n)},$$

where Tr is the trace operator and the Fisher information matrix is defined by (2.15). To compute the score, as well as the Fisher information, we numerically evaluate the integral of the spectral density (and its derivatives) using the Paxson method described in [Fukasawa and Tetsuya, 2019].

For each set of parameters, we perform $M = 10000$ Monte Carlo simulations for samples of size $n = 1000$. The number of Fourier frequencies for the initial estimation is fixed at $m = \left[n^{0.6}\right]$.

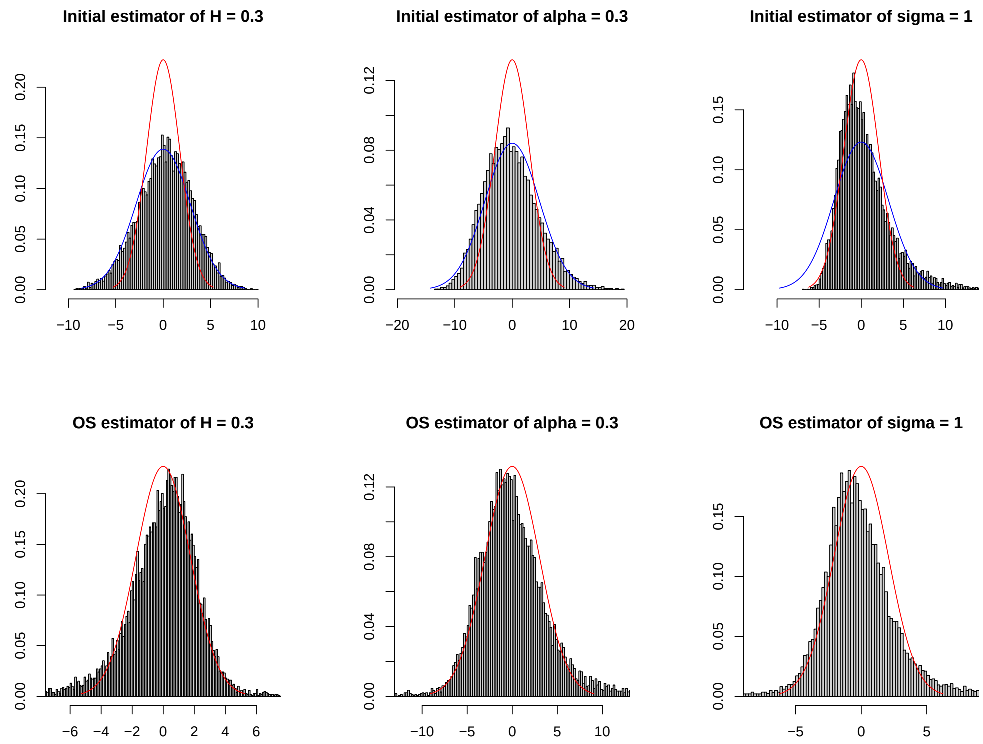

FIGURE 1. Statistical error of the estimators multiplied by $\sqrt{n}$ speed, where $\vartheta = (0.3, 0.3, 1)$. The first line corresponds to the initial estimator and the second line to the one-step estimator. The blue curves correspond to the density of the centred normal distribution where the theoretical variance is approximated by the empirical variance. The curves in red correspond to the limit law of the one-step estimator.

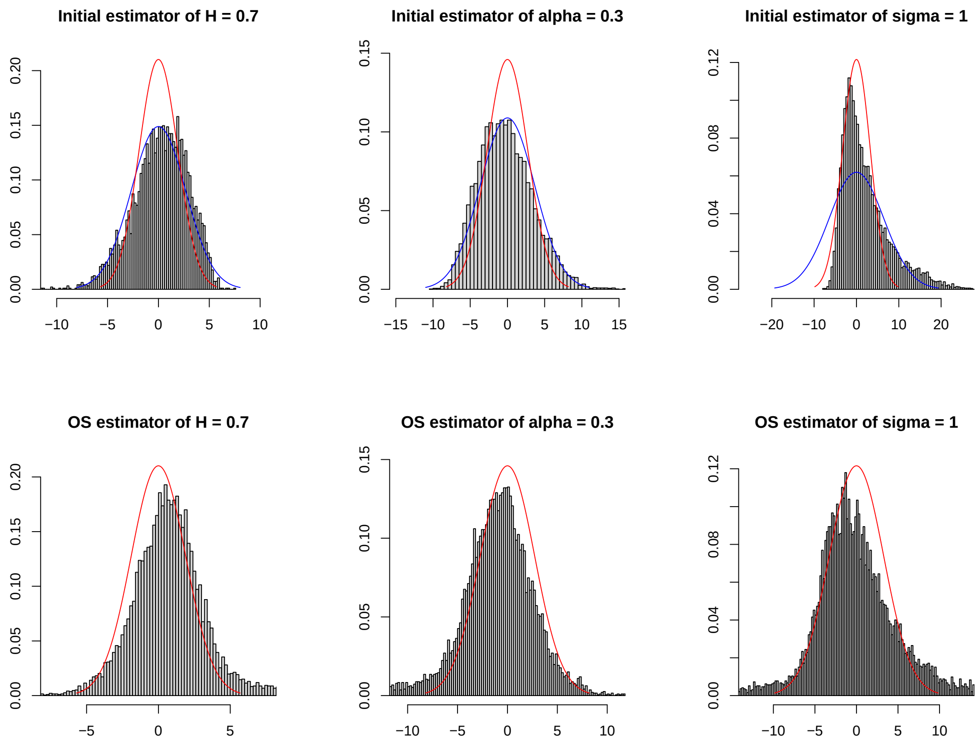

FIGURE 2. Statistical error of the estimators multiplied by $\sqrt{n}$ speed, where $\vartheta = (0.7, 0.3, 1)$. The first line corresponds to the initial estimator and the second line to the one-step estimator. The blue curves correspond to the density of the centred normal distribution where the theoretical variance is approximated by the empirical variance. The curves in red correspond to the limit law of the one-step estimator.

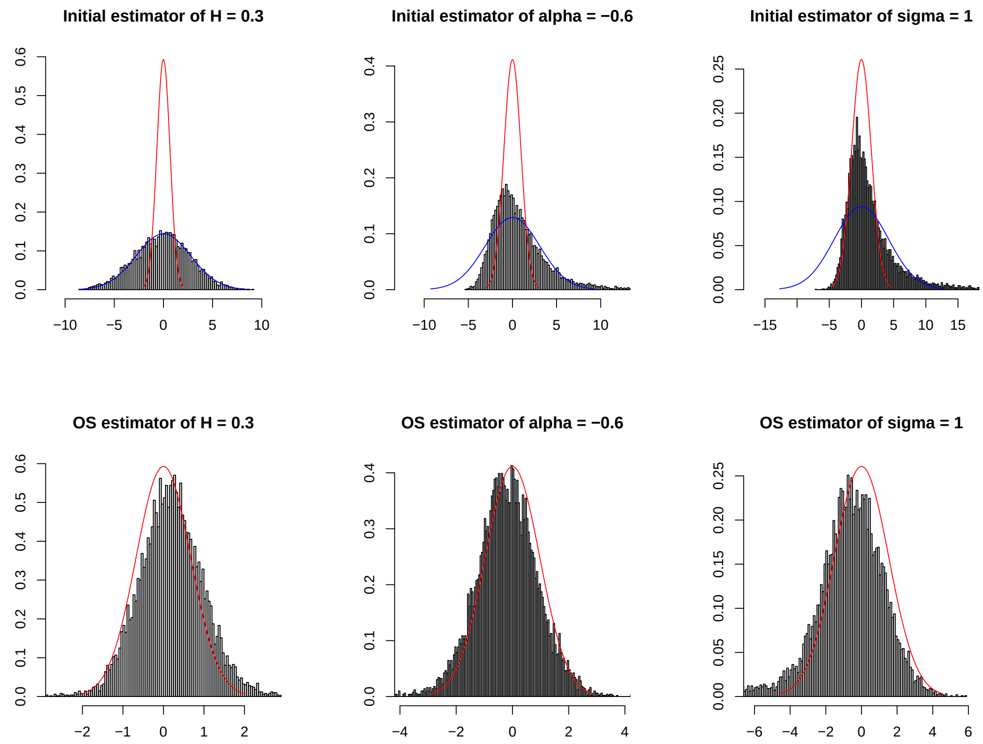

FIGURE 3. Statistical error of the estimators multiplied by $\sqrt{n}$ speed, where $\vartheta = (0.3, -0.6, 1)$. The first line corresponds to the initial estimator and the second line to the one-step estimator. The blue curves correspond to the density of the centred normal distribution where the theoretical variance is approximated by the empirical variance. The curves in red correspond to the limit law of the one-step estimator.

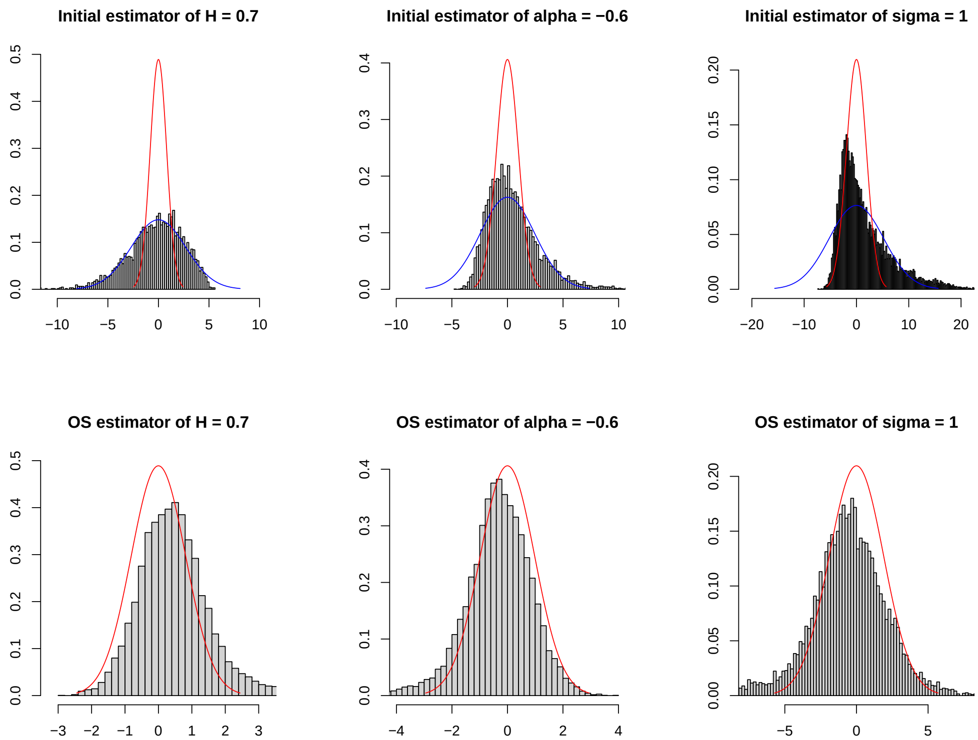

FIGURE 4. Statistical error of the estimators multiplied by $\sqrt{n}$ speed, where $\vartheta = (0.7, -0.6, 1)$. The first line corresponds to the initial estimator and the second line to the one-step estimator. The blue curves correspond to the density of the centred normal distribution where the theoretical variance is approximated by the empirical variance. The curves in red correspond to the limit law of the one-step estimator.

4. Conclusion and perspectives

In this paper, we propose a simple and easy-to-implement estimator for the parameters of an $AR(1)$ process with dependent errors. We prove the joint asymptotic normality of the vector of estimators to a Gaussian degenerate law. Furthermore, we observe that the Hurst parameter estimators drive and dominate other estimators.

Using a gradient descent approach, we improve on the first estimator with one iteration. The resulting estimator, called one step, achieves the $\sqrt{n}$ rate with optimal variance. Hence, we reproduce the performance of the maximum likelihood estimator. As such, not only does the one-step process improve upon the rate of the initial estimator, but it also achieves the inverse of the Fisher information as a limiting variance. This is especially relevant when dealing with large sample sizes.

Our result can be extended to other process classes, including ARFIMA models. Furthermore, it is notable that the initial poor estimator could be replaced by anyone with a sufficient speed of convergence.

In addition, a procedure for detecting breaks in parameters, as achieved by [Brouste et al., 2020a], could be developed for the autoregressive parameter. On the application level, we could implement functions in R that perform this estimation, as managed by the OneStep package described in [Brouste et al.,].

Acknowledgments

This research benefited from the support of the Chair Risques Emergents ou Atypiques en Assurance, under the sponsorship of Fondation du Risque, a joint initiative by Le Mans University, Ecole Polytechnique and MMA, a member of the Covea Group.

5. Auxiliary results

For clarity, we separate the technical results into two subsections. The first details the technical lemmas related to the initial estimator and the second details those related to the one-step estimator.

5.1. Technical results related to Section 2.1.

Lemma 5.1. *For any $t \in \mathbb{Z}$, the process*

$$X_t = \sum_{j=0}^{+\infty} \alpha^j \varepsilon_{t-j}^H, \text{ a.s.} \tag{5.1}$$

is a stationary and ergodic process.

Proof. As we know, for any time series $\{Y_t\}$ under the monotone convergence theorem, we have

$$\mathbb{E}\left(\sum_{t=-\infty}^{\infty} |Y_t|\right) = \lim_{n\to\infty} \mathbb{E}\left(\sum_{t=-n}^{n} |Y_t|\right) = \sum_{t=-\infty}^{+\infty} \mathbb{E}|Y_t|. \tag{5.2}$$

With this equation and the Schwarz inequality we have

$$\mathbb{E}\left(\sum_{j=0}^{+\infty} \left|\alpha^j \varepsilon_{t-j}^H\right|\right) = \sum_{t=0}^{\infty} |\alpha^j| \mathbb{E}|\varepsilon_{t-j}^H| \leq C \sum_{t=0}^{\infty} |\alpha^j| < \infty.$$

This demonstrates that the right side of Equation 5.1 converges absolutely and thus converges a.s.. Considering that

$$\left|\sum_{j=0}^{n} \alpha^j \varepsilon_{t-j}^H\right| \leq \sum_{j=0}^{\infty} \left|\alpha^j \varepsilon_{t-j}^H\right|,$$

with the dominated convergence theorem,

$$\mathbb{E}\left(X_t\right) = \lim_{n\to\infty} \mathbb{E}\left(\sum_{j=0}^{n} \alpha^j \varepsilon_{t-j}^H\right) = 0.$$

Now, for $t, s \in \mathbb{Z}$ we define

$$\xi_n = \sum_{j=0}^{n} \alpha^j \varepsilon_{t-j}^H,\ \eta_n = \sum_{k=0}^{n} \alpha^k \varepsilon_{s-k}^H.$$

From the previous conclusion, we have $\xi_n \eta_n \to X_t X_s$ a.s. and $|\xi_n \eta_n| \leq V$, where V is defined by

$$V = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \left|\alpha^j \alpha^k \varepsilon_{t-j}^H \varepsilon_{s-k}^H\right|.$$

Now, from Equation (5.2) we know that

$$\mathbb{E}\left(V\right) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} |\alpha^j \alpha^k| \mathbb{E}|\varepsilon_{t-j}^H \varepsilon_{s-k}^H| \leq \sigma^2 \left(\sum_{j=0}^{\infty} |\alpha^j|\right)^2 < \infty.$$

Finally, Theorem 5.3.8 in [Stout, 1974] ensures that (X_t) is stationary and ergodic. ∎

Lemma 5.2. *The process $\{X_t\}$ defined in (5.1) is a Gaussian process.*

Proof. Here, we only need to prove that for any $m \in \mathbb{N}_+$ we have the vector

$$\mathbf{X} = (X_1, X_2, \cdots, X_m)^T \sim \mathcal{N}(\mathbf{0}, \Sigma_m), \tag{5.3}$$

where $\Sigma_m = (\gamma_{j-k})_{m \times m}$ and γ_k is the auto-covariance of X_t and X_{t+k}. In fact, if we define

$$\eta_k(n) = \sum_{j=0}^{n} \alpha^j \varepsilon_{k-j}^H,$$

then it follows that

$$\mathbb{E}|\eta_k(n) - X_k| \to 0,\ n \to \infty.$$

Now, for any real vector $\mathbf{b} = (b_1, b_2, \cdots, b_m)^T$, we define

$$Y = \mathbf{b}^T \mathbf{X} = \sum_{k=1}^{m} b_k X_k,\ \eta_n = \sum_{k=1}^{m} b_k \eta_k(n).$$

When $n \to \infty$, we have

$$\mathbb{E}(|Y - \eta_n|) = \mathbb{E}\left(\left|\sum_{k=1}^{m} b_k (X_k - \eta_k(n))\right|\right) \leq \sum_{k=1}^{m} |b_k| \mathbb{E}[|X_k - \eta_k(n)|] \to 0,$$

where L^1 convergence implies convergence in distribution, such that η_n converges in distribution to Y. As η_n satisfies the normal distribution and $Y \sim \mathcal{N}(\mathbb{E}(Y), \mathbb{V}ar(Y))$, where $\mathbb{E}(Y) = 0$ and $\mathbb{V}ar(Y) = \mathbf{b}^T \Sigma_m \mathbf{b}$, we have the conclusion of (5.3). ☐

Lemma 5.3. *Letting $Y_n \in \mathbb{R}^n$ be a centred Gaussian vector with covariance matrix Σ_n and any symmetric matrix A_n of size $n \times n$, then*

$$\phi(Y_n) = Y_n^* A_n Y_n.$$

As a result,

$$\mathbb{E}\left(\phi(Y_n)\right) = Tr\left(A_n \Sigma_n\right)\ \text{ and } \mathbb{V}ar\left(\phi(Y_n)\right) = 2Tr\left(\left(A_n \Sigma_n\right)^2\right).$$

Proof. See Lemma A1 of [Cohen et al., 2013]. ☐

Lemma 5.4. *Under the hypothesis on the parametric space we have the following conclusions:*

1) *For any $H \in [a;b]$ and $j \in \{0,1,2,3\}$, $\frac{\partial}{\partial \lambda} \frac{\partial^j}{\partial^j H} f_H(\lambda)$ are continuous functions on $[\,a;b] \times [-\pi;\pi], \backslash\{0\}$.*
2) *For any $j \in \{0,1,2,3\}$ the functions $\frac{\partial^j}{\partial^j H} f_H(\lambda)$ are symmetric with respect to λ.*
3) *For any $\delta > 0$ and all $(H,\lambda) \in [a;b] \times [-\pi;\pi]\backslash\{0\}$,*
   a) *$K_{1,\delta}|\lambda|^{1-2H+\delta} \leqslant f_H(\lambda) \leqslant K_{2,\delta}|\lambda|^{1-2H-\delta}$,*
   b) *$\left|\frac{\partial}{\partial \lambda} f_H(\lambda)\right| \leqslant K_{3,\delta}|\lambda|^{-2H-\delta}$*
   c) *and for any $j \in \{1,2,3\}$, $\left|\frac{\partial^j}{\partial^j H} f_H(\lambda)\right| \leqslant K_{4,\delta}|\lambda|^{1-2H-\delta}$.*

   Where $K_{i,\delta}$ are some finite positive constants which only depend upon δ and the parametric space of H, namely $[\,a;b]$.,
4) *$\frac{\partial^j \rho(k)}{\partial^j H} = \int_{-\pi}^{\pi} \exp(ik\lambda) \frac{\partial^j f(\lambda)}{\partial^j H}\, d\lambda$*

Proof. **We start by proving the assertion *3) a).*** We write

$$f_H(\lambda) = C_{H,\sigma_2}(1 - \cos\lambda)\left(\frac{1}{|\lambda|^{2H+1}} + \sum_{k \in \mathbb{Z}^*} f_k(\lambda)\right),$$

where

$$f_k(\lambda) = \frac{1}{|\lambda + 2k\pi|^{2H+1}}.$$

We observe that

$$f_k(\lambda) \leqslant C_H |k|^{-1-2H}$$

and so $\sum_{k \in \mathbb{Z}^*} f_k(\lambda)$ converge to a continuous bounded function uniformly in $[-\pi;\pi]$. Then, for some $C > 0$

$$C_{H,\sigma_2}(1 - \cos\lambda)\left(\frac{1}{|\lambda|^{2H+1}} - C\right) \leqslant f_H(\lambda) \leqslant C_{H,\sigma_2}(1 - \cos\lambda)\left(\frac{1}{|\lambda|^{2H+1}} + C\right).$$

As $\lambda^2 = O\left(|\lambda|^{1-2H}\right)$, we have the conclusion of ***3)a)***.

Assertion ***3)b)*** is demonstrated in the same way.

Now we prove assertion *3) c)*. For all $\lambda \in [-\pi;\pi]$, f_k is continuous with respect to λ and H with
$$\frac{\partial^j}{\partial^j H} f_k(\lambda) = \frac{\prod_{\ell=1}^{j}\left(-2\log\left(|\lambda+2k\pi|\right)\right)}{|\lambda+2k\pi|^{2H+1}}.$$
For all $\delta_1 > 0$ there exists a positive constant C_{δ_1}, such that, for a sufficiently large k,
$$\left|\frac{\partial^j}{\partial^j H} f_k(\lambda)\right| \leqslant C_{\delta_1}\frac{k^{\delta_1}}{k^{2a+1}} = O\left(k^{-1-2a+\delta_1}\right).$$
Choosing δ_1 such that $\delta_1 - 2a < 0$, then $\sum_{k\in\mathbb{Z}^*}\frac{\partial^j f_k(\lambda)}{\partial^j H}$ converge uniformly in (λ, H) to a continuous, bounded function, and
$$(1-\cos\lambda)\frac{\partial^j}{\partial^j H}\left(\sum_{k\in\mathbb{Z}^*} f_k(\lambda)\right) = (1-\cos\lambda)\sum_{k\in\mathbb{Z}^*}\frac{\partial^j}{\partial^j H} f_k(\lambda) = O(\lambda^2).$$
In the same way, for almost all λ,
$$\frac{\partial^j}{\partial^j H}\left(\frac{1-\cos\lambda}{|\lambda|^{2H+1}}\right) = \frac{(1-\cos\lambda)\prod_{\ell=1}^{j} -2\log(|\lambda|)}{|\lambda|^{2H+1}} = O\left(|\lambda|^{1-2b-\delta_2}\right)$$
for any $\delta_2 > 0$ and sufficiently small λ. Note that the constant C_{H,σ_2} depends on H very regularly and therefore does not affect the next estimates. We now choose δ_2 such that $2b + \delta_2 < 2$ in order to produce a positive function, h. As such, h is integrable with respect to λ and, for any $j \in \{1,2,3\}$,
$$\left|\frac{\partial^j}{\partial^j H} f_H(\lambda)\right| \leqslant h(\lambda) \text{ and } \left|\frac{\partial^j}{\partial^j H} f_H(\lambda)\right| \leqslant K_{4,\delta}|\lambda|^{1-2H-\delta} \text{ for any } \delta > 0.$$

Assertion ***1)*** is demonstrated using uniform convergence, as in the proof of **3)a)** and **3) c)**.

Assertion ***2)*** is demonstrated by nothing that, for any $j \in \{0,1,2,3\}$, $\frac{\partial^j}{\partial^j H} f_k(-\lambda) = \frac{\partial^j}{\partial^j H} f_{-k}(\lambda)$.

Assertion ***4)*** is demonstrated using the derivation theorem under the integral, which is possible because of the last estimation on the proof of ***3)c)***. ☐

Lemma 5.5. *We let g be a function defined on $[-\pi;\pi]$, such that*

$$g(\lambda) = \left|\sin\left(\frac{\lambda}{2}\right)\right|^{1-2H} g^*(\lambda). \tag{5.4}$$

We assume that

1) $g^*(\lambda) > 0\ \forall\lambda \in [-\pi;\pi]$.
2) g^* *is continuous and bounded in a neighbourhood of* 0.
3) $\frac{\partial}{\partial\lambda} g^*$ *is well defined, continuous and bounded in a neighbourhood of* 0.
4) $m \longrightarrow \infty, n \longrightarrow \infty$ *with* $\frac{m}{n} \longrightarrow 0$ *and* $\frac{m\log m}{n} \longrightarrow 0$.

Then,

$$\left|-\frac{1}{2S_m}\sum_{j=1}^{m}(a_j - \overline{a}_m)\log g^*(\lambda_j)\right| = O\left(\frac{m}{n}\right). \tag{5.5}$$

Proof. This proof is based on that of Lemma 1 in [Hurvich et al., 1998]. We have
$$-\frac{1}{2S_m}\sum_{j=1}^{m}(a_j - \overline{a}_m)\log g^*(\lambda_j) = -\frac{1}{2S_m}\sum_{j=1}^{m}(a_j - \overline{a}_m)\frac{\frac{\partial}{\partial\lambda} g^*(\omega_j)}{g^*(\omega_j)}\lambda_j,$$

where for all $j \in \{1, \ldots, m\}$, $0 < \omega_j < \lambda_j$. We can find a finite constant, such that $\left|\frac{\frac{\partial}{\partial \lambda} g^*(\omega_j)}{g^*(\omega_j)}\right| \leqslant C$ for all j. The Cauchy-Schwarz inequality implies that

$$\frac{1}{2S_m}\left|\sum_{j=1}^{m}(a_j - \overline{a}_m)\, C\lambda_j\right| \leqslant \frac{C}{2S_m}\left(\sum_{j=1}^{m}(a_j - \overline{a}_m)^2\right)^{\frac{1}{2}}\left(\sum_{j=1}^{m}\lambda_j^2\right)^{\frac{1}{2}}.$$

It is shown in Lemma 1 of [Hurvich and Beltrao, 1994] that

$$S_m = m\,(1 + o(1))\,;$$

then

$$\begin{aligned}
\frac{C}{2S_m}\left(\sum_{j=1}^{m}(a_j - \overline{a}_m)^2\right)^{\frac{1}{2}}\left(\sum_{j=1}^{m}\lambda_j^2\right)^{\frac{1}{2}} &= \frac{C}{2}\left(\frac{\sum_{j=1}^{m}\lambda_j^2}{S_m}\right)^{\frac{1}{2}} \\
&= C\left(\frac{\sum_{j=1}^{m} j^2}{m(1+o(1))n^2}\right)^{\frac{1}{2}} \\
&= C\left(\frac{(m+1)(2m+1}{n^2(1+o(1))}\right)^{\frac{1}{2}} \\
&= O\left(\frac{m}{n}\right).
\end{aligned}$$

□

5.2. **Technical results related to Section 2.2.** The following lemmas are needed to derive the asymptotic properties of the one-step estimator in a general centred stationary Gaussian process. The lemmas of this subsection are valid under the regularity conditions.

Lemma 5.6. *We let $\vartheta_0 \in \Theta^*$ and $\delta > 0$, such that $B(\vartheta_0, \delta) \subset \Theta^*$. For any $\vartheta \in B(\vartheta_0, \delta)$,*

$$\|\mathcal{I}(\vartheta) - \mathcal{I}(\vartheta_0)\| \leqslant K\|\vartheta - \vartheta_0\|$$

for some constant K.

Proof. We verified Conditions(A.1) and (A.2) of [Cohen et al., 2013] during the proof of Lemma 5.4. These imply that, for any $k, j \in \{1, 2, \ldots, d\}$, with the mean value inequality,

$$\left|\frac{1}{4\pi}\left(\int_{-\pi}^{\pi}\frac{\partial \log g_\vartheta(\lambda)}{\partial \vartheta_k}\frac{\partial \log g_\vartheta(\lambda)}{\partial \vartheta_j}\,d\lambda\right) - \frac{1}{4\pi}\left(\int_{-\pi}^{\pi}\frac{\partial \log g_{\vartheta_0}(\lambda)}{\partial \vartheta_{0,k}}\frac{\partial \log g_{\vartheta_0}(\lambda)}{\partial \vartheta_{0,j}}\,d\lambda\right)\right| \leqslant K_{\{k,j\}}\,\|\vartheta - \vartheta_0\|,$$

where

$$K_{\{k,j\}} = \sup_{\vartheta \in B(\vartheta_0,\delta)}\left\|\left(\frac{\partial}{\partial \vartheta_i}\left(\int_{-\pi}^{\pi}\frac{\partial \log g_\vartheta(\lambda)}{\partial \vartheta_k}\frac{\partial \log g_\vartheta(\lambda)}{\partial \vartheta_j}\,d\lambda\right)\right)_{1\leqslant i\leqslant d}\right\|.$$

Note that $B(\vartheta_0, \delta)$ is a convex subset of $\mathbb{R}^3$. Owing to Conditions (A.1) and (A.2) of [Cohen et al., 2013] that for any $k, j \in \{1, 2, \ldots, d\}$, $K_{\{k,j\}} < \infty$, the proof is finished. □

Lemma 5.7. *For any ϑ in Θ, under the law induced by ϑ,*

$$\frac{\Delta\ell_n(\vartheta)}{\sqrt{n}} + \sqrt{n}\mathcal{I}(\vartheta) = O_{\mathbb{P}}(1).$$

Proof. The proof of Lemma 3.6 in [Cohen et al., 2013] shows that, under the law induced by ϑ,

$$\mathbb{E}\left(\frac{\Delta\ell_n(\vartheta)}{n}\right) \xrightarrow[n\to\infty]{} -\mathcal{I}(\vartheta).$$

The above convergence is obtained from Theorem 2.3 in [Cohen et al., 2013] and we need a convergence rate in this Theorem to prove this Lemma. Lemmas 3 and 4 in [Lieberman et al., 2012] lead us to state that

$$\mathbb{E}\left(\frac{\Delta\ell_n(\vartheta)}{n}\right) + \mathcal{I}(\vartheta) = O\left(n^{-1+\delta}\right)$$

for any $\delta > 0$, and hence

$$\mathbb{E}\left(\frac{\Delta\ell_n(\vartheta)}{\sqrt{n}}\right) + \sqrt{n}\mathcal{I}(\vartheta) = O\left(n^{-\frac{1}{2}+\delta}\right).$$

Again using the proof of Lemma 3.6 in [Cohen et al., 2013], we have

$$\mathbb{V}ar\left(\frac{\Delta\ell_n(\vartheta)}{\sqrt{n}}\right) = O(1),$$

which completes the proof. □

Lemma 5.8. *We let $\left(\overline{\vartheta}_n\right)_n$ be a random sequence, such that, under the law induced by ϑ, $\overline{\vartheta}_n - \vartheta = o_{\mathbb{P}}(1)$. Then, under the law induced by ϑ, for any $\kappa > 0$,*

$$\frac{\Delta\ell_n\left(\overline{\vartheta}_n\right)}{n} - \frac{\Delta\ell_n(\vartheta)}{n} = O_{\mathbb{P}}\left(n^{\kappa}\left(\overline{\vartheta}_n - \vartheta\right)\right).$$

Proof. We let $\kappa > 0$ and $C_{\vartheta,\kappa}$ be a convex compact set which depends on κ and contains ϑ. Using the proof of Lemma 3.7 in [Cohen et al., 2013], we can state that, conditional upon $\overline{\vartheta}_n \in C_{\vartheta,\kappa}$,

$$\sup_{\overline{\vartheta}_n \in C_{\vartheta,\kappa}} \left|\frac{\partial^3}{\partial^{i_1}\vartheta_1\partial^{i_2}\vartheta_2\dots\partial^{i_d}\vartheta_d}\frac{\ell_n\left(\overline{\vartheta}_n\right)}{n^{1+\kappa}}\right| = O(1)\ a.s..$$

For any $(i_1, i_2, \dots, i_d) \in \{0,1,2,3\}^d$, such that $i_1 + i_2 + \dots + i_d = 3$. Using the mean value inequality, for some positive finite random value K, we produce

$$\mathbb{P}\left(\left\|\frac{\Delta\ell_n\left(\overline{\vartheta}_n\right)}{n} - \frac{\Delta\ell_n(\vartheta)}{n}\right\| \leqslant K n^{\kappa}\left\|\overline{\vartheta}_n - \vartheta\right\|\right) \geqslant \mathbb{P}\left(\overline{\vartheta}_n \in C_{\vartheta,\kappa}\right),$$

which finishes the proof. □

6. Proofs of the main results

6.1. **Proof of proposition 2.1.** We use Theorem 4.4.1 in [Brockwell and Davis, 1981] together with Lemma 5.1 to directly prove Proposition 2.1.

6.2. **Proof of theorem 2.1. Consistency of $\widehat{H}_n$** We start by giving the process (X_n) a representation of spectral density identical to that of Equation (1) in [Hurvich et al., 1998]. We note that $1 - \cos\lambda = 2\sin\left(\frac{\lambda}{2}\right)^2$ and we rewrite g_{H,α,σ_2} in the form

$$g_{H,\alpha,\sigma_2} = \left(2\sin\left(\frac{\lambda}{2}\right)\right)^{-2d} g^*_{H,\alpha,\sigma_2}(\lambda), \tag{6.1}$$

where $d = H - \frac{1}{2}$ and

$$g^*_{H,\alpha,\sigma_2}(\lambda) = \frac{2C_{H,\sigma_2}}{1 - 2\alpha\cos(\lambda) + \alpha^2} \times \left(\frac{\sin\left(\frac{\lambda}{2}\right)}{\frac{\lambda}{2}}\right)^{2H+1} \times \left(1 + \sum_{k\in\mathbb{Z}^*}\frac{|\lambda|^{2H+1}}{|\lambda + 2k\pi|^{2H+1}}\right). \tag{6.2}$$

We note that the first two terms of (6.2) are very regular in a neighbourhood of 0 and the last term is differentiable at first order with a bounded derivative in the neighbourhood of 0 (with respect to λ). This can be shown using uniform convergence of the derivate as in the proof of Lemma 5.4. Unfortunately, g^*_{H,θ,σ_2} does not satisfy Condition 2 in [Hurvich et al., 1998], as the function $|\lambda|^{2H+1}$ is not second order derivable for any H. The bias terms convergence rate of 0 is therefore slower than in [Hurvich et al., 1998] and the choice of m is consequently restricted. It follows that

$$\widehat{d}_n - d = -\frac{1}{2S_m}\sum_{j=1}^{m}(a_j - \overline{a}_m)\log\left(g^*_{H,\alpha,\sigma_2}(\lambda_j)\right) - \frac{1}{2S_m}\sum_{j=1}^{m}(a_j - \overline{a}_m)\,\varepsilon_j, \tag{6.3}$$

where (ε_j) is the error process as defined in Equation (3) of [Hurvich et al., 1998]. Condition 2 in [Hurvich et al., 1998] is required to treat only the behaviour of the first term in the decomposition of $\widehat{d}_n - d$. The others are treated by only making assumptions about the behaviour for m and n. Lemma 5.5 and Theorem 1 in [Hurvich et al., 1998] lead to

$$\widehat{d}_n \xrightarrow[n\to\infty]{\mathbb{P}} d$$

and hence

$$\widehat{H}_n \xrightarrow[n\to\infty]{\mathbb{P}} H.$$

We also require an estimate of convergence rate for $\widehat{H}_n$ (in probability) for the rest of the proof. We therefore apply Theorem 2 from [Hurvich et al., 1998] by treating the last term of Relation (A13) in [Hurvich et al., 1998] as in the Lemma from 5.4) in order to deduce that

$$\widehat{d}_n - d = O_{\mathbb{P}}\left(m^{-\frac{1}{2}}\right)$$

and hence

$$\widehat{H}_n - H = O_{\mathbb{P}}\left(m^{-\frac{1}{2}}\right).$$

Consistency of $\widehat{\alpha}_n$ Letting

$$\widetilde{\Phi}_i^j = \left(\varepsilon_i^H, \dots, \varepsilon_j^H\right)^* \text{ for } i \leqslant j,$$

we have

$$\widehat{\alpha}_n - \alpha = \frac{\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{n-1}}{\Phi_1^{n-1^*}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{n-1}}. \tag{6.4}$$

We now focus on the asymptotic behaviour of the numerator. A Taylor expansion of $\Gamma_{n-1}\left(\widehat{H}_n\right)^{-1}$ at H leads to

$$\begin{aligned}\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{n-1} =&\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}^{-1}(H)\,\Phi_1^{n-1} + \widetilde{\Phi}_2^{n^*}A_n^{(1)}(H)\,\Phi_1^{n-1}\left(\widehat{H}_n - H\right)\\ &+\frac{1}{2}\widetilde{\Phi}_2^{n^*}A_n^{(2)}(H)\,\Phi_1^{n-1}\left(\widehat{H}_n - H\right)^2 + \frac{1}{6}\widetilde{\Phi}_2^{n^*}A_n^{(3)}\left(\overline{H}_n\right)\Phi_1^{n-1}\left(\widehat{H}_n - H\right)^3,\end{aligned}$$

where

$$\begin{aligned}A_n^{(1)}(H) =& -\Gamma_{n-1}^{-1}(H)\,\frac{\partial\Gamma_{n-1}(H)}{\partial H}\Gamma_{n-1}^{-1}(H),\\ A_n^{(2)}(H) =&\Gamma_{n-1}^{-1}(H)\,\frac{\partial^2\Gamma_{n-1}(H)}{\partial^2 H}\Gamma_{n-1}^{-1}(H)\\ &+2\Gamma_{n-1}^{-1}(H)\,\frac{\partial\Gamma_{n-1}(H)}{\partial H}\Gamma_{n-1}^{-1}(H)\,\frac{\partial\Gamma_{n-1}(H)}{\partial H}\Gamma_{n-1}^{-1}(H),\\ A_n^{(3)}(H) =& -6\Gamma_{n-1}^{-1}(H)\,\frac{\partial\Gamma_{n-1}(H)}{\partial H}\Gamma_{n-1}^{-1}(H)\,\frac{\partial\Gamma_{n-1}(H)}{\partial H}\Gamma_{n-1}^{-1}(H)\,\frac{\partial\Gamma_{n-1}(H)}{\partial H}\Gamma_{n-1}^{-1}(H)\\ &-3\Gamma_{n-1}^{-1}(H)\,\frac{\partial^2\Gamma_{n-1}(H)}{\partial^2 H}\Gamma_{n-1}^{-1}(H)\,\frac{\partial\Gamma_{n-1}(H)}{\partial H}\Gamma_{n-1}^{-1}(H)\\ &-3\Gamma_{n-1}^{-1}(H)\,\frac{\partial\Gamma_{n-1}(H)}{\partial H}\Gamma_{n-1}^{-1}(H)\,\frac{\partial^2\Gamma_{n-1}(H)}{\partial^2 H}\Gamma_{n-1}^{-1}(H)\\ &-\Gamma_{n-1}^{-1}(H)\,\frac{\partial^3\Gamma_{n-1}(H)}{\partial^3 H}\Gamma_{n-1}^{-1}(H),\end{aligned}$$

and $\overline{H}_n \in B\left(H, |\widehat{H}_n - H|\right)$. We now use the generic notation

$$\phi_{A_n^{(i)}}\left(X^{(n-1)}, Y^{(n-1)}\right) = X^{(n-1)^*}A_n^{(i)}Y^{(n-1)}$$

for any $\left(X^{(n-1)}, Y^{(n-1)}\right) \in \mathbb{R}^{n-1}\times\mathbb{R}^{n-1}$ and $i \in \{1,2,3\}$. Then,

$$\phi_{A_n^{(i)}}\left(\widetilde{\Phi}_2^n, \Phi_1^{n-1}\right) = \frac{1}{2}\left(\phi_{A_n^{(i)}}\left(\Phi_1^{n-1} + \widetilde{\Phi}_2^n, \Phi_1^{n-1} + \widetilde{\Phi}_2^n\right) - \phi_{A_n^{(i)}}\left(\Phi_1^{n-1}, \Phi_1^{n-1}\right) - \phi_{A_n^{(i)}}\left(\widetilde{\Phi}_2^n, \widetilde{\Phi}_2^n\right)\right) \tag{6.5}$$

We observe that each term of the right-hand side of the above equality is a quadratic form of a Gaussian process whose matrix is expressed as a product of Toeplitz matrices (inverse of Toeplitz matrices) in which the coefficients are Fourier coefficients of a certain function owing to Assertion 4) of Lemma 5.4. We use Lemmas 5.3 and 5.4 with Theorem 2.3 of [Cohen et al., 2013] to obtain the asymptotic behaviour of these terms (suitably renormalized).

***Remark* 6.1.** *Owing to the almost perfect linear representation given by Lemma 5.1, the process* $(X_{n-1}+\varepsilon_n^H)$, (X_n) *has a spectral density which exhibits exactly the same behaviour as that of the fractional Gaussian noise in the neighbourhood of* 0. *Indeed, the spectral density of* $(X_{n-1}+\varepsilon_n^H)$, (X_n) *satisfies Conditions (1) and (2) of Lemma 5.4 (although only* $j=0$ *is necessary) and satisfies the estimates given in Condition (3c) of Lemma 5.4.*

***Remark* 6.2.** *With the notations of Theorem 2.3 in* [Cohen et al., 2013], *it is evident that, in our case, the third assumption of this Theorem is satisfied for all* $p\in\mathbb{N}^*$ *owing to Estimation (3) of Lemma 5.4 and Remark 6.1.*

To simplify the proof, we detail the treatment of one term, with the others following the same reasoning. We let

$$\frac{1}{n}\phi_{A_n^{(2)}}\left(\widetilde{\Phi}_2^n,\Phi_1^{n-1}\right)=\frac{1}{2n}\left(\phi_{A_n^{(2)}}\left(\Phi_1^{n-1}+\widetilde{\Phi}_2^n,\Phi_1^{n-1}+\widetilde{\Phi}_2^n\right)\right)-\phi_{A_n^{(2)}}\left(\Phi_1^{n-1},\Phi_1^{n-1}\right)-\phi_{A_n^{(2)}}\left(\widetilde{\Phi}_2^n,\widetilde{\Phi}_2^n\right)\Big)\Big)$$

and let $\Gamma_n\left(H,\theta,\sigma_2\right)$ (respectively $\widetilde{\Gamma}_n\left(H,\theta,\sigma_2\right)$) be the covariance matrix of the Gaussian process $(X_j)_{\{1\leqslant j\leqslant n\}}$ (respectively $(X_j+\varepsilon_{j+1}^H)_{\{1\leqslant j\leqslant n\}}$). From Lemma 5.3, we have

$$\frac{1}{n}\mathbb{E}\left(\phi_{A_n^{(2)}}\left(\Phi_1^{n-1}+\widetilde{\Phi}_2^n,\Phi_1^{n-1}+\widetilde{\Phi}_2^n\right)\right)=\frac{1}{n}Tr\left(A_n^{(2)}\left(H\right)\widetilde{\Gamma}_{n-1}\left(H,\alpha,\sigma_2\right)\right)\tag{6.6}$$

$$\frac{1}{n}\mathbb{E}\left(\phi_{A_n^{(2)}}\left(\Phi_1^{n-1},\Phi_1^{n-1}\right)\right)=\frac{1}{n}Tr\left(A_n^{(2)}\left(H\right)\Gamma_n\left(H,\alpha,\sigma_2\right)\right)\tag{6.7}$$

$$\frac{1}{n}\mathbb{E}\left(\phi_{A_n^{(2)}}\left(\widetilde{\Phi}_2^n,\widetilde{\Phi}_2^n\right)\right)=\frac{1}{n}Tr\left(A_n^{(2)}\left(H\right)\sigma_2\Gamma_{n-1}\left(H\right)\right)\tag{6.8}$$

$$\frac{1}{n^2}\mathbb{V}ar\left(\phi_{A_n^{(2)}}\left(\Phi_1^{n-1}+\widetilde{\Phi}_2^n,\Phi_1^{m-1}+\widetilde{\Phi}_2^n\right)\right)=\frac{1}{n^2}Tr\left(\left(A_n^{(2)}\left(H\right)\widetilde{\Gamma}_{n-1}\left(H,\alpha,\sigma_2\right)\right)^2\right)\tag{6.9}$$

$$\frac{1}{n^2}\mathbb{V}ar\left(\phi_{A_n^{(2)}}\left(\Phi_1^{n-1},\Phi_1^{n-1}\right)\right)=\frac{1}{n^2}Tr\left(\left(A_n^{(2)}\left(H\right)\Gamma_n\left(H,\alpha,\sigma_2\right)\right)^2\right)\tag{6.10}$$

and

$$\frac{1}{n^2}\mathbb{V}ar\left(\phi_{A_n^{(2)}}\left(\widetilde{\Phi}_2^n,\widetilde{\Phi}_2^n\right)\right)=\frac{1}{n^2}Tr\left(\left(A_n^{(2)}\left(H\right)\sigma_2\Gamma_{n-1}\left(H\right)\right)^2\right)\tag{6.11}$$

In view of Remarks 6.1, 6.2 and Lemma 5.4, we can apply Theorem 2.3 in [Cohen et al., 2013] and deduce that

$$\frac{1}{n}Tr\left(A_n^{(2)}\left(H\right)\widetilde{\Gamma}_{n-1}\left(H,\alpha,\sigma_2\right)\right)\xrightarrow[n\to\infty]{}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\frac{\partial^2}{\partial^2H}f_H(\lambda)\widetilde{f}_{H,\theta,\sigma_2}(\lambda)}{f_H(\lambda)^2}+2\frac{\frac{\partial}{\partial H}f_H(\lambda)^2\widetilde{f}_{H,\alpha,\sigma_2}(\lambda)}{f_H(\lambda)^3}\right)d\lambda$$

and that

$$\frac{1}{n}Tr\left(\left(A_n^{(2)}\left(H\right)\widetilde{\Gamma}_{n-1}\left(H,\alpha,\sigma_2\right)\right)^2\right)\xrightarrow[m\to\infty]{}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\frac{\partial^2}{\partial^2H}f_H(\lambda)\widetilde{f}_{H,\alpha,\sigma_2}(\lambda)}{f_H(\lambda)^2}+2\frac{\frac{\partial}{\partial H}f_H(\lambda)^2\widetilde{f}_{H,\theta,\sigma_2}(\lambda)}{f_H(\lambda)^3}\right)^2d\lambda,$$

where $\widetilde{f}_{H,\alpha,\sigma_2}$ is the spectral density of the process $(X_{n-1}+\varepsilon_n^H)$. Consequently, (6.9) converge to 0 and

$$\frac{1}{n}\phi_{A_n^{(2)}}\left(\Phi_1^{n-1}+\widetilde{\Phi}_2^m,\Phi_1^{m-1}+\widetilde{\Phi}_2^n\right)\xrightarrow[m\to\infty]{\mathbb{P}}\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\frac{\partial^2}{\partial^2 H}f_H(\lambda)\widetilde{f}_{H,\alpha,\sigma_2}(\lambda)}{f_H(\lambda)^2}+2\frac{\frac{\partial}{\partial H}f_H(\lambda)^2\widetilde{f}_{H,\alpha,\sigma_2}(\lambda)}{f_H(\lambda)^3}\right)d\lambda.$$

The generalization of this reasoning leads to

$$\frac{\widetilde{\Phi}_2^{m^*}A_n^{(1)}(H)\,\Phi_1^{n-1}}{n}\xrightarrow[n\to\infty]{\mathbb{P}}K^{(1)}_{H,\alpha,\sigma_2}\tag{6.12}$$

and

$$\frac{\widetilde{\Phi}_2^{n^*}A_n^{(2)}(H)\,\Phi_1^{n-1}}{n}\xrightarrow[n\to\infty]{\mathbb{P}}K^{(2)}_{H,\alpha,\sigma_2}\tag{6.13}$$

where $K^{(1)}_{H,\theta,\sigma_2}$ and $K^{(2)}_{H,\theta,\sigma_2}$ are constants which can be expressed in integral form from Theorem 2.3 in [Cohen et al., 2013]. We now turn to the most delicate term, which is

$$\widetilde{\Phi}_2^{n^*}A_n^{(3)}\left(\overline{H}_n\right)\Phi_1^{n-1}=\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}(H)^{-\frac{1}{2}}\Gamma_{n-1}(H)^{\frac{1}{2}}A_n^{(3)}\left(\overline{H}_n\right)\Gamma_{n-1}(H)^{\frac{1}{2}}\Gamma_{n-1}(H)^{-\frac{1}{2}}\Phi_1^{n-1}.$$

The Cauchy-Schwarz inequality implies that

$$\left|\widetilde{\Phi}_2^{n^*}A_n^{(3)}\left(\overline{H}_n\right)\Phi_1^{m-1}\right|\leqslant\left|\Gamma_{n-1}^{-\frac{1}{2}}(H)\widetilde{\Phi}_2^{n^*}\right|\left|\Gamma_{n-1}(H)^{\frac{1}{2}}A_n^{(3)}\left(\overline{H}_n\right)\Gamma_{n-1}^{\frac{1}{2}}(H)\right|\left|\Gamma_{n-1}^{-\frac{1}{2}}(H)\Phi_1^{n-1}\right|.$$

The quantities $\left|\Gamma_{n-1}(H)^{-\frac{1}{2}}\widetilde{\Phi}_2^{n^*}\right|^2$ and $\left|\Gamma_{n-1}(H)^{-\frac{1}{2}}\Phi_1^{n-1}\right|^2$ are quadratic forms of Gaussian process. Using Theorem 2.3 in [Cohen et al., 2013], we produce

$$\frac{1}{\sqrt{n}}\left|\Gamma_{n-1}(H)^{-\frac{1}{2}}\widetilde{\Phi}_2^{n^*}\right|=O_{\mathbb{P}}(1)\ \text{ and }\ \frac{1}{\sqrt{n}}\left|\Gamma_{n-1}(H)^{-\frac{1}{2}}\Phi_1^{n-1}\right|=O_{\mathbb{P}}(1).$$

Now,

$$\begin{aligned}\left|\Gamma_{n-1}(H)^{\frac{1}{2}}A_n^{(3)}\left(\overline{H}_n\right)\Gamma_{m-1}(H)^{\frac{1}{2}}\right|\leqslant&Tr\left(\left(\Gamma_{n-1}(H)^{\frac{1}{2}}A_n^{(3)}\left(\overline{H}_n\right)\Gamma_{n-1}(H)^{\frac{1}{2}}\right)^2\right)\\=&\ Tr\left(A_m^{(3)}\left(\overline{H}_n\right)\Gamma_{n-1}(H)\,A_m^{(3)}\left(\overline{H}_n\right)\Gamma_{n-1}(H)\right).\end{aligned}$$

We let $\delta>0$ and $A_\delta=\{h\in\mathbb{R}\big|h\in[H-\delta;H+\delta]\}$. Fixing $K>0$, then

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}Tr\left(A_n^{(3)}\left(\overline{H}_n\right)\Gamma_{m-1}(H)\,A_n^{(3)}\left(\overline{H}_n\right)\Gamma_{n-1}(H)\right)\leqslant K\right)\geqslant&\mathbb{1}_{\left\{\frac{1}{n}Tr\left(A_n^{(3)}\left(\overline{H}_n\right)\Gamma_{n-1}(H)A_n^{(3)}\left(\overline{H}_n\right)\Gamma_{n-1}(H)\right)\leqslant K\right\}}\\&\times\mathbb{P}\left(\overline{H}_n\in A_\delta\right)\end{aligned}$$

Note that A_δ is a compact set, so we can choose a δ small enough to verify the third hypothesis of Theorem 2.3 in [Cohen et al., 2013] which works for uniform convergence. Owing to this Theorem, we have

$$\lim_{n\to\infty}\sup_{\overline{H}_n\in A_\delta}\frac{1}{n}Tr\left(A_n^{(3)}\left(\overline{H}_n\right)\Gamma_{n-1}(H)\,A_n^{(3)}\left(\overline{H}_n\right)\Gamma_{n-1}(H)\right)\leqslant K_{\Theta^*},$$

where K_{Θ^*} is a constant depending on the set A_δ. This constant is finite because it expresses itself as a supremum of an integral (i.e. it is distributed over a compact set) with respect to the λ of continuous function with respect to H for almost any λ bounded by an integrable function not depending on H. We deduce that

$$\lim_{n\to\infty}\mathbb{P}\left(\frac{1}{n}Tr\left(A_n^{(3)}\left(\overline{H}_n\right)\Gamma_{n-1}(H)\,A_n^{(3)}\left(\overline{H}_n\right)\Gamma_{n-1}(H)\right)\leqslant K_{\Theta^*}\right)=1$$

and hence

$$n^{-\frac{3}{2}}\widetilde{\Phi}_2^{n^*}A_n^{(3)}\left(\overline{H}_n\right)\Phi_1^{n-1}=O_{\mathbb{P}}(1).\tag{6.14}$$

It is shown in [Esstafa, 2019] (see Chapter 4, Lemma 4.4) that

$$\frac{1}{n}\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}^{-1}(H)\,\Phi_1^{n-1}\xrightarrow[n\to\infty]{\mathbb{P}}0.\tag{6.15}$$

The combination of (6.12),(6.13), (6.14) and (6.15) allows us to deduce that

$$\frac{1}{n}\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{n-1}\xrightarrow[n\to\infty]{\mathbb{P}}0. \tag{6.16}$$

We now consider the asymptotic behaviour of the denominator. We know that

$$\begin{aligned}\Phi_1^{n-1}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{n-1} =&\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}^{-1}\left(H\right)\Phi_1^{n-1}+\Phi_1^{n-1}A_n^{(1)}\left(H\right)\Phi_1^{n-1}\left(\widehat{H}_n-H\right)\\&+\frac{1}{2}\Phi_1^{n-1}A_n^{(2)}\left(\overline{H}_n\right)\Phi_1^{n-1}\left(\widehat{H}_n-H\right)^2.\end{aligned}$$

The same reasoning which made it possible to obtain (6.12) and (6.13) also makes it possible to show that

$$\frac{1}{n}\Phi_1^{n-1}A_n^{(1)}\left(H\right)\Phi_1^{n-1}\xrightarrow[n\to\infty]{\mathbb{P}}K_{H,\alpha,\sigma_2}^{(3)}, \tag{6.17}$$

where $K_{H,\alpha,\sigma_2}^{(3)}$ is a positive constant. In the same manner as for the treatment of (6.14), we have

$$\frac{1}{n^{\frac{3}{2}}}\Phi_1^{n-1}A_n^{(2)}\left(\overline{H}_n\right)\Phi_1^{n-1}=O_{\mathbb{P}}\left(1\right). \tag{6.18}$$

It is shown in [Esstafa, 2019] (see Chapter 4, Lemma 4.3) that

$$\frac{1}{n}\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}^{-1}\left(H\right)\Phi_1^{n-1}\xrightarrow[n\to\infty]{\mathbb{P}}\frac{\sigma_2}{1-\alpha^2}. \tag{6.19}$$

Equations (6.17), (6.18) and (6.19) lead us to

$$\frac{1}{n}\Phi_1^{n-1}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{n-1}\xrightarrow[n\to\infty]{\mathbb{P}}\frac{\sigma_2}{1-\alpha^2}. \tag{6.20}$$

Finally, Equations (6.16) and (6.20) give us the consistency of $\widehat{\alpha}_n$.

Consistency of $\widehat{\sigma}_{2,n}$. We know that

$$\begin{aligned}\widehat{\sigma}_{2,n}-\sigma_2 =&\frac{1}{n-1}\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\widetilde{\Phi}_2^{n}-\sigma_2-\left(\widehat{\alpha}_n-\alpha\right)\frac{2}{n-1}\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{n-1}\\&+\left(\widehat{\alpha}_n-\alpha\right)^2\frac{1}{n-1}\Phi_1^{n-1^*}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{n-1}.\end{aligned}$$

Using similar arguments as the proof of the consistency of $\widehat{\alpha}_n$ (i.e. by expanding $\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)$ at H to the third order), we produce

$$\frac{1}{n-1}\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\widetilde{\Phi}_2^{n}\xrightarrow[n\to\infty]{\mathbb{P}}\sigma_2, \tag{6.21}$$

$$\frac{1}{n-1}\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{n-1}\xrightarrow[n\to\infty]{\mathbb{P}}0, \tag{6.22}$$

and

$$\frac{1}{n-1}\Phi_1^{n-1^*}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{n-1}\xrightarrow[n\to\infty]{\mathbb{P}}\frac{\sigma_2}{1-\alpha^2}. \tag{6.23}$$

Combining (6.21), (6.22) and (6.23) with the consistency of $\widehat{\alpha}_n$, we find that $\widehat{\sigma}_{2,n}-\sigma_2$ converges in probability to 0.

Asymptotic joint distribution of the estimators. We show that the statistical error of each component is a function of the statistical error of H modulo a remainder term which is asymptotically negligible. The first estimator is satisfied (owing to Theorem 2 in [Hurvich et al., 1998]) as

$$\sqrt{m}\left(\widehat{H}_n-H\right)\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}\left(0;\frac{\pi^2}{24}\right) \tag{6.24}$$

The second estimator is shown in [Esstafa, 2019] (see Chapter 4, Lemma 4.5) as

$$\frac{1}{\sqrt{n}}\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}^{-1}\left(H\right)\Phi_1^{n-1}\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}\left(0,\frac{\sigma_2}{1-\alpha^2}\right). \tag{6.25}$$

The combination of (6.25) and the decomposition of $\widetilde{\Phi}_2^{m^*}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{m-1}$ lead to

$$\sqrt{m}\left(\widehat{\alpha}_n-\alpha\right)=\sqrt{m}\frac{\widetilde{\Phi}_2^{n^*}A_n^{(1)}\left(H\right)\Phi_1^{n-1}}{\Phi_1^{n-1}\Gamma_{n-1}^{-1}\left(\widehat{H}_n\right)\Phi_1^{n-1}}\left(\widehat{H}_n-H\right)+R_n^{(1)} \tag{6.26}$$

where

$$R_n^{(1)}\xrightarrow[n\to\infty]{\mathbb{P}}0.$$

To treat the third estimator, we show via a cumulant method (like in Theorem 2 of [Avram, 1988]) that

$$\frac{\widetilde{\Phi}_2^{n^*}\Gamma_{n-1}^{-1}\left(H\right)\widetilde{\Phi}_2^{n}}{\sqrt{n}}-\sigma_2\sqrt{n}\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}\left(0;V_\vartheta\right) \tag{6.27}$$

for $V_\vartheta>0$, which can be expressed in integral form as in Theorem 2.3 of [Cohen et al., 2013]. It only remains to apply (6.27) and (6.22) together with the expansion of $\widehat{\sigma}_{2,n}-\sigma_2$ to deduce that

$$\sqrt{m}\left(\widehat{\sigma}_{2,n}-\sigma_2\right)=\sqrt{m}\frac{\widetilde{\Phi}_2^{n^*}A_n^{(1)}\left(H\right)\widetilde{\Phi}_2^{n}}{n}\left(\widehat{H}_n-H\right)+R_n^{(2)} \tag{6.28}$$

where

$$R_n^{(2)}\xrightarrow[n\to\infty]{\mathbb{P}}0$$

and

$$\frac{\widetilde{\Phi}_2^{n^*}A_n^{(1)}\left(H\right)\widetilde{\Phi}_2^{n}}{n}\xrightarrow[n\to\infty]{\mathbb{P}}C_\vartheta^{(1)}$$

for some finite constant $C_\vartheta^{(1)}>0$. Then

$$\widetilde{\Sigma}_\vartheta=U_\vartheta U_\vartheta^* \tag{6.29}$$

with

$$U_\vartheta=\begin{pmatrix}1\\ \frac{\sigma_2K_{H,\alpha,\sigma_2}^{(1)}}{1-\alpha^2}\\ C_\vartheta^{(1)}\end{pmatrix}$$

and (2.11) is proved.

6.3. **Proof of proposition 2.2.** This result is a direct consequence of Theorem 3.4 from [Cohen et al., 2013].

6.4. **Proof of proposition 2.3.** Direct computations lead to

$$\sqrt{n}\left(\widetilde{\vartheta}_n^{(1)}-\vartheta\right)=\sqrt{n}\left(\widehat{\vartheta}_n^{(1)}-\vartheta\right)+\mathcal{I}^{-1}(\widehat{\vartheta}_n)\frac{\nabla\ell_n(\widehat{\vartheta}_n^{(1)})}{\sqrt{n}}. \tag{6.30}$$

The mean value theorem yields

$$\nabla\ell_n(\widehat{\vartheta}_n^{(1)})=\nabla\ell_n(\vartheta)+\left(\widehat{\vartheta}_n^{(1)}-\vartheta\right)\int_0^1\Delta\ell_n(\vartheta+v(\widehat{\vartheta}_n^{(1)}-\vartheta))dv$$

and, substituting in (6.30), we produce

$$\sqrt{n}\left(\widetilde{\vartheta}_n^{(1)}-\vartheta\right)=\sqrt{n}\left(\widehat{\vartheta}_n^{(1)}-\vartheta\right)\mathcal{I}^{-1}(\widehat{\vartheta}_n^{(1)})\left(\mathcal{I}(\widehat{\vartheta}_n^{(1)})+\frac{\int_0^1\Delta\ell_n(\vartheta_v)dv}{n}\right)+\mathcal{I}^{-1}(\widehat{\vartheta}_n^{(1)})\frac{\nabla\ell_n(\vartheta)}{\sqrt{n}},$$

where $\vartheta_v=\vartheta+v(\widehat{\vartheta}_n^{(1)}-\vartheta)$ for some $0<v<1$.
We then consider the first term, as follows.

$$\begin{aligned}A_n &= \sqrt{n}\left(\widehat{\vartheta}_n^{(1)}-\vartheta\right)\mathcal{I}(\widehat{\vartheta}_n^{(1)})^{-1}\left(\mathcal{I}(\widehat{\vartheta}_n^{(1)})+\frac{1}{\sqrt{n}}\int_0^1\frac{\Delta\ell_n(\vartheta_v)}{\sqrt{n}}dv\right)\\ &= \sqrt{n^\delta}\left(\widehat{\vartheta}_n^{(1)}-\vartheta\right)\mathcal{I}(\widehat{\vartheta}_n^{(1)})^{-1}\sqrt{n^{1-\delta}}\left(\mathcal{I}(\widehat{\vartheta}_n^{(1)})+\frac{1}{\sqrt{n}}\int_0^1\frac{\Delta\ell_n(\vartheta_v)}{\sqrt{n}}dv\right)\end{aligned}$$

Then,

$$\begin{aligned}\mathcal{I}(\widehat{\vartheta}_n)^{(1)} + \frac{1}{\sqrt{n}}\int_0^1 \frac{\Delta \ell_n(\vartheta_v)}{\sqrt{n}} dv &= \left(\mathcal{I}(\widehat{\vartheta}_n^{(1)}) - \mathcal{I}(\vartheta)\right) + \left(\mathcal{I}(\vartheta) + \frac{\Delta \ell_n(\vartheta)}{n}\right) \\ &\quad + \frac{1}{\sqrt{n}}\int_0^1 \left(\frac{\Delta \ell_n(\vartheta_v)}{\sqrt{n}} - \frac{\Delta \ell_n(\vartheta)}{\sqrt{n}}\right) dv\end{aligned}$$

The first subterm needs more regularity than the usual uniform continuity of the Sweeting conditions [Sweeting, 1980]. As in [Kutoyants and Motrunich, 2016], we impose that the Fisher information matrix is locally Lipschitz, as shown in Lemma 5.6. Consequently,

$$\left\|\mathcal{I}(\widehat{\vartheta}_n^{(1)}) - \mathcal{I}(\vartheta)\right\| \leq K \left\|\widehat{\vartheta}_n^{(1)} - \vartheta\right\|$$

and this subterm is $O_{\mathbb{P}}\left(n^{-\frac{\delta}{2}}\right)$.
Lemma 5.7 indicates that the renormalized second term,

$$\sqrt{n}\left(\mathcal{I}(\vartheta) + \frac{\Delta \ell_n(\vartheta)}{n}\right),$$

is $O_{\mathbb{P}}(1)$, or that the second term is $O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$.
Lemma 5.8 produces

$$\frac{1}{\sqrt{n}}\int_0^1 \left(\frac{\Delta \ell_n(\vartheta_v)}{\sqrt{n}} - \frac{\Delta \ell_n(\vartheta)}{\sqrt{n}}\right) dv = O_{\mathbb{P}}\left(n^{\kappa - \frac{\delta}{2}}\right)$$

for any $\kappa > 0$. Then, combining with the previous result,

$$A_n = n^{\frac{1}{2} - \frac{\delta}{2}}\left(O_{\mathbb{P}}\left(n^{-\frac{\delta}{2}}\right) + O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right) + O_{\mathbb{P}}\left(n^{\kappa - \frac{\delta}{2}}\right)\right).$$

As $\frac{\sqrt{n}}{n^\delta} \longrightarrow 0$, we get that $A_n \to 0$ in probability.
The second term is

$$\begin{aligned}B_n &= \mathcal{I}(\widehat{\vartheta}_n^{(1)})^{-1}\frac{\nabla \ell_n(\vartheta)}{\sqrt{n}} \\ &= \mathcal{I}(\vartheta)^{-1}\frac{\nabla \ell_n(\vartheta)}{\sqrt{n}} + \left(\mathcal{I}(\widehat{\vartheta}_n^{(1)})^{-1} - \mathcal{I}(\vartheta)^{-1}\right)\frac{\nabla \ell_n(\vartheta)}{\sqrt{n}}.\end{aligned}$$

We show that the second subterm tends to 0 in probability as $\mathcal{I}_n(\cdot)$ is supposed to be non-degenerate and uniformly continuous. The central limit theorem for the first subterm gives the result.

References

[Avram, 1988] Avram, F. (1988). On bilinear forms in Gaussian random variables, Toeplitz matrices and Parseval's relation. *Probability Theory and Related Fields*, 79:37–45.

[Brockwell and Davis, 1981] Brockwell, J. and Davis, R. A. (1981). Time series: Theory and methods.

[Brouste et al., 2014] Brouste, A., Cai, C., and Kleptsyna, M. (2014). Asymptotic properties of the MLE for the autoregressive process coefficients under stationary Gaussian noise. *Mathematical Methods of Statistics*, 23(2):103–115.

[Brouste et al., 2020a] Brouste, A., Cai, C., Soltane, M., and Wang, L. (2020a). Testing for the change of the mean-reverting parameter of an autoregressive model with stationary gaussian noise. *Statistical Inference for Stochastic Processes*, 28:301–308.

[Brouste et al.,] Brouste, A., Dutang, C., and Ntousa Meniedou, D. OneStep - Le Cam's one-step estimation procedure. *To appear in R Journal*.

[Brouste and Iacus, 2013] Brouste, A. and Iacus, S. (2013). Parameter estimation for the discretely observed fractional Ornstein Uhlenbeck process and the Yuima R package. *Computational Statistics*, 28(4):1529–1547.

[Brouste and Masuda, 2018] Brouste, A. and Masuda, H. (2018). Efficient estimation of stable lévy process with symmetric jumps. *Statistical Inference for Stochastic Processes*, 21:289–307.

[Brouste et al., 2020b] Brouste, A., Soltane, M., and Votsi, E. (2020b). One-step estimation for the fractional Gaussian noise model at high-frequency. *ESAIM Probability and Statistics*, 24:827–841.

[Cohen et al., 2013] Cohen, S., Gamboa, F., Lacaux, C., and Loubes, J.-M. (2013). LAN property for some fractional type Brownian motion. *ALEA : Latin American Journal of Probability and Mathematical Statistics*, 10(1):91–106.

[Dahlhaus, 1989] Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 17(4):1749–1766.

[Dahlhaus, 2006] Dahlhaus, R. (2006). Correction efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 34(2):1045–1047.

[Esstafa, 2019] Esstafa, Y. (2019). *Long-memory time series models with dependent innovations*. Theses, Université Bourgogne Franche-Comté.

[Fox and Taqqu, 1986] Fox, R. and Taqqu, M. (1986). Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, 14(2):517–532.

[Fukasawa and Tetsuya, 2019] Fukasawa, M. and Tetsuya, T. (2019). Asymptotically efficient estimators for self-similar stationary Gaussian noises under high frequency observations. *Bernoulli*, 25(3):1870 – 1900.

[Gatheral et al., 2018] Gatheral, J., Jaisson, T., and Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, 18(6):933–949.

[Geweke and Porter-Hudak, 1983] Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4):221–238.

[Gloter and Yoshida, 2021] Gloter, A. and Yoshida, N. (2021). Adaptive estimation for degenerate diffusion processes. *Electronic Journal of Statistics*, 15(1):1424 – 1472.

[Hurvich and Beltrao, 1994] Hurvich, C. M. and Beltrao, K. I. (1994). Automatic semiparametric estimation of the memomy parameter of a long memory time series. *Journal of Time Series Analysis*, 15(3):285–302.

[Hurvich et al., 1998] Hurvich, C. M., Deo, R., and Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis*, 19(1):19–46.

[Ibragimov and Has'minskii, 1981] Ibragimov, I. A. and Has'minskii, R. Z. (1981). Statistical estimation: Asymptotic theory. *Springer-Verlag*, 16.

[Istas and Lang, 1997] Istas, J. and Lang, G. (1997). Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Annales de l'I.H.P. section B*, 33(4):407–436.

[Kamatani and Masayuki, 2015] Kamatani, K. A. and Masayuki, Y. (2015). Hybrid multi-step estimators for stochastic differential equations based on sampled data. *Statistical Inference for Stochastic Processes*, 18:177–204.

[Kutoyants and Motrunich, 2016] Kutoyants, Y. A. and Motrunich, A. (2016). On multi-step mle-process for markov sequences. *Metrika*, 79:705–724.

[Le Cam, 1956] Le Cam, L. (1956). On the asymptotic theory of estimation and testing hypothesis. *Proceedings of the 3rd Berkeley Symposium*, 1:355–368.

[Lieberman et al., 2012] Lieberman, O., Rosemarin, R., and Rousseau, J. (2012). Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary gaussian processes. *Econometric Theory*, 28(2):457–470.

[Robinson, 1995a] Robinson, P. (1995a). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23(5):1630–1661.

[Robinson, 1995b] Robinson, P. (1995b). Log-periodogram regression of time series with long-range dependence. *The Annals of Statistics*, 23(3):1048–1072.

[Shimotsu and Phillips, 2005] Shimotsu, K. and Phillips, P. C. B. (2005). Exact local whittle estimation of fractional integration. *The Annals of Statistics*, 33(4):1890–1933.

[Soltane, 2018] Soltane, M. (2018). Asymptotic efficiency in autoregressive processes driven by stationary gaussian noise. *In Revision*.

[Stout, 1974] Stout, W. F. (1974). *Almost sure convergence*, volume 24 of *Probability and Mathematical Statistics*. Academic Press, New-York-London.

[Sweeting, 1980] Sweeting, T. (1980). Uniform asymptotic normality of the maximum likelihood estimator. *The Annals of Statistics*, 8(6):1375–1381.

Laboratoire Manceau de Mathématiques, Le Mans Université, Avenue O. Messiaen, 72085 Le Mans Cedex 9, France.

Email address: samir.ben_hariz@univ-lemans.fr

Laboratoire Manceau de Mathématiques, Le Mans Université, Avenue O. Messiaen, 72085 Le Mans Cedex 9, France.

Email address: Alexandre.Brouste@univ-lemans.fr

School of Mathematics, Shanghai University of Finance and Economics, No.777 Guoding Road, Yangpu District, Shanghai, China.

Email address: caichunhao@mail.shufe.edu.cn

Laboratoire Manceau de Mathématiques, Le Mans Université, Avenue O. Messiaen, 72085 Le Mans Cedex 9, France.

Email address: marius.soltane@univ-lemans.fr