Parameter estimation of Gaussian stationary processes using the generalized method of moments

Luis A. Barboza

Centro de Investigación en Matemática Pura y Aplicada (CIMPA)
Universidad de Costa Rica
San José, Costa Rica
e-mail: luisalberto.barboza@ucr.ac.cr

and

Frederi G. Viens

Department of Statistics and Probability, Michigan State University
East Lansing, MI, USA
e-mail: viens@stt.msu.edu

Abstract: We consider the class of all stationary Gaussian process with explicit parametric spectral density. Under some conditions on the autocovariance function, we defined a GMM estimator that satisfies consistency and asymptotic normality, using the Breuer-Major theorem and previous results on ergodicity. This result is applied to the joint estimation of the three parameters of a stationary Ornstein-Uhlenbeck (fOU) process driven by a fractional Brownian motion. The asymptotic normality of its GMM estimator applies for any H in $(0, 1)$ and under some restrictions on the remaining parameters. A numerical study is performed in the fOU case, to illustrate the estimator's practical performance when the number of data-points is moderate.

Contents

1. Introduction

In this article we are interested in the parameter estimation of a stationary Gaussian process with explicit spectral density, which applies in particular to the stationary solution of the Ornstein-Uhlenbeck equation driven by the fractional Brownian motion, also known as the fOU process. The main assumption on the data is that a single trajectory of the process is observed at predetermined discrete times, with a fixed length of observation time, and an increasing horizon. Thus this study, which applies to continuous-time Gaussian stochastic processes, also falls within the realm of Gaussian time series with complex parametric auto-covariances.

It is well known that the Maximum Likelihood estimation (MLE) is the preferred method of estimation because of its asymptotic optimality among densities with regular properties and independent samples (see for example, [42]). If the samples are generated from a single path of a stationary process, they can be far from independent, and the asymptotic properties of the MLE can be deduced, but this would depend on their distributional assumptions. For example, one of the most studied classes of stochastic processes is the Gaussian one, mainly because of its wide use in modeling events which have some degree of temporal dependence, particularly when joint-distributional information need only rely on the first two moments, since no further information is needed in the Gaussian case. In a particular instance, the MLE estimator of a stationary ARMA process is strongly consistent and asymptotically efficient (see section 10.8 in [7]). As an interesting generalization of the above fact, [41] studies the parameter estimation of a Gaussian process with rational spectral density, using a modification of the likelihood function.

Another interesting example in which the MLE uses spectral information is the estimation of self-similar processes. The best known example of such processes in a Gaussian setting is the fractional Brownian motion (fBm). Its applications have been widely studied in many areas including hydrology, finance, climatology, and others. Jan Beran explains in more detail its definitions and properties in [3] as well as another example of interest in applications: the fARIMA process. The estimation of the fractal dimension H of the fBm process, also known as Hurst parameter, which turns out to be the same parameter determining the slow power speed of decay of the auto-correlation function for fBm's stationary increments, namely n^{2H-2}, has been researched from different points of view: exact and approximate Maximum Likelihood using spectral in-

formation ([19], [15] and [32]), discrete variations of the sample paths ([27], [14], [43], [45] and [12]), and wavelets ([2], [18], [35]).

A natural generalization of the above example is to consider processes whose variogram is (see [27]):

$$v(t) = C|t|^s + o(|t|^s).$$

If $s = 2H$, a sufficient condition for the above to be true is that the spectral density can be written as a power law with degree $-1 - 2H$ plus a remainder (see equation (5) in [4]). It is possible to estimate the parameter H of such processes using the quadratic variation of a filtered process, which will be a fundamental idea in our suggested method, though the interpretation of H as a memory parameter for discrete observation, as in our study (where what matters is the small-t behavior of the variogram), is not the same as its interpretation for use with in-fill asymptotics, where how H determines self-similarity or path regularity is what relates it to estimation schemes. The fractional Onrstein-Uhlenbeck (fOU) process satisfies the above-mentioned condition from [27]; see [10] for an exhaustive definition and initial study. We also give a brief review of its definition in section 3.1.

Kleptsyna and Le Breton [29] proved that a maximum likelihood estimator of the drift parameter λ for the fOU process (in the non-stationary case) can be obtained using similar techniques to the case where $H = \frac{1}{2}$. This is attained by considering $H \in [\frac{1}{2}, 1)$. The main difference between [29] and [34] is the presence of the fractional kernels within the likelihood-estimate formulas, due to the generalization of the Girsanov formula to the fBm case. They use this formula to obtain an MLE of λ which is strongly consistent. Furthermore, they compute approximations of the bias and MSE of their estimator. Tudor and this paper's second author [44] extended the previous result to fractional SDEs where the drift parameter is not necessarily linear. They proved strong consistency of the MLE estimator even in the case $H < \frac{1}{2}$, using Malliavin calculus techniques, together with bias and MSE approximations for their estimate of λ. They also provide a discretized estimator which is still strongly consistent under certain conditions on the drift. A more recent estimate of λ (assuming $\lambda > 0$) was developed by Hu and Nualart [25]. They propose a least-squares estimate which is written in terms of an Itô stochastic integral, but its numerical manipulation is difficult when $H > \frac{1}{2}$; the estimator cannot be implemented directly when $H < \frac{1}{2}$ since then, the stochastic integral would have to be interpreted in the Skorohod sense, and the latter cannot be evaluated measurably with respect to the fOU process's sigma-field, as it would be if the integral were a classical Wiener or Wiener/Ito integral; in other words, because the Skorohod integral is not a stochastic integral in the classical sense, but rather is a divergence operator (it is the adjoint of a derivative operator, and cannot be interpreted as an antiderivative in the case of fBm-based models), information from the entire path of the noise term driving of the fOU process would be required, which is not an explicit functional of the observations of the fOU process itself. The estimate is strongly consistent for all $H \geq \frac{1}{2}$ and it turns out that for $H \in [\frac{1}{2}, \frac{3}{4})$ the least-squares estimate satisfies a classical CLT with a variance

depending on the unknown parameter λ. Moreover, they improved the numerical applicability of the previous estimate by introducing a new one in terms of a Riemann integral of the observed process X_t ($H \geq \frac{1}{2}$), and this estimate turns out to be strongly consistent and satisfies a CLT for $H \in \left[\frac{1}{2}, \frac{3}{4}\right)$. The proofs of strong consistency and the CLT for both estimates rely on Malliavin calculus techniques, in particular, the Nualart-Pecatti-Ortiz characterization of normal convergence for multiple stochastic integrals (see [39]). In a more recent study, Brouste and Iacus (see [8]) proved that the variogram of a non-stationary fOU process satisfies the regularity assumptions of Istas and Lang (see [27]) and hence they were able to give a joint estimator of (H, σ^2) when λ is fixed. In fact, they proved that it is strongly consistent and asymptotically normal. When $H \in (\frac{1}{2}, \frac{3}{4})$ they also gave an estimator of λ based on Hu and Nualart's results in [25], under a certain combination of increasing-domain and infill asymptotics. For H and σ fixed, their estimate is strongly consistent and asymptotically normal as well. Recently [17] generalized the previous results to the estimation of drift and scaling parameters of essentially any long-memory stationary Gaussian process. That paper was written in response to their own paper jointly with El Onsy [16] and several other recent papers cited in [17] which are other special cases of their framework. However, their method does not appear to handle any memory parameter, nor do any of the methods cited in their work.

As we mentioned from the outset, our interest is to estimate the parameters in stochastic differential equations using discrete observations, which can be particularly useful in the areas of econometrics and finance; indeed these fields' long-memory stochastic modeling are typically limited in how frequently data can be observed. Section section 4 contains a brief and specific analysis of such a situation in the case of financial data. Unlike the cases described in the references above, the asymptotics are taken when the number of observations increases, but the time length among observations remains fixed. In the case of a fOU process this problem has been studied by [26] for the estimation of λ when the Hurst parameter is fixed and is greater than $\frac{1}{2}$. The joint estimation of the pair (λ, σ) has been solved for discrete observations in [46] when H is kept fixed in $\left(\frac{1}{2}, \frac{3}{4}\right)$. It is important to note that [47] solves the parameter estimation of the entire set of parameters (H, λ, σ) for a fixed interval and infill asymptotics, i.e. drawing on availability of data with arbitrarily high sampling frequency.

In the present article, using the Generalized Method of Moments, we show how to achieve the joint estimation of any finite vector of parameters for stationary Gaussian sequences at the highest level of generality, using discrete observations, without requiring infill asymptotics, and we show that our setting applies directly to the joint estimation of H, λ and σ for the fOU process. Normal asymptotics are established thanks to the Breuer-Major theorem, and comments on efficiency are included. Our work is the first attempt to solve the joint estimation of the fOU process for its three parameters using a classical technique, without using infill asymptotics at all. We show via numerics for the fOU process that our technique is easily implementable, with good empirical precision and robustness properties even when operating with a moderate and realistic dataset size.

An interesting phenomenon occurs in the case of long-memory sequences such as discretely-observed fOU: if using only a Method of Moments based directly on the terms of the sequence itself, not its finite-differences, any estimator based on quadratic variations fails to be asymptotically normal when $H > 3/4$. This is a relatively well-known phenomenon, since that range of H falls beyond the Breuer-Major theorem's scope, though typically authors do not seek results in that range. Arguably, such results would be unnecessary since one only needs to consider one order filtered observations to avoid the non-normal asymptotics. However, it is instructive to investigate this case a bit further. Here we provide some calculations for the fOU case showing that the quadratic variation without any filtering cannot converge to a normal law, or even, which is more surprising, to a second-chaos law, but that it remains the basis for a strongly consistent estimator of the variance nonetheless. This requires the use of the Malliavin calculus.

This article is structured as follows: section 2 defines our GMM estimator for any stationary Gaussian process and states its general properties: consistency, asymptotic normality and efficiency. Section 3 is an application of the above results to the joint parameter estimation of the stationary solution of the fractional Ornstein-Uhlenbeck SDE. Finally, section 4 analyzes the numerical performance of the GMM estimator under a simulation study. Appendix A contain additional numerics. Appendix B contains mathematical proofs of technical lemmas and other results in the paper. Appendix C contains the convergence results, with proofs, for the non-finite-differenced estimator for fOU when $H > 3/4$.

2. Joint estimation of Gaussian stationary processes

Let X_t be a real-valued centered gaussian stationary process with spectral density $f_{\theta_0}(x)$, where $f_\theta(x)$ is a continuous function with respect to x, continuously differentiable with respect to θ (θ represents a vector of paremeters and it belongs to a compact set $\Theta \subset \mathbb{R}^p$) and $\theta_0 \in \text{Interior}(\Theta)$. Hereafter, θ_0 will be the true parameter of X_t. By Bochner's theorem, the autocovariance function of X_t can be written as:

$$\rho_\theta(s) := \text{Cov}(X_{t+s}, X_t) = \int_{\mathbb{R}} e^{isx} f_\theta(x) dx = \int_{\mathbb{R}} \cos(sx) f_\theta(x) dx \qquad (2.1)$$

for $s \geq 0$ and $\theta \in \Theta$.

Note that if we assume that $\rho_\theta(s)$ is a continuous function of s, then by [21] (page 257), the process X_t is ergodic. Take a positive integer $L \geq p$, and define:

$$\boldsymbol{\rho}_\theta(\alpha) := (\rho_\theta(\alpha \cdot 0), \dots, \rho_\theta(\alpha \cdot L))'. \qquad (2.2)$$

Now we have the condition that let us to identify the parameter θ.

Assumption 2.1 (Identifiability condition)**.** Take $\alpha > 0$. Then the subvector of $\boldsymbol{\rho}_\theta(\alpha)$:

$$\boldsymbol{\rho}_{\theta,p}(\alpha) = (\rho_\theta(\alpha \cdot 0), \dots, \rho_\theta(\alpha \cdot (p-1)))'.$$

is an injective map in θ. Assume also that the map $\boldsymbol{\rho}_{\theta,p}(\alpha)$ is differentiable.

Remark. Note that the above assumption can be obtained if $\nabla_\theta \boldsymbol{\rho}_{\theta,p}(\alpha)$ is a non-singular matrix or $\nabla_\theta \boldsymbol{\rho}_\theta(\alpha)$ is a full-column rank matrix for any $\theta \in \Theta$. Here we are using the notation $\nabla_\theta \cdot$ to mean the gradient of its argument with respect to θ.

The concept of filter will be used along this article. Its definition is as follows:

Definition 2.1. A filter $\mathbf{a} = (a_0, \ldots, a_L)$ of length $L + 1$ and order l is a sequence of $L + 1$ real numbers such that:

$$\sum_{q=0}^{L} a_q q^r = 0, \qquad 0 \le r \le l - 1,\ r \in \mathbb{Z}$$

$$\sum_{q=0}^{L} a_q q^l \ne 0$$

when $l > 0$. When $l = 0$, we will assume that $a_0 = 1$ and $a_q = 0$ for $0 < q \le L$.

In this article we will employ a family of discrete filters of length L and orders in $\{0, \ldots, L\}$. To simplify notation, we define for a filter a with order l:

$$b_k := \begin{cases} \sum_{q=0}^{L} a_q^2 & \text{if } k = 0 \\ 2 \sum_{j=0}^{L-k} a_{k+j} a_j & \text{if } k > 0 \end{cases} \tag{2.3}$$

and $\mathbf{b} := (b_0, \ldots, b_L)'$. It is straightforward to prove that given two different orders l_i, l_j, the corresponding vectors $\mathbf{b}_i$ and $\mathbf{b}_j$ are linearly independent. For these reason we can assume the following:

Assumption 2.2. Assume that we can choose L filters $\mathbf{a}_1, \ldots, \mathbf{a}_L$ with respective orders $l_1, \ldots, l_L$ such that the $L \times L + 1$ matrix:

$$B = \begin{bmatrix} \mathbf{b}_1' \\ \mathbf{b}_2' \\ \vdots \\ \mathbf{b}_L' \end{bmatrix}$$

satisfies that $p \le rank(B) \le L$.

Using this last assumption, we can define the function $V(\theta)$, for any $\theta \in \Theta$, as:

$$V(\theta) := \sum_{k=0}^{L} b_k \rho_\theta(\alpha k),$$

and the *filtered process with step size* $\alpha > 0$ (fixed) at $t \ge 0$ as:

$$\varphi(t) := \sum_{q=0}^{L} a_q X_{t-\alpha q}.$$

for a fixed filter $\mathbf{a}$. This filtered process was already considered in previous estimation studies, for example [4], [11], [14] and [27]. Due to the stationarity of

X_t, we can note that the variable $\varphi(t)$ is centered gaussian, hence the expected value of $\varphi(t)$ is zero and its variance is $E[\varphi(t)^2] = V(\theta_0)$, as follows:

$$
\begin{aligned}
E[\varphi(t)^2] &= \sum_{q,q'=0}^{L} a_q a_{q'} E[X_{t-\alpha q} X_{t-\alpha q'}] \\
&= \sum_{q,q'=0}^{L} a_q a_{q'} \rho_{\theta_0}(\alpha \cdot |q - q'|) \\
&= \sum_{k=0}^{L} b_k \rho_{\theta_0}(\alpha k) = V(\theta_0).
\end{aligned}
$$

Following the notation of [36] we define the vector of differences among the squared filtered observations for each of the L filters and its corresponding $V(\theta)$, for any $\theta \in \Theta$ as:

$$
\mathbf{g}(t, \theta) := (g_1(t, \theta), \dots, g_L(t, \theta))'
$$

where

$$
g_\ell(t, \theta) = \varphi_\ell(t)^2 - V_\ell(\theta), \qquad \text{for } 1 \le \ell \le L
$$

and the subscript ℓ means that we use the filter $\mathbf{a}_\ell$ in the computation of the filtered process and its variance entrywise. Clearly, in the jargon of Method of Moments' estimation, $\mathbf{g}(t, \theta)$ *satisfies a population moment condition* (see [22] and [36]):

$$
E[\mathbf{g}(t, \theta_0)] = 0
$$

for all $t \ge 0$. Also note that the vector $\mathbf{g}$ is a second-Wiener chaos vector (see [38]).

Let us assume that we have observed the process X_t at times $0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = T$ and fix $\alpha = t_i - t_{i-1}$.

Let A be a symmetric positive-definite matrix. We will denote $\|A\|$ as its matrix norm induced by the Euclidean norm in $\mathbb{R}^{L+1}$. It is important to note that A can be chosen to ensure that the GMM estimate is efficient (see section 2.3). Denote, for each $\theta \in \Theta$, and for an arbitrary time $t \ge 0$:

- $\mathbf{g}_0(\theta) := E[\mathbf{g}(t, \theta)]$ (vector of expected differences)
- $\hat{\mathbf{g}}_N(\theta) := \frac{1}{N-L+1} \sum_{i=L}^{N} \mathbf{g}(t_i, \theta)$ (vector of sampled differences)
- $Q_0(\theta) := \mathbf{g}_0(\theta)' A \mathbf{g}_0(\theta)$ (squared distance of expected differences)
- $\hat{Q}_N(\theta) := \hat{\mathbf{g}}_N(\theta)' A \hat{\mathbf{g}}_N(\theta)$ (sampled version of the above distance)

Note that $\mathbf{g}_0(\theta)$ does not depend on time t, due to the stationarity assumption on X_t. Finally, we define the GMM (Generalized Method of Moments) estimator of θ_0 as (see [24], [22]):

$$
\hat{\theta}_N = \mathrm{argmin}_{\theta \in \Theta} \hat{Q}_N(\theta). \tag{2.4}
$$

The main idea behind the GMM estimation is that since the function $Q_0(\theta)$ attains a unique zero at θ_0 under Assumptions 2.1 and 2.2 (see Lemma 2.2 below),

we would want to find a value $\hat{\theta}_N$ (that depends on the current observations) such that our weighted empirical approximation to Q_0, namely $\hat{Q}_N$, should be as small as possible.

It is easy to prove that all the remaining results are still valid if we substitute the fixed matrix A with a sequence of random matrices (perhaps depending on the data) with a deterministically bounded eigenstructure. In particular, we can choose such sequence in order to attain convergence in probability to the efficient alternative of A. For more details see section 2.3.

2.1. Consistency

In this section we will prove the strong consistency of the estimator defined in (2.4). Our first lemma shows that the sequence $\hat{Q}_N(\theta)$ is uniformly convergent for any $\theta \in \Theta$.

Lemma 2.1. *It holds that:*

$$\sup_{\theta \in \Theta} |\hat{Q}_N(\theta) - Q_0(\theta)| \overset{a.s}{\to} 0.$$

Proof. See Appendix B.   □

Under the injectivity assumptions on $\boldsymbol{\rho}_\theta(\alpha)$, we can prove:

Lemma 2.2. *Under Assumptions 2.1 and 2.2:*

$$Q_0(\theta) = 0.$$

if and only if $\theta = \theta_0$.

Proof. See Appendix B.   □

And these two lemmas allow us to prove:

Theorem 2.1 (Strong Consistency). *Under Assumptions 2.1 and 2.2, it holds that:*

$$\hat{\theta}_N \overset{a.s.}{\to} \theta_0.$$

Proof. Because of Lemma 2.1 and 2.2 we can apply Theorem 2.1 in [36] and the result holds.   □

2.2. Asymptotic normality

In this section we use the Breuer-Major theorem (see [6]) to prove the asymptotic normality of $\hat{\theta}_N$, under some specific assumptions on $\boldsymbol{\rho}_\theta(\alpha)$ and $f_\theta(x)$. First denote for any $\theta \in \Theta$:

$$\hat{G}_N(\theta) := \nabla_\theta \hat{\mathbf{g}}_N(\theta)$$
$$G(\theta) := E[\nabla_\theta \mathbf{g}(\cdot, \theta)].$$

Note that $G(\theta)$ does not depend on t because of the stationarity of X_t. Now we can state the following lemma:

Lemma 2.3. *Let $\theta \in \Theta$. The following is true:*

(i) $\hat{G}_N(\theta) = G(\theta) = -\nabla_\theta \mathbf{V}(\theta)$. *Moreover, $G(\theta)$ is a continuous function of θ.*

(ii) *Under Assumptions 2.1 and 2.2 the matrix $G(\theta)'DG(\theta)$ is non-singular for any positive-definite matrix D.*

Proof. See Appendix B. $\square$

The fact that X_t is a stationary process allows us to write this gaussian process as a Wiener-Itô integral with respect to a complex centered Gaussian random measure W over $[-\frac{\pi}{\alpha}, \frac{\pi}{\alpha}]$ as follows (see Section 3.2 in [33], Theorem 2.7.7 in [38] and Section 2.1 in [5]),

$$X_{t_k} = \int_{-\frac{\pi}{\alpha}}^{\frac{\pi}{\alpha}} e^{it_k x} \bar{f}_{\theta_0}(x)^{1/2} dW(x) = I_1(A_k(\cdot|\theta_0)) \tag{2.5}$$

where $A_k(x|\theta_0) := e^{it_k x} \bar{f}_{\theta_0}(x)^{1/2}$ is the corresponding kernel under this representation for $k \in \{0, \ldots, N\}$ and $\bar{f}_{\theta_0}(x)$ is the spectral density of X_t (for $t_k = k \cdot \alpha$, $k \in \mathbb{Z}$):

$$\bar{f}_{\theta_0}(u) := \sum_{p \in \mathbb{Z}} f_{\theta_0}\left(u + \frac{2\pi}{\alpha}p\right), \qquad u \in \left[-\frac{\pi}{\alpha}, \frac{\pi}{\alpha}\right]. \tag{2.6}$$

Assume that we choose n and m such that $t_n < t_m$. The covariance between X_{t_m} and X_{t_n} can be computed as:

$$\begin{aligned}
\mathrm{Cov}(X_{t_m}, X_{t_n}) &= \rho_{\theta_0}(|t_m - t_n|) \\
&= \frac{\alpha}{2\pi} \int_{-\frac{\pi}{\alpha}}^{\frac{\pi}{\alpha}} e^{i|m-n|\alpha x} \bar{f}_{\theta_0}(x) dx \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i|m-n|u} \bar{f}_{\theta_0}\left(\frac{u}{\alpha}\right) du. \tag{2.7}
\end{aligned}$$

For the next lemma, we need to generalize the definition of b_k in (2.3) in the following way:

$$b_k^{i,j} := \sum_{q=0}^{L-k} a_q^i a_{q+k}^j + \sum_{q=0}^{L-k} a_q^j a_{q+k}^i$$

where $i, j = 1, \ldots, L$ and a_k^i is the k-th entry of the filter $\mathbf{a}_i$ (recall the notation in Assumption 2.2). The main condition in the Breuer-Major theorem is the summability of the second-power of the autocovariance function at integer times (see [6]). In order to validate this argument we can make the following assumption, which is equivalent to the above fact:

Assumption 2.3. Assume that for any $i, j \in \{1, \ldots, L\}$ we have that:

$$R(u|i,j) := \min\left(1, |u|^{l_i + l_j}\right) \bar{f}_{\theta_0}(u) \in L^2[(-\pi, \pi)],$$

where l_i is the order of the filter $\mathbf{a}_i$.

The assumption above and the fact that the vector $\hat{\mathbf{g}}_N(\theta_0)$ can be rescaled to obtain a sum of Hermite processes of second order, are sufficient conditions to prove the asymptotic normality of the sample moment equations, as follows:

Lemma 2.4. *Under Assumptions 2.1-2.3:*

$$\sqrt{N}\hat{\mathbf{g}}_N(\theta_0) \overset{d}{\to} \mathcal{N}(0, \Omega)$$

where Ω is a $L \times L$ matrix whose components are:

$$\Omega_{ij} = 2 \sum_{k \in \mathbb{Z}} \left[\sum_{p=0}^{L} b_p^{i,j} \rho_{\theta_0} \left[\alpha(k+p) \right] \right]^2. \tag{2.8}$$

Proof. See Appendix B.   □

The next natural step in this analysis is to study the asymptotic behavior of the Mean Squared error (MSE) of $\hat{\theta}_N$. Note that from equation (2.5) the filtered process can be represented for any $\ell \in \{1, \ldots, L\}$ and $i \in \{0, 1, \ldots, N\}$ as:

$$\varphi_\ell(t_i) = \sum_{q=0}^{L} a_q X_{t_i - \alpha q} = I_1(SA_{i,\ell}(\cdot|\theta_0)) \tag{2.9}$$

where $SA_{i,\ell}(u|\theta_0) := \sum_{q=0}^{L} a_q A_{i-q}(u|\theta_0)$, $u \in \mathbb{R}$. Hence, using the multiplication rule of Wiener integrals (see [40]), the componentwise average of the sample moment equations is:

$$(\hat{\mathbf{g}}_N(\theta))_\ell = \frac{1}{N-L+1} \sum_{j=L}^{N} [\varphi_\ell(t_j)^2 - V_\ell(\theta)]$$

$$= \frac{1}{N-L+1} \sum_{j=L}^{N} I_2(SA_{\ell,j} \otimes SA_{\ell,j}) = I_2[B_{\ell,j}(\cdot|\theta)]$$

where $B_{\ell,j}(\cdot|\theta) := \frac{1}{N-L+1} \sum_{j=L}^{N} SA_{\ell,j} \otimes SA_{\ell,j}$ and $\ell \in \{1, \ldots, L\}$. Then we can conclude that $\hat{\mathbf{g}}_N(\theta)$ is a vector of L integrals belonging to the second Wiener-chaos of W. Because of Lemma 2.4, we already know that:

$$\sqrt{N}(\hat{\mathbf{g}}_N(\theta_0))_\ell \overset{d}{\longrightarrow} \mathcal{N}(0, \Omega_{\ell,\ell})$$

and therefore $E[|(\hat{\mathbf{g}}_N(\theta_0))_\ell|^2] = O(N^{-1})$. Hence there exists a constant $K_\ell > 0$ such that:

$$E[|(\hat{\mathbf{g}}_N(\theta_0))_\ell|^2] \leq \frac{K_\ell}{N}.$$

so, we can use the fact that $(\hat{\mathbf{g}}_N(\theta_0))_\ell$ belongs to the second-Wiener chaos generated by the process $W(x)$ and furthermore we can use the equivalence of all the L^s-norms of $(\hat{\mathbf{g}}_N(\theta_0))_\ell$ to get: (see Theorem 3.50 in [28])

$$E[|(\hat{\mathbf{g}}_N(\theta_0))_\ell|^{4s}] \leq c_s[E[|(\hat{\mathbf{g}}_N(\theta_0))_\ell|^2]]^{2s} \leq \frac{c_s \cdot K_\ell}{N^{2s}}$$

where $c_s > 0$. Using Jensen's inequality:

$$E[\|\hat{\mathbf{g}}_N(\theta_0)\|^{4s}] = E\left[\left|\sum_{\ell=1}^{L} |(\hat{\mathbf{g}}_N(\theta_0))_\ell|^2\right|^{2s}\right] \leq L^{2s-1} \sum_{\ell=1}^{L} E\left[|(\hat{\mathbf{g}}_N(\theta_0))_\ell|^{4s}\right]$$
$$\leq \frac{L^{2s} c_s}{N^{2s}} \max_{1 \leq \ell \leq L} \{K_\ell\}.$$

All these calculations allow us to study the behavior of the MSE of $\hat{\theta}_N$. Let $\varepsilon_N := \hat{\theta}_N - \theta_0$. First note that because of the multivariate mean value theorem (see Proposition F.6.1 in [31] and more details in [36]):

$$\varepsilon_N := \hat{\theta}_N - \theta_0 = -[\hat{G}_N(\hat{\theta}_N)' A \hat{G}_N(\bar{\theta}_N)]^{-1} \cdot \hat{G}_N(\hat{\theta}_N)' A \hat{g}_N(\theta_0)$$
$$= -\psi_N(\bar{\theta}_N, \hat{\theta}_N) \cdot \hat{g}_N(\theta_0)$$

where $\bar{\theta}_N$ is a value in the interval $[\theta_0, \hat{\theta}_N]$,

$$\psi_N(\bar{\theta}_N, \hat{\theta}_N) := [G(\hat{\theta}_N)' A G(\bar{\theta}_N)]^{-1} \cdot G(\hat{\theta}_N)' A \qquad (2.10)$$

and $\theta \in \Theta$ (see Lemma 2.3). The next lemma states that all moments of ψ_N can be uniformly bounded:

Lemma 2.5. *Let ψ_N as in equation (2.10). Under the remark after assumption 2.1 there exists $R_{s,L} > 0$ such that:*

$$E[\|\psi_N(\bar{\theta}_N, \hat{\theta}_N)\|^{4s}] < R_{s,L}.$$

Proof. See Appendix B. $\qquad\qquad\square$

Note also that because of the continuity of G:

$$\psi_N(\bar{\theta}_N, \hat{\theta}_N) \stackrel{a.s.}{\to} [G(\theta_0)' A G(\theta_0)]^{-1} G(\theta_0)' A.$$

and due to Lemma 2.5, for any $s > 0$:

$$E[\|\varepsilon_N\|^{2s}] = E\left[\|\psi_N(\bar{\theta}_N, \hat{\theta}_N) \cdot \hat{\mathbf{g}}_N(\theta_0)\|^{2s}\right] = E\left[\|\psi_N(\bar{\theta}_N, \hat{\theta}_N)\|^{2s} \cdot \|\hat{\mathbf{g}}_N(\theta_0)\|^{2s}\right]$$
$$\leq E\left[\|\psi_N(\bar{\theta}_N, \hat{\theta}_N)\|^{4s}\right]^{1/2} E\left[\|\hat{\mathbf{g}}_N(\theta_0)\|^{4s}\right]^{1/2} < \frac{\mathcal{R}_{s,L}}{N^s}$$

where $\mathcal{R}_{s,L} = (L \cdot R_{s,L} \cdot c_s \cdot \max_\ell \{K_\ell\})^{1/2}$, and taking $s = 1$ we obtain an upper bound for the MSE of θ_N. Furthermore, if we choose $s > 1$, we can apply Borel-Cantelli lemma to get:

$$\varepsilon_N \xrightarrow{a.s} 0$$

If we take $\gamma > 0$, and under the same reasoning as before, we can use Chebyshev's inequality to get:

$$P[\|\varepsilon_N\| > N^{-\gamma}] \leq \frac{E[\|\varepsilon_N\|^{2s}]}{N^{-2\gamma s}} < \frac{\mathcal{R}_{s,L}}{N^{(1-2\gamma)s}}$$

and hence for any $\gamma < \frac{1}{2}$ there exists s such that $s > \frac{1}{1-2\gamma}$ and, by Borel-Cantelli lemma there exists $N_0(\omega) < \infty$ such that for all $N > N_0(\omega)$:

$$N^\gamma \|\varepsilon_N\| < c \qquad \text{a.s.}$$

where c is an arbitrary positive constant. Hence, we conclude the following theorem:

Theorem 2.2. *Under Assumptions 2.1 and 2.2 we have:*

(i) $E[\|\hat{\theta}_N - \theta_0\|^2] = O(N^{-1})$
(ii) $N^\gamma \|\hat{\theta}_N - \theta_0\| \xrightarrow{a.s.} 0$ *for any* $\gamma < \frac{1}{2}$.

Now we can state a CLT for $\hat{\theta}_N$, under Assumptions 2.1-2.3:

Theorem 2.3 (Asymptotic Normality)**.** *Let* X_t *be a Gaussian stationary process with parameter* θ_0. *If* $\hat{\theta}_N$ *is the GMM estimator given in (2.4), then under Assumptions 2.1-2.3, it holds that:*

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{C}(\theta_0)\Omega\mathcal{C}(\theta_0)')$$

where $\mathcal{C}(\theta_0) = [G(\theta_0)'AG(\theta_0)]^{-1} G(\theta_0)'A$ *and* Ω *is defined as in Lemma 2.4.*

Proof. Lemmas 2.3 and 2.4 gives sufficient conditions to apply Theorem 3.4 in [36], and conclude the asymptotic normality of $\hat{\theta}_N$. □

2.3. Practical considerations

If the number of moment conditions L is greater than the number of parameters p, the asymptotic variance of $\sqrt{N}(\hat{\theta}_N - \theta_0)$ is minimized when $A = \Omega^{-1}$ (see Theorem 3.4 of [22] and [36] for a detailed proof of this fact). In the case of a number of moment equations L equal to the number of parameters p, the estimation problem is equivalent to the Method of Moments and hence $A = I$.

There are some numerical methods to compute a sequence of positive-definite matrices such that $\hat{A}_N \xrightarrow{d} A$ in such a way that $\hat{A}_N$ depends on the data available. This type of procedure gives numerical advantages, specifically when $A = \Omega^{-1}$, since the previous matrix depends on the unknown parameter θ, while the sequence $\hat{A}_N$ do not depend on this parameter. Three methods that are commonly used to achieve the above are:

1. Hansen's Two-Step estimator (see [24]),
2. Iterative estimator and
3. Continuous-Updating estimator (CUE) of Hansen, Heaton and Yaron (see [23]).

For ease in our calculations in sections 3 and 4, we decided to use a constant matrix $A = \Omega^{-1}$ under the case where $L > p$ and $A = I$ otherwise.

3. Parameter estimation of the fractional Ornstein-Uhlenbeck process

The main objective of this section is to apply the GMM method in the parameter estimation of a particular gaussian stationary process: the fractional Ornstein-Uhlenbeck process (fOU), under an increasing domain approach. First we will give a brief introduction of the fOU process.

3.1. Introduction

Cheridito [10] defines the fractional Ornstein-Uhlenbeck process with initial condition $\xi \in L^0(\Omega)$, as the unique almost surely solution of the Langevin equation:

$$X_t = \xi - \lambda \int_0^t X_s ds + \sigma B_t^H, \qquad t \geq 0,$$

where $\lambda, \sigma > 0$, $H \in (0,1]$ and we use the notation $\theta = (H, \lambda, \sigma)$. The stationary solution of the above equation is written as:

$$X_t = \sigma \int_{-\infty}^t e^{-\lambda(t-u)} dB_u^H, \qquad t \in \mathbb{R}, \tag{3.1}$$

where the integral is understood in the sense of Riemann-Stieltjes. Since this process is centered-gaussian and stationary, we can describe its distributional properties by using only its autocovariance function:

$$\rho_\theta(t) = 2\sigma^2 c_H \int_0^\infty \cos(tx) \frac{x^{1-2H}}{\lambda^2 + x^2} dx \tag{3.2}$$

where $H \in (0,1)$ and $c_H = \frac{\Gamma(2H+1)\sin(\pi H)}{2\pi}$. Note that when $H \geq \frac{1}{2}$ we can obtain an alternative expression for $\rho_\theta(t)$ (see Lemma B.1 and formula (B.4) in the Appendix B). Note that using equation (3.2), we can write an explicit expression of the spectral density of X_t in continuous time:

$$f_\theta(x) = \sigma^2 c_H \frac{x^{1-2H}}{x^2 + \lambda^2}.$$

Throughout this section we will assume that there exists a closed rectangle $\Theta \subset \mathbb{R}^2$ such that $\theta \in \Theta$.

3.2. Consistency

For ease of computation we will study the joint estimation of the three parameters in θ when $H \geq \frac{1}{2}$, and we will employ the autocovariance function obtained in formula (B.4) only. Note that $f_\theta(x)$ is continuous with respect to x, H, λ and σ and its partial derivatives are also continuous functions of its parameters.

Assumption 2.1 is verified locally because of Lemma B.2 in Appendix B. Note that the sufficient conditions in order to the above assumption to hold are (1) $\alpha < 1$, (2) σ is of known sign and (3) $0 < \lambda < \exp(\Psi(3))$. The condition (2) does not change the magnitude of the variance of the process X_t nor its scale, since the process is centered. Finally, the joint GMM estimator defined in (2.4) is consistent under the previous limitations.

The ergodicity of X_t is deduced from the expansion of the function $\rho_\theta(x)$ when $H \neq \frac{1}{2}$ (see [10]):

$$\rho_\theta(x) = \frac{H(2H-1)}{\lambda} x^{2H-2} + O(x^{2H-4}), \qquad \text{as } x \to \infty. \tag{3.3}$$

3.3. Asymptotic normality

Assume that we have a collection of at least 3 filters satisfying Assumption 2.2. Hence, we can study under which cases Assumption 2.3 is satisfied:

Lemma 3.1. *Let X_t be the stationary fOU process with parameters $\theta = (H, \lambda, \sigma)$. The Assumption 2.3 holds under the following two cases:*

Case 1 *If $l + l' \geq 1$ then it holds for all $H \in (0, 1)$,*
Case 2 *If $l + l' = 0$ then it holds if $H \in (0, \frac{3}{4})$,*

where $L \geq 3$.

Proof. See Appendix B. $\qquad\qquad\square$

Since the case 1 in the above lemma guarantees the asymptotic normality of the moment equations for any $H \in (0, 1)$, we decided to use filters with order greater or equal to 1. If this is the case we can use the previous lemma together with Lemma 2.4 to obtain the asymptotics of $\hat{\mathbf{g}}_N(\theta)$ as follows:

$$\sqrt{N}\hat{\mathbf{g}}_N(\theta) \xrightarrow{d} N(0, \Omega(\theta)) \tag{3.4}$$

where the covariance matrix $\Omega(\theta)$ has entries according to equation (2.8).

The behavior of

$$(\hat{\mathbf{g}}_N(\theta))_0 := \frac{1}{N - L + 1} \sum_{i=1}^{N} \varphi_0(t_i)^2 - V_0(\theta)$$

deserves interest by itself due to its asymptotic behavior when $H \geq \frac{3}{4}$ and whose proofs rely heavily on Malliavin calculus techniques. We summarized these results in Appendix C.

We finish this section by applying the general results of the previous section to obtain consistency and asymptotic normality for the fOU's GMM estimator. Assumption 2.1 on injectivity of the estimator is satsified, as shown in Lemma B.2 in Appendix B; this lemma establishes this assumption for the fOU observations under an upper bound on λ and for α sufficiently small, where these conditions do not depend on N. First, we can use the general result in Theorem 2.2 to conclude the following rates of convergence of the GMM estimator:

Proposition 3.1. *Let X_t be a fOU process with parameters $\theta = (H, \lambda, \sigma)$. Then, for any positive-definite matrix A and under the conditions of Lemma B.2, the GMM estimator $\hat{\theta}_N$ satisfies:*

(i)

$$E[\|\hat{\theta}_N - \theta\|^2] = O(N^{-1})$$

(ii) As N goes to infinity:

$$N^\gamma \|\hat{\theta}_N - \theta\| \xrightarrow{a.s.} 0$$

for all $\gamma < \frac{1}{2}$.

Now we are ready to state the asymptotic distribution of our joint estimator, using the limiting behavior of $\hat{\mathbf{g}}_N$:

Proposition 3.2. *Let X_t be the stationary fOU process with parameters $\theta = (H, \lambda, \sigma)$. Then, for any positive-definite matrix A and under the conditions of Lemma B.2, the GMM estimator $\hat{\theta}_N$ of θ in equation (2.4) is consistent for any $H \in (0, 1)$ and:*

$$\sqrt{N}(\hat{\theta}_N - \theta) \xrightarrow{d} N(0, \mathcal{C}(\theta)\Omega\mathcal{C}(\theta)') \tag{3.5}$$

where $\mathcal{C}(\theta) = [G(\theta)'AG(\theta)]^{-1}G(\theta)'A$ and Ω according to equation (2.8).

Proof. The consistency is immediate from Theorem 2.1, since X_t satisfies Assumption 2.1 locally. The asymptotic normality is attained since assumption 2.3 holds for any pair of filter lengths whose sum is greater or equal to 1, which is the case in this example. Therefore, we can apply Theorem 2.3 to get the desired result. □

As we noted in section 2.3, we chose $A = \Omega^{-1}$ because this ensures that the asymptotic variance is minimal. If this is the case the asymptotic variance can be simplified to:

$$[G(\theta)'\Omega^{-1}G(\theta)]^{-1}.$$

4. Numerical performance of the GMM estimators for the fractional OU process

In this section we test empirically the properties of the GMM estimator of (H, λ, σ) for the fOU case. We are especially interested in checking the performance of the joint estimator for a fixed value of N, particularly when N is not sufficiently large that the constants coming from the asymptotic variances become secondary concerns. As noted below, we run tests for $N = 1000$ and $N = 600$ observation times of a single path of the stochastic process, which represents a realistic dataset size for such questions as the estmation of memory length in financial markets under stochastic volatility (see [13]) or in late-Holocene paleoclimatology (See [1]), where single observation paths at discrete times are all that is available. These are examples of situations which are not

data-rich, but where the amount of data should be sufficient to provide evidence of long memory. Specifically in the case of financial markets with stochastic volatility memory, evidence exists that such memory does not persist for longer than a few weeks at a given level. Assuming for instance that one has access to 20 working days of stock-, option-, and/or volatility-market data, which corresponds to calibrating models to one-month options, in order to stay in the realm of continuous-time models, i.e. to avoid having to consider jump models to account for market microstructure, the best one typically hopes for (liberal estimate) is to observe prices every 5 to 15 minutes depending on option liquidity. This would correspond to roughly 600 to 2000 datapoints before one would need to recalibrate or change the model, hence our order of magnitudes for N. This section provides empirical estimates of the bias and variance of our joint estimator on all three parameters, and a comparison to the the joint estimation of the pair (H, σ) when λ is fixed.

First we choose a fixed step size among observations (α). In order to simulate the fOU process we employ the formula (B.4) together with the Choleski decomposition of the autocovariance matrix of this gaussian process. This procedure gives us m independently-generated paths of the fOU process, each such path, sampled at the $N = 1000$ observation times, representing one replication of the single-trajectory dataset available in practice. In our simulations we use $m = 1000$ replications. We decided to use finite-difference filters since they are the most typical p-vanishing-moments filters in the literature (see [14, 11, 8] for example). The filters were selected sequentially, i.e. we choose $l_i = i$ for $i = 1, \ldots, L$, for the purpose of ensuring that the respective matrix B satisfies Assumption 2.2 in page 6.

In order to solve numerically for $\hat{\theta}_N = (\hat{H}_N, \hat{\lambda}_N, \hat{\sigma}_N)$ in (2.4), we use the quasi-Newton L-BFGS-R algorithm contained in the R-routine `optim` (see [9] and [37]). This optimization algorithm is designed for solving non-linear problems with simple bounds. In order to check that there was an acceptable level of variability produced by the optimization method, we compared the theoretical bounds induced in formula (3.5) with the empirical variance $e(Var)$ defined the largest eigenvalue of the empirical covariance matrix of the estimator vector. 99% of the confidence regions constructed with the empirical variance contained the true parameter for each scenario.

Finally, we ran different scenarios for the purpose of verification of our GMM procedure. We chose four different values of the parameter H (0.55, 0.65, 0.75 and 0.85) so that it is possible to assess the effect of increasing levels of long-memory in our estimates' performance. In addition, for each of the above cases we worked with subscenarios with an increase of 10% in λ and a decrease of 10% in the value of σ, as a quick check on the robustness the empirical statistics. We noted no significant aberrant behavior in these various scenarios; though all results are reported below, readers may safely concentrate on the case $\lambda = \sigma = 1$. We chose a maximum number of filters equal to 7, and since there are three parameters to estimate, we selected an initial order equal to 3. The scenarios were chosen so that we could analyze the impact of small changes on the actual values of the parameters in a given fOU process.

TABLE 1
Comparison of Scenarios ($\alpha = 0.1$).

		$\lambda = 1, \sigma = 1$			$\lambda = 1.1, \sigma = 1$			$\lambda = 1, \sigma = 0.9$		
H	L	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$
	3	0.058	0.057	0.000	0.054	0.053	0.000	0.048	0.047	0.000
	4	0.049	0.047	0.000	0.048	0.047	0.000	0.042	0.041	0.000
0.25	5	0.045	0.043	0.000	0.045	0.043	0.000	0.040	0.039	0.000
	6	0.043	0.042	0.000	0.045	0.043	0.000	0.038	0.037	0.000
	7	0.043	0.042	0.000	0.045	0.043	0.000	0.038	0.037	0.000
	3	0.065	0.062	0.001	0.082	0.078	0.002	0.066	0.063	0.001
	4	0.058	0.056	0.001	0.074	0.070	0.002	0.062	0.060	0.001
0.40	5	0.053	0.051	0.000	0.067	0.064	0.002	0.058	0.056	0.001
	6	0.052	0.049	0.000	0.068	0.064	0.002	0.056	0.054	0.001
	7	0.052	0.049	0.000	0.068	0.064	0.002	0.057	0.054	0.001
	3	0.061	0.056	0.002	0.080	0.074	0.003	0.062	0.059	0.002
	4	0.060	0.055	0.002	0.079	0.074	0.003	0.063	0.059	0.002
0.55	5	0.060	0.055	0.002	0.080	0.075	0.003	0.064	0.060	0.002
	6	0.061	0.055	0.002	0.081	0.075	0.003	0.063	0.059	0.002
	7	0.061	0.055	0.002	0.081	0.075	0.003	0.064	0.060	0.002
	3	0.115	0.103	0.008	0.136	0.124	0.007	0.112	0.102	0.006
	4	0.115	0.103	0.007	0.130	0.119	0.007	0.105	0.096	0.005
0.70	5	0.109	0.099	0.006	0.126	0.116	0.006	0.102	0.094	0.004
	6	0.104	0.095	0.005	0.120	0.111	0.005	0.099	0.091	0.004
	7	0.106	0.097	0.005	0.118	0.109	0.005	0.096	0.088	0.004
	3	0.258	0.203	0.042	0.248	0.202	0.034	0.233	0.191	0.031
	4	0.219	0.197	0.009	0.244	0.225	0.008	0.224	0.205	0.009
0.85	5	0.179	0.162	0.005	0.205	0.189	0.005	0.185	0.171	0.004
	6	0.160	0.145	0.004	0.177	0.163	0.003	0.163	0.150	0.003
	7	0.156	0.141	0.004	0.169	0.155	0.003	0.154	0.141	0.003

Tables 1 and 2 comprise three empirical metrics of comparison among scenarios:

$$\widehat{MSE} := \frac{1}{m} \sum_{i=1}^{m} \|\hat{\theta}_{N,i} - \theta\|^2$$

$$e(\widehat{Var}) := \text{maximum eigenvalue of } \widehat{Var}(\theta_N)$$

$$\widehat{Bias}^2 := \left\| \frac{1}{m} \sum_{i=1}^{m} \hat{\theta}_{N,i} - \theta \right\|^2 .$$

where $\hat{\theta}_{N,i}$ is the GMM estimate of θ using a sample size of N which was obtained using the i-th fOU replication. $\widehat{Var}(\theta_N)$ is the empirical covariance matrix of the GMM estimator based on the m replications. These metrics are computed as the number of filters increases (i.e. as L increases). The calculations of Table 1 use a sample size $N = 1000$ and table 2 use $N = 600$. The second table also uses an increased time-step of $\alpha = 0.5$ which in principle might affect estimator accuracy, though this seems to also depend on the value of H. We have also computed the empirical statistics when λ is assumed to be known, and when

TABLE 2
Comparison of Scenarios ($\alpha = 0.5$).

H	L	$\lambda = 1, \sigma = 1$			$\lambda = 1.1, \sigma = 1$			$\lambda = 1, \sigma = 0.9$		
		$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$
	3	0.314	0.295	0.017	0.355	0.325	0.028	0.337	0.312	0.023
	4	0.275	0.259	0.015	0.368	0.335	0.031	0.289	0.265	0.022
0.25	5	0.264	0.248	0.014	0.348	0.320	0.027	0.272	0.250	0.021
	6	0.261	0.245	0.014	0.344	0.318	0.025	0.264	0.244	0.019
	7	0.265	0.248	0.016	0.349	0.321	0.027	0.264	0.243	0.019
	3	0.147	0.143	0.002	0.169	0.166	0.001	0.145	0.141	0.003
	4	0.137	0.133	0.002	0.154	0.152	0.000	0.135	0.132	0.002
0.40	5	0.132	0.128	0.002	0.149	0.147	0.000	0.133	0.130	0.002
	6	0.131	0.128	0.002	0.151	0.149	0.000	0.130	0.127	0.002
	7	0.132	0.129	0.002	0.150	0.148	0.000	0.130	0.127	0.001
	3	0.198	0.169	0.030	0.256	0.224	0.043	0.207	0.174	0.027
	4	0.197	0.167	0.028	0.254	0.211	0.041	0.202	0.166	0.035
0.55	5	0.186	0.157	0.027	0.240	0.198	0.041	0.189	0.156	0.031
	6	0.173	0.148	0.023	0.212	0.178	0.032	0.179	0.148	0.029
	7	0.173	0.149	0.023	0.225	0.188	0.036	0.180	0.148	0.030
	3	0.174	0.169	0.001	0.254	0.230	0.020	0.170	0.166	0.000
	4	0.207	0.201	0.002	0.237	0.229	0.005	0.216	0.210	0.003
0.70	5	0.179	0.174	0.002	0.218	0.211	0.004	0.184	0.180	0.001
	6	0.170	0.166	0.001	0.208	0.202	0.003	0.178	0.175	0.001
	7	0.173	0.168	0.001	0.208	0.203	0.002	0.175	0.172	0.001
	3	0.201	0.188	0.002	0.204	0.193	0.002	0.168	0.158	0.001
	4	0.195	0.181	0.004	0.206	0.191	0.006	0.167	0.154	0.005
0.85	5	0.177	0.164	0.003	0.189	0.175	0.006	0.156	0.144	0.005
	6	0.168	0.156	0.002	0.175	0.163	0.004	0.142	0.131	0.004
	7	0.166	0.155	0.002	0.175	0.163	0.004	0.145	0.134	0.004

using another class of filters, with results shown in Appendix A. Below are some conclusions:

1. The empirical MSE ($\widehat{MSE}$) seems to decrease for a fixed scenario when the number of filters (number of moment conditions L) increases. For $\alpha = 0.1$, we observe an increase in the empirical MSE as the Hurst parameter increases; this fact does not persist when $\alpha = 0.5$, where the empirical MSEs are larger by up to a factor of 3, particularly when H is small. The case of only 600 datapoints with a larger time interval appears to reach the limit of our joint estimation method's reliability when trying to estimate all three paramters together, since the MSE there exceeds 10% of the value of the parameter routinely in this case. See however, the comments below when λ is assumed known.

2. As the drift and diffusion parameters λ and σ are perturbed, the empirical MSE seems to change mainly due to the change in the magnitude of the variance when that happens.

3. Sensitivity of $\widehat{Bias}^2$ and $e(\widehat{Var})$ to λ and σ are noted as moderate, and for each case the change persists as the number of filters increases. Dependence

of $e(\widehat{Var})$ on σ is easily explained in theory since the asymptotic covariance in Proposition 3.2 is proportional to σ^2. The two empirical statistics appear to be largely insensitive to the number of filters, however.

4. If we compare Tables 1 and 2 with Tables 3 and 4 in Appendix A, where the estimation is performed keeping the parameter λ fixed on each scenario, we notice that most of the empirical MSE is due to the variation in λ estimation. The level of precision in essentially all scenarios is higher than 100 to 1, and the estimators can be considered as empirically unbiased. This holds even in the unfavorable case of 600 observation times and larger time step (Table 4). We also note that sensitivity of the estimators' performances to the number of filters is very low, except for large values of H where it seems preferably to choose at least $L = 3$ even though there are only two parameters to estimate.

5. Finally Tables 5 and 6 in Appendix A contain the joint estimation of the three parameters under the same scenarios when the filters are Daubechies wavelets. We note that if we compare those tables with Tables 1 and 2, performance is essentially unchanged from one filter-class to the other. The slight improvement from Daubechies filters compared to finite-difference filters when the Hurst parameter is larger than $\frac{1}{2}$ does not seem to be significant.

Appendix A: Additional tables

TABLE 3
Comparison of Scenarios ($\alpha = 0.1$). Estimation of (H, σ) with λ fixed.

H	L	$\lambda = 1, \sigma = 1$			$\lambda = 1.1, \sigma = 1$			$\lambda = 1, \sigma = 0.9$		
		$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$
	2	0.00174	0.00166	1.26×10^{-06}	0.00180	0.00171	1.58×10^{-06}	0.00153	0.00145	1.30×10^{-06}
	3	0.00134	0.00125	5.79×10^{-07}	0.00130	0.00121	3.81×10^{-07}	0.00116	0.00108	4.23×10^{-07}
0.25	4	0.00128	0.00119	2.62×10^{-07}	0.00128	0.00119	2.87×10^{-07}	0.00113	0.00104	5.19×10^{-07}
	5	0.00128	0.00120	2.92×10^{-07}	0.00128	0.00119	2.05×10^{-07}	0.00113	0.00105	5.36×10^{-07}
	6	0.00128	0.00120	1.11×10^{-07}	0.00128	0.00119	1.57×10^{-07}	0.00113	0.00104	5.29×10^{-07}
	2	0.00275	0.00268	1.15×10^{-06}	0.00290	0.00281	2.32×10^{-05}	0.00235	0.00227	6.58×10^{-07}
	3	0.00218	0.00211	3.49×10^{-06}	0.00225	0.00217	1.15×10^{-05}	0.00175	0.00167	1.54×10^{-07}
0.40	4	0.00215	0.00207	4.15×10^{-06}	0.00218	0.00210	9.43×10^{-06}	0.00170	0.00163	1.72×10^{-07}
	5	0.00213	0.00206	4.25×10^{-06}	0.00218	0.00210	9.19×10^{-06}	0.00170	0.00163	2.50×10^{-07}
	6	0.00214	0.00206	4.49×10^{-06}	0.00217	0.00209	8.77×10^{-06}	0.00169	0.00161	2.73×10^{-07}
	2	0.00442	0.00435	1.62×10^{-05}	0.00413	0.00404	1.65×10^{-05}	0.00372	0.00364	1.38×10^{-05}
	3	0.00318	0.00311	1.76×10^{-05}	0.00307	0.00299	1.54×10^{-05}	0.00273	0.00267	1.44×10^{-06}
0.55	4	0.00312	0.00304	1.71×10^{-05}	0.00306	0.00298	1.87×10^{-05}	0.00268	0.00262	1.12×10^{-06}
	5	0.00312	0.00305	1.60×10^{-05}	0.00306	0.00298	1.84×10^{-05}	0.00267	0.00261	1.05×10^{-06}
	6	0.00311	0.00304	1.74×10^{-05}	0.00306	0.00298	1.85×10^{-05}	0.00266	0.00260	9.60×10^{-07}
	2	0.00817	0.00793	1.90×10^{-04}	0.00837	0.00827	4.23×10^{-05}	0.00669	0.00652	1.16×10^{-04}
	3	0.00426	0.00421	5.33×10^{-06}	0.00462	0.00458	2.40×10^{-06}	0.00352	0.00347	4.90×10^{-06}
0.70	4	0.00426	0.00422	5.53×10^{-06}	0.00458	0.00454	2.75×10^{-06}	0.00353	0.00349	4.87×10^{-06}
	5	0.00426	0.00422	5.92×10^{-06}	0.00458	0.00453	3.18×10^{-06}	0.00353	0.00349	4.92×10^{-06}
	6	0.00425	0.00421	6.12×10^{-06}	0.00457	0.00452	3.10×10^{-06}	0.00355	0.00351	4.79×10^{-06}
	2	0.03361	0.03300	5.44×10^{-04}	0.03137	0.03053	7.81×10^{-04}	0.03046	0.02985	5.11×10^{-04}
	3	0.01167	0.01165	4.39×10^{-06}	0.01035	0.01032	1.72×10^{-05}	0.00957	0.00954	1.74×10^{-05}
0.85	4	0.01172	0.01170	4.01×10^{-06}	0.01034	0.01031	1.74×10^{-05}	0.00947	0.00943	1.95×10^{-05}
	5	0.01170	0.01169	3.99×10^{-06}	0.01025	0.01022	1.68×10^{-05}	0.00947	0.00943	2.26×10^{-05}
	6	0.01171	0.01169	3.98×10^{-06}	0.01026	0.01023	1.71×10^{-05}	0.00945	0.00942	1.91×10^{-05}

TABLE 4
Comparison of Scenarios ($\alpha = 0.5$). Estimation of (H, σ) with λ fixed.

		$\lambda = 1,\ \sigma = 1$			$\lambda = 1.1,\ \sigma = 1$			$\lambda = 1,\ \sigma = 0.9$		
H	L	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$
	2	0.00179	0.00121	8.34×10^{-07}	0.00189	0.00131	2.47×10^{-06}	0.00159	0.00106	5.11×10^{-07}
	3	0.00179	0.00122	1.12×10^{-06}	0.00186	0.00129	3.50×10^{-06}	0.00159	0.00106	6.27×10^{-07}
0.25	4	0.00180	0.00122	1.18×10^{-06}	0.00185	0.00128	3.79×10^{-06}	0.00158	0.00106	6.82×10^{-07}
	5	0.00178	0.00121	1.41×10^{-06}	0.00185	0.00128	3.72×10^{-06}	0.00158	0.00105	7.32×10^{-07}
	6	0.00179	0.00121	1.26×10^{-06}	0.00186	0.00128	3.53×10^{-06}	0.00158	0.00106	1.05×10^{-06}
	2	0.00169	0.00124	7.24×10^{-06}	0.00172	0.00125	2.14×10^{-06}	0.00152	0.00111	5.87×10^{-08}
	3	0.00171	0.00125	7.11×10^{-06}	0.00172	0.00125	1.99×10^{-06}	0.00153	0.00112	6.10×10^{-08}
0.40	4	0.00170	0.00124	7.28×10^{-06}	0.00172	0.00125	1.90×10^{-06}	0.00153	0.00112	7.78×10^{-08}
	5	0.00170	0.00124	7.72×10^{-06}	0.00171	0.00125	1.77×10^{-06}	0.00152	0.00111	1.34×10^{-07}
	6	0.00170	0.00124	7.88×10^{-06}	0.00170	0.00124	1.70×10^{-06}	0.00152	0.00111	1.77×10^{-07}
	2	0.00183	0.00150	1.02×10^{-05}	0.00179	0.00149	3.18×10^{-07}	0.00152	0.00123	1.95×10^{-07}
	3	0.00180	0.00149	1.07×10^{-05}	0.00175	0.00145	3.71×10^{-07}	0.00147	0.00119	1.60×10^{-07}
0.55	4	0.00180	0.00149	1.05×10^{-05}	0.00175	0.00145	3.68×10^{-07}	0.00146	0.00118	1.54×10^{-07}
	5	0.00179	0.00148	9.46×10^{-06}	0.00173	0.00143	4.20×10^{-07}	0.00146	0.00118	2.30×10^{-07}
	6	0.00179	0.00148	9.85×10^{-06}	0.00175	0.00145	3.33×10^{-07}	0.00146	0.00118	2.49×10^{-07}
	2	0.00305	0.00279	2.82×10^{-05}	0.00334	0.00313	6.76×10^{-06}	0.00288	0.00265	2.06×10^{-05}
	3	0.00270	0.00248	7.47×10^{-06}	0.00308	0.00288	4.75×10^{-06}	0.00250	0.00231	3.99×10^{-06}
0.70	4	0.00270	0.00248	6.60×10^{-06}	0.00308	0.00288	4.76×10^{-06}	0.00251	0.00231	4.01×10^{-06}
	5	0.00269	0.00247	6.13×10^{-06}	0.00308	0.00288	4.64×10^{-06}	0.00249	0.00229	3.51×10^{-06}
	6	0.00269	0.00247	6.17×10^{-06}	0.00306	0.00286	4.71×10^{-06}	0.00247	0.00227	3.06×10^{-06}
	2	0.01554	0.01513	3.02×10^{-04}	0.01501	0.01457	3.46×10^{-04}	0.01298	0.01233	5.28×10^{-04}
	3	0.01008	0.00999	2.00×10^{-05}	0.01027	0.01018	1.46×10^{-05}	0.00838	0.00826	4.72×10^{-05}
0.85	4	0.00998	0.00989	1.51×10^{-05}	0.01016	0.01007	1.43×10^{-05}	0.00830	0.00818	5.12×10^{-05}
	5	0.00974	0.00965	1.41×10^{-05}	0.00991	0.00983	1.25×10^{-05}	0.00806	0.00794	5.39×10^{-05}
	6	0.00974	0.00965	1.37×10^{-05}	0.00962	0.00954	1.24×10^{-05}	0.00798	0.00786	4.55×10^{-05}

TABLE 5
Comparison of Scenarios ($\alpha = 0.1$). Estimation of (H, λ, σ) with Daubechies wavelets.

		$\lambda = 1,\ \sigma = 1$			$\lambda = 1.1,\ \sigma = 1$			$\lambda = 1,\ \sigma = 0.9$		
H	L	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$
	3	0.045	0.044	0.000	0.053	0.052	0.000	0.045	0.044	0.000
	4	0.042	0.041	0.000	0.047	0.046	0.000	0.042	0.040	0.000
0.25	5	0.042	0.040	0.000	0.047	0.046	0.000	0.041	0.040	0.000
	6	0.041	0.040	0.000	0.047	0.045	0.000	0.040	0.039	0.000
	7	0.041	0.040	0.000	0.047	0.045	0.000	0.040	0.039	0.000
	3	0.061	0.059	0.000	0.077	0.074	0.001	0.064	0.061	0.001
	4	0.059	0.056	0.000	0.074	0.071	0.001	0.063	0.060	0.001
0.40	5	0.060	0.057	0.001	0.074	0.071	0.001	0.062	0.059	0.001
	6	0.059	0.057	0.001	0.074	0.071	0.001	0.063	0.059	0.001
	7	0.060	0.058	0.001	0.075	0.072	0.001	0.063	0.060	0.001
	3	0.071	0.064	0.004	0.087	0.081	0.003	0.068	0.063	0.003
	4	0.072	0.065	0.004	0.084	0.079	0.002	0.069	0.064	0.003
0.55	5	0.071	0.064	0.004	0.086	0.081	0.002	0.069	0.063	0.003
	6	0.070	0.064	0.003	0.086	0.081	0.002	0.069	0.063	0.003
	7	0.070	0.064	0.004	0.087	0.081	0.002	0.068	0.063	0.003
	3	0.136	0.125	0.007	0.174	0.157	0.012	0.137	0.123	0.010
	4	0.116	0.107	0.004	0.137	0.127	0.005	0.110	0.101	0.005
0.70	5	0.105	0.098	0.002	0.126	0.118	0.004	0.102	0.095	0.004
	6	0.101	0.094	0.002	0.123	0.115	0.003	0.097	0.091	0.003
	7	0.102	0.095	0.002	0.121	0.113	0.003	0.099	0.092	0.003
	3	0.268	0.240	0.003	0.284	0.247	0.001	0.235	0.202	0.007
	4	0.184	0.167	0.006	0.208	0.190	0.007	0.172	0.156	0.006
0.85	5	0.168	0.152	0.005	0.189	0.173	0.006	0.151	0.138	0.004
	6	0.157	0.142	0.004	0.175	0.160	0.005	0.142	0.128	0.005
	7	0.154	0.139	0.004	0.173	0.157	0.005	0.138	0.125	0.004

TABLE 6

Comparison of Scenarios ($\alpha = 0.5$). Estimation of (H, λ, σ) with Daubechies wavelets.

		$\lambda = 1,\ \sigma = 1$			$\lambda = 1.1,\ \sigma = 1$			$\lambda = 1,\ \sigma = 0.9$		
H	L	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$	$\widehat{MSE}$	$e(\widehat{Var})$	$\widehat{Bias}^2$
	3	0.217	0.203	0.012	0.216	0.213	0.002	0.212	0.202	0.009
	4	0.305	0.280	0.023	0.344	0.332	0.010	0.338	0.315	0.022
0.25	5	0.306	0.283	0.022	0.349	0.335	0.011	0.322	0.301	0.019
	6	0.299	0.276	0.021	0.360	0.345	0.013	0.318	0.297	0.020
	7	0.306	0.280	0.024	0.358	0.343	0.012	0.319	0.297	0.021
	3	0.126	0.124	0.000	0.129	0.125	0.002	0.120	0.118	0.000
	4	0.134	0.133	0.000	0.160	0.158	0.000	0.133	0.131	0.000
0.40	5	0.137	0.135	0.000	0.156	0.154	0.000	0.130	0.128	0.000
	6	0.134	0.132	0.000	0.159	0.157	0.000	0.128	0.126	0.000
	7	0.136	0.134	0.000	0.157	0.155	0.000	0.126	0.125	0.000
	3	0.197	0.169	0.027	0.249	0.212	0.047	0.191	0.169	0.028
	4	0.195	0.167	0.026	0.246	0.201	0.043	0.189	0.162	0.026
0.55	5	0.194	0.166	0.026	0.235	0.192	0.041	0.178	0.152	0.024
	6	0.194	0.165	0.028	0.224	0.184	0.038	0.174	0.148	0.025
	7	0.194	0.165	0.027	0.221	0.184	0.035	0.167	0.145	0.020
	3	0.202	0.167	0.031	0.232	0.194	0.035	0.192	0.161	0.029
	4	0.187	0.182	0.002	0.228	0.222	0.002	0.179	0.175	0.001
0.70	5	0.181	0.177	0.001	0.216	0.211	0.001	0.169	0.166	0.001
	6	0.176	0.172	0.001	0.212	0.208	0.001	0.166	0.163	0.001
	7	0.176	0.172	0.001	0.214	0.210	0.001	0.165	0.161	0.001
	3	0.202	0.188	0.004	0.212	0.195	0.007	0.175	0.161	0.006
	4	0.198	0.183	0.004	0.204	0.187	0.007	0.172	0.157	0.007
0.85	5	0.192	0.177	0.005	0.195	0.179	0.006	0.165	0.151	0.007
	6	0.194	0.180	0.004	0.195	0.179	0.006	0.165	0.150	0.007
	7	0.189	0.175	0.005	0.193	0.178	0.005	0.167	0.152	0.007

Appendix B: Proofs

Proof of Lemma 2.1

First recall that $E[\varphi(t)^2] = V(\theta_0)$, where stationarity allows the above expression does not depend on t, and the data's true parameter is θ_0. Note that, for any $\theta \in \Theta$ and using the ergodicity of X_t:

$$\|\hat{\mathbf{g}}_N(\theta) - \mathbf{g}_0(\theta)\| = \left\|\frac{1}{N - L + 1} \sum_{i=L}^{N} [\boldsymbol{\varphi}(t_i)^2 - \mathbf{V}(\theta)] - [\mathbf{V}(\theta_0) - \mathbf{V}(\theta)]\right\|$$

$$= \left\|\frac{1}{N - L + 1} \sum_{i=L}^{N} \boldsymbol{\varphi}(t_i)^2 - \mathbf{V}(\theta_0)\right\| \overset{a.s}{\to} 0$$

where $\boldsymbol{\varphi}(t)^2 := (\varphi_1(t)^2, \ldots, \varphi_L(t)^2)'$ and $\mathbf{V}(\theta) := (V_1(\theta), \ldots, V_L(\theta))'$. Then we have:

$$\sup_{\theta \in \Theta} \|\hat{\mathbf{g}}_N(\theta) - \mathbf{g}_0(\theta)\| \overset{a.s.}{\to} 0. \tag{B.1}$$

Furthermore, based on the proof of Theorem 2.6 in [36]:

$$|\hat{Q}_N(\theta) - Q_0(\theta)|$$
$$\le |(\hat{\mathbf{g}}_N(\theta) - \mathbf{g}_0(\theta))'A(\hat{\mathbf{g}}_N(\theta) - \mathbf{g}_0(\theta))| + 2|(\hat{\mathbf{g}}_N(\theta) - \mathbf{g}_0(\theta))'A\mathbf{g}_0(\theta)|$$
$$\le \|\hat{\mathbf{g}}_N(\theta) - \mathbf{g}_0(\theta)\| \cdot \|A\|_2 + 2\|\hat{\mathbf{g}}_N(\theta) - \mathbf{g}_0(\theta)\| \cdot \|A\| \cdot \|g_0(\theta)\|$$
$$\overset{a.s.}{\to} 0$$

for any $\theta \in \Theta$. Here we used the following facts: (1) the matrix A are deterministic, (2) $\mathbf{g}_0(\theta)$ is uniformly bounded (because $\rho_\theta(t)$ is continuous over the compact set Θ) and (3) equation (B.1). Finally if we take the supremum over all $\theta \in \Theta$, the lemma holds. $\qquad\square$

Proof of Lemma 2.2

Note that $Q_0(\theta)$ can be written as (for any $\theta \in \Theta$):

$$\begin{aligned}
Q_0(\theta) &= \mathbf{g}_0(\theta)'A\mathbf{g}_0(\theta) \\
&= [\mathbf{V}(\theta_0) - \mathbf{V}(\theta)]'A[\mathbf{V}(\theta_0) - \mathbf{V}(\theta)] \\
&= \begin{bmatrix} \mathbf{b}_1'(\boldsymbol{\rho}_{\theta_0}(\alpha k) - \boldsymbol{\rho}_\theta(\alpha k)) \\ \vdots \\ \mathbf{b}_L'(\boldsymbol{\rho}_{\theta_0}(\alpha k) - \boldsymbol{\rho}_\theta(\alpha k)) \end{bmatrix}' A \begin{bmatrix} \mathbf{b}_1'(\boldsymbol{\rho}_{\theta_0}(\alpha k) - \boldsymbol{\rho}_\theta(\alpha k)) \\ \vdots \\ \mathbf{b}_L'(\boldsymbol{\rho}_{\theta_0}(\alpha k) - \boldsymbol{\rho}_\theta(\alpha k)) \end{bmatrix} \\
&= (\boldsymbol{\rho}_{\theta_0}(\alpha) - \boldsymbol{\rho}_\theta(\alpha))'B'AB(\boldsymbol{\rho}_{\theta_0}(\alpha) - \boldsymbol{\rho}_\theta(\alpha))
\end{aligned}$$

Since $A > 0$ and B has column-rank between p and L (see assumption 2.2) we can diagonalize $B'AB$ to get:

$$Q_0(\theta) = \sum_{i=1}^{p} \lambda_i[\rho_{\theta_0}(\alpha \cdot (i-1)) - \rho_\theta(\alpha \cdot (i-1))]^2 + J(\theta)$$

where $\{\lambda_i\}_{i=1}^{p}$ are non-zero eigenvalues of $B'AB$ and $J(\theta)$ is a non-injective function of θ such that $J(\theta_0) = 0$. Because of Assumption 2.1 the lemma holds. $\qquad\square$

Proof of Lemma 2.3

(i) Let $1 \le j \le p$ and $\theta \in \Theta$. Then the j-th column of $\hat{G}_N(\theta)$ is:

$$\begin{aligned}
(\hat{G}_N(\theta))_j &= \frac{1}{N-L+1} \sum_{i=L}^{N} \frac{\partial}{\partial \theta_j}\left(\boldsymbol{\varphi}(t_i)^2 - \mathbf{V}(\theta)\right) \\
&= \frac{1}{N-L+1} \sum_{i=L}^{N} -\frac{\partial}{\partial \theta_j}\mathbf{V}(\theta) \\
&= \frac{1}{N-L+1} \sum_{i=L}^{N} \frac{\partial}{\partial \theta_j}\left(\mathbf{V}(\theta_0) - \mathbf{V}(\theta)\right) = -\frac{\partial}{\partial \theta_j}\mathbf{V}(\theta)
\end{aligned}$$

$$= E\left[\frac{\partial}{\partial \theta_j}\hat{\mathbf{g}}_N(\theta)\right]$$

$$= (G(\theta))_j.$$

and the continuity is inherited from the continuous differentiability of $\rho_\theta(\cdot)$.

(ii) Using (i), we have:

$$G(\theta) = -B\nabla_\theta \boldsymbol{\rho}_\theta(\alpha)$$

and this matrix is a full rank-matrix due to the remark after Assumption 2.1 and the fact that B is full-rank too. Since D is positive definite, the result holds. $\square$

Proof of Lemma 2.4

First note that by rescaling the vector $\hat{\mathbf{g}}_N(\theta_0)$ with the variance of each component of the vector $\mathbf{g}(t, \theta_0)$, we can write the sample moment equations in the following way:

$$\sqrt{N}V_D(\theta_0)^{-1}\hat{\mathbf{g}}_N(\theta_0) = \frac{\sqrt{N}}{N-L+1}\sum_{i=L}^{N}\left(\frac{\varphi_\ell(t_i)^2 - V_\ell(\theta_0)}{V_\ell(\theta_0)}\right)_{\ell\in\{1,\dots,L\}}$$

$$= \frac{\sqrt{N}}{N-L+1}\sum_{i=L}^{N}\left(Z_{\ell,t_i}^2 - 1\right)_{\ell\in\{1,\dots,L\}}$$

$$= \frac{\sqrt{N}}{N-L+1}\sum_{i=L}^{N}\left(H_2(Z_{\ell,t_i})\right)_{\ell\in\{1,\dots,L\}} \qquad (B.2)$$

where $V_D(\theta_0) := \mathrm{Diag}\,(V_\ell(\theta_0))_{\ell\in\{1,\dots,L\}}$, $H_2(x) = x^2 - 1$ is the second Hermite polynomial (see [30]) and

$$Z_{\ell,t_i} := \frac{\varphi_\ell(t_i)}{\sqrt{V_\ell(\theta_0)}}.$$

So the main idea of the proof is to use the vector-valued version of the Breuer-Major theorem with spectral-information conditions (see Theorem 3.1, [5]) to study the asymptotic behavior of (B.2). In order to achieve this, we need to analyze the spectral behavior of Z_{ℓ,t_i}. First recall that for $n, m \in \{0, \dots, N\}$ we have (see equation (2.7)):

$$\mathrm{Cov}(X_{t_m}, X_{t_n}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i|m-n|u}\bar{f}_{\theta_0}\left(\frac{u}{\alpha}\right)du.$$

where $\bar{f}_{\theta_0}(\cdot)$ is the spectral density of the discretized version of X_t at $t \in \{t_0, \dots, t_N\}$ (see equation (2.6)). Now, choose $\ell, \ell' \in \{1, \dots, L\}$ and note that

the covariance among Z_{ℓ,t_m} and Z_{ℓ',t_n} is: [1]

$$
\begin{aligned}
\beta(m&-n|\ell,\ell')\\
&:= \mathrm{Cov}(Z_{\ell,t_m}, Z_{\ell',t_n})\\
&= \frac{1}{\sqrt{V_\ell(\theta_0)V_{\ell'}(\theta_0)}} \sum_{k,k'=0}^{L} a_k^\ell a_{k'}^{\ell'} \mathrm{Cov}(X_{t_m-\alpha k}, X_{t_n-\alpha k'})\\
&= \frac{1}{\sqrt{V_\ell(\theta_0)V_{\ell'}(\theta_0)}} \sum_{k,k'=0}^{L} a_k^\ell a_{k'}^{\ell'} \cdot \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i[(m-n)-(k-k')]u} \bar{f}_{\theta_0}(u/\alpha) du\\
&= \int_{-\pi}^{\pi} e^{i(m-n)u} |P_{\mathbf{a}_\ell}(e^{-iu})| \cdot \overline{|P_{\mathbf{a}_{\ell'}}(e^{-iu})|} \frac{\bar{f}_{\theta_0}(u/\alpha)}{\sqrt{V_\ell(\theta_0)V_{\ell'}(\theta_0)}} du\\
&= \int_{-\pi}^{\pi} e^{i(m-n)u} \bar{h}_{\theta_0,\alpha}(u|\ell,\ell') du,
\end{aligned}
$$

where $P_{\mathbf{a}_\ell}(x) := \sum_{k=0}^{L} a_k^\ell x^k$ and

$$
\bar{h}_{\theta_0,\alpha}(u|\ell,\ell') := |P_{a_\ell}(e^{-iu})| \cdot \overline{|P_{a_{\ell'}}(e^{-iu})|} \frac{\bar{f}_{\theta_0}(u/\alpha)}{\sqrt{V_\ell(\theta_0)V_{\ell'}(\theta_0)}}.
$$

If we denote $\bar{h}_{\theta_0,\alpha}(\cdot) := \left(\bar{h}_{\theta_0,\alpha}(u|\ell,\ell')\right)_{\ell,\ell'\in\{1,...,L\}}$ the symmetric matrix of spectral densities among different filters, it is easy to conclude that this matrix does not depend on the sample size N, because α is fixed. Therefore, in order to apply Theorem 3.1 of [5], we only need to check that $\bar{h}_{\theta_0,\alpha}(\cdot) \in L^2((-\pi,\pi))$ uniformly. But this is true due to the fact that if l_ℓ and $l_{\ell'}$ are the corresponding filter orders for ℓ and ℓ' respectively, the polynomial $P_{\mathbf{a}_\ell}(x)$ has l_ℓ vanishing moments and:

$$
\begin{aligned}
\|\bar{h}_{\theta_0,\alpha}&(u|\ell,\ell')\|_{L^2((-\pi,\pi))}^2\\
&= \int_{-\pi}^{\pi} |P_{\mathbf{a}_\ell}(e^{-iu})|^2 \cdot \overline{|P_{\mathbf{a}_{\ell'}}(e^{-iu})|}^2 \frac{\bar{f}_{\theta_0}(u/\alpha)^2}{V_\ell(\theta_0)V_{\ell'}(\theta_0)} du\\
&\le \frac{1}{V_\ell(\theta_0)V_{\ell'}(\theta_0)} \int_{-\pi}^{\pi} \min\left(1, |u|^{2l_\ell+2l_{\ell'}}\right) \bar{f}_{\theta_0}(u/\alpha)^2 du < \infty
\end{aligned}
$$

by Assumption 2.3. Then:

$$
\sqrt{N} V_D(\theta_0)^{-1} \hat{\mathbf{g}}_N(\theta_0) \xrightarrow{d} N(0, U)
$$

where, by Plancherel theorem: $U_{ij} = 2\sum_{k\in\mathbb{Z}} \beta(k|i,j)^2 = 2\|\bar{h}_{\theta_0,\alpha}(u|i,j)\|_{L^2((-\pi,\pi))}^2$ and finally the lemma holds with:

$$
\Omega_{ij} = V_i(\theta_0) U_{ij} V_j(\theta_0) = 2 \sum_{k\in\mathbb{Z}} E[\varphi_i(t_{n+k})\varphi_j(t_n)]^2
$$

---

[1]Without loss of generality assume that $n, m > L$.

$$= 2 \sum_{k \in \mathbb{Z}} \left[\sum_{q,q'=0}^{L} a_q^i a_{q'}^j E \left[X_{\alpha(n+k)-\alpha q} X_{\alpha n - \alpha q'} \right] \right]^2$$

$$= 2 \sum_{k \in \mathbb{Z}} \left[\sum_{p=0}^{L} b_p^{i,j} \rho_{\theta_0} \left[\alpha(k+p) \right] \right]^2$$

for $i, j = 1 \ldots, L$. $\qquad \square$

Proof of Lemma 2.5

For any matrix norm $\| \cdot \|$ and $N \in \mathbb{N}$, we have that:

$$\|\psi_N\| = \|(G'(\theta_N) A G'(\theta_N))^{-1}\| \cdot \|G'\|_{op} \cdot \|A\|,$$

where $\| \cdot \|_{op}$ is the operator norm. Because of the remark after Assumption 2.1 and Lemma 2.3, the operator norm in the above equation is bounded for all N. Also note that if $F_N := G'(\theta_N) A G(\theta_N)$ and if $\lambda_{\max}(A)$ and $\lambda_{\min}(A)$ are the maximum and minimum eigenvalue of A:

$$\|F_N^{-1}\| \leq p \cdot \lambda_{\min}(F_N)^{-1} = p \cdot \left[\inf_{\|x\|=1} x' G'(\theta_N) A G(\theta_N) x \right]^{-1}$$

$$\leq p \left[\inf_{\|u\| \geq 1} u' A u \right]^{-1} = p \cdot \lambda_{\min}(A)^{-1},$$

where $u = G(\theta_N) x$. Moreover,

$$\|A\| \leq (L+1) \cdot \lambda_{\max}(A)$$

and hence there exists $R_{s,L} > 0$ such that $\|\psi_N\|^{4s} < R_{s,L}$ uniformly for all N. The lemma holds. $\qquad \square$

Proof of Lemma 3.1

Take $l, l' \in \{l_0, \ldots, l_L\}$, hence:

$$\int_{-\pi}^{\pi} R(u|l,l')^2 du = \int_{-\pi}^{\pi} \min\left(1, |u|^{2(l+l')}\right) \left[c_H \sum_{p \in \mathbb{Z}} \frac{\left|\frac{u+2\pi p}{\alpha}\right|^{1-2H}}{\left(\frac{u+2\pi p}{\alpha}\right)^2 + \lambda^2} \right]^2 du$$

$$= \alpha^{4H+2} c_H^2 \int_{1 \leq |u| \leq \pi} \left(\sum_{p \in \mathbb{Z}} \frac{|u+2\pi p|^{1-2H}}{(u+2\pi p)^2 + (\lambda\alpha)^2} \right)^2 du +$$

$$\alpha^{4H+2} c_H^2 \int_{|u| \leq 1} |u|^{2(l+l')} \left(\sum_{p \in \mathbb{Z}} \frac{|u+2\pi p|^{1-2H}}{(u+2\pi p)^2 + (\lambda\alpha)^2} \right)^2 du.$$

$$\text{(B.3)}$$

First note that for any $0 < H < 1$ and $p \neq 0$:

$$\sum_{p \neq 0, p \in \mathbb{Z}} \frac{|u + 2\pi p|^{1-2H}}{(u + 2\pi p)^2 + (\lambda \alpha)^2} \leq 2 \sum_{p=1}^{\infty} (u + 2\pi p)^{-1-2H} < \infty$$

and this implies that the two summands in equation (B.3) are finite for all p if $\pi < |u| < 1$. The same applies if $|u| < 1$ and $p \neq 0$. If $p = 0$ and $|u| < 1$ then the situation is slightly more complicated. Note that:

$$\int_{|u|<1} |u|^{2(l'+l)} \frac{|u|^{2-4H}}{[u^2 + (\lambda \alpha)^2]^2} du \leq \frac{2}{(\lambda \alpha)^4} \int_0^1 u^{2(l'+l)+2-4H} du < \infty$$

if and only if $2(l + l') > 4H - 3$. Then the lemma holds.　　□

Lemma B.1. *Assume that X_t is the stationary solution of the Ornstein-Uhlenbeck SDE. Then, for $t \geq 0$, the autocovariance function $\rho_\theta(t)$ of $\theta = (H, \lambda, \sigma)$ is[2]:*

$$\rho_\theta(t) = e^{-\lambda t} \, Var(X_0) \left[\frac{1 + e^{2\lambda t}}{2} - \lambda B_\theta(t) \right] \tag{B.4}$$

where:

$$Var(X_0) = \sigma^2 \lambda^{-2H} H \Gamma(2H),$$

$$B_\theta(t) = \int_0^t e^{2\lambda v} F_H(\lambda v) dv$$

$$F_H(x) = \frac{1}{\Gamma(2H-1)} \int_0^x e^{-s} s^{2H-2} ds.$$

Proof. Note that:

$$\rho_\theta(t) = E[X_t X_0] = \sigma^2 E \left[\int_{-\infty}^0 e^{\lambda u} dB_u^H \int_{-\infty}^t e^{-\lambda(t-v)} dB_v^H \right]$$

$$= e^{-\lambda t} \left[Var(X_0) + \sigma^2 E \left[\int_{-\infty}^0 e^{\lambda u} dB_u^H \int_0^t e^{-\lambda v} dB_v^H \right] \right].$$

Using the proof of Lemma 5.2 in [25] and Lemma 2.1 of [10]:

$$Var(X_0) = \sigma^2 \lambda^{-2H} H \Gamma(2H) \tag{B.5}$$

and

$$\sigma^2 E \left[\int_{-\infty}^0 e^{\lambda u} dB_u^H \int_0^t e^{-\lambda v} dB_v^H \right]$$

$$= \sigma^2 H(2H-1) \int_{-\infty}^0 \int_0^t e^{\lambda(u+v)} |u-v|^{2H-2} dv du$$

---

[2]Note that $F_H(\cdot)$ is the cdf of a $\Gamma(2H-1, 1)$ random variable when $H \geq 1/2$.

$$= \sigma^2 H(2H-1) \int_0^t \int_v^\infty e^{2\lambda v - \lambda x} x^{2H-2} dx dv$$

$$= \sigma^2 H \Gamma(2H) \lambda^{1-2H} \int_0^t e^{2\lambda v} dv - \sigma^2 H(2H-1) \int_0^t e^{2\lambda v} \int_0^v e^{-\lambda x} x^{2H-2} dx dv$$

$$= \mathrm{Var}(X_0) \left[\frac{e^{2\lambda t} - 1}{2} - \frac{\lambda}{\Gamma(2H-1)} \int_o^t e^{2\lambda v} \int_0^{\lambda v} e^{-s} s^{2H-2} ds dv \right]$$

and hence:

$$\rho_\theta(t) = e^{-\lambda t} \mathrm{Var}(X_0) \left[1 + \frac{e^{2\lambda t} - 1}{2} - \lambda B_\theta(t) \right]$$

and (B.4) holds. Note that [25] proves formula (B.5) in the case $H \geq \frac{1}{2}$, but using the analytical continuation of the gamma function its formula holds also for $0 < H < \frac{1}{2}$. $\square$

Lemma B.2. *Let X_t be the fOU process and $\theta = (H, \lambda, \sigma)$. Consider the mapping:*

$$\boldsymbol{\rho}_{\theta,3}(\alpha) = (\rho_\theta(\alpha \cdot 0), \rho_\theta(\alpha \cdot 1), \rho_\theta(\alpha \cdot 2))'.$$

in a closed rectangle $\Upsilon \subset \Theta$ where it holds that $\lambda < \exp(\Psi(3))$ and σ is of known sign. Then Assumption 2.1 is satisfied for α sufficiently small[3].

Proof. Following the proof of Lemma 5.2 in [25] we have that:

$$\rho_\theta(t) \approx \sigma^2 \lambda^{-2H} H \Gamma(2H) - \frac{\sigma^2 t^{2H}}{2}$$

for t sufficiently small. Define:

$$h_1(\theta) := \rho_\theta(0) = \frac{\sigma^2}{2} \lambda^{-2H} \Gamma(2H+1)$$

$$h_2(\theta) := \rho_\theta(\alpha) = \frac{\sigma^2}{2} \lambda^{-2H} \Gamma(2H+1) - \frac{\sigma^2}{2} \alpha^{2H}$$

$$h_3(\theta) := \rho_\theta(2\alpha) = \frac{\sigma^2}{2} \lambda^{-2H} \Gamma(2H+1) - \frac{\sigma^2}{2} (2\alpha)^{2H}.$$

It is not hard to verify that if $\lambda < \exp(\Psi(2H+1)) < \exp(\Psi(3))$ then $\frac{\partial h_1}{\partial H} > 0$. Also it is straightforward to compute that:

$$\left| \begin{matrix} \frac{\partial h_1}{\partial H} & \frac{\partial h_1}{\partial \lambda} \\ \frac{\partial h_2}{\partial H} & \frac{\partial h_2}{\partial \lambda} \end{matrix} \right| = -H\sigma^4 \Gamma(2H+1) \lambda^{-2H-1} \alpha^{2H} \log \alpha$$

and the latter is strictly positive iff $\alpha < 1$. Finally:

$$\left| \begin{matrix} \frac{\partial h_1}{\partial H} & \frac{\partial h_1}{\partial \lambda} & \frac{\partial h_1}{\partial \sigma} \\ \frac{\partial h_2}{\partial H} & \frac{\partial h_2}{\partial \lambda} & \frac{\partial h_2}{\partial \sigma} \\ \frac{\partial h_3}{\partial H} & \frac{\partial h_3}{\partial \lambda} & \frac{\partial h_3}{\partial \sigma} \end{matrix} \right| = -H\sigma^5 \Gamma(2H+1) \lambda^{-2H-1} 2^{2H} \alpha^{4H} \log 2.$$

---

[3]$\Psi(x) = \frac{d}{dx} \log \Gamma(x)$: digamma function. Note that $\exp(\Psi(3)) \approx 2.52$.

and this expression is positive if $\text{sign}(\sigma) < 0$. Using the Theorem 4 in [20] we can deduce that the mapping $\boldsymbol{\rho}_{\theta,3}(\alpha)$ is injective for α small. If the $\text{sign}(\sigma) > 0$ we can deduce the injectivity from the autocovariance function ρ_θ related to the process $-X$. Also note that the mapping satisfies the remark after assumption 2.1 as a direct consequence of the above calculations. $\square$

Appendix C: Behavior of $(\hat{\mathbf{g}}_N(\theta))_0$ when $H \geq 3/4$

The last result we present and prove in this paper, Proposition C.1 below, shows that if one does not perform any filtering of the observations of an fOU process, but one attempts to construct a Method-of-Moments estimator based directly on its the quadratic variations, for the purpose of estimating its variance, for instance, then the estimator cannot be asymptotically normal when $H > 3/4$, or even asymptotically of second-chaos type in law; this is the first item in the next proposition. However, the second item in this proposition shows that the estimator is strongly consistent nonetheless, with an L^2-rate of convergence of order N^{2H-2} (which is slower than $N^{-1/2}$). The standard Borel-Cantelli argument (such as the one employed to prove Lemma 2.5) then implies that the estimator is strongly consistent with rate $N^{-\gamma}$ for any $\gamma < 2 - 2H$. The proof of Proposition C.1 requires elements of the Malliavin calculus, which are also given below.

Proposition C.1. *Assume that X_t is a stationary fOU process with parameters $\theta = (H, \lambda, \sigma)$ and $H \geq \frac{3}{4}$. Then:*

(i) If $H = \frac{3}{4}$:

$$\sqrt{\frac{N}{\log N}}(\hat{\mathbf{g}}_N(\theta))_0 \xrightarrow{d} N(0, 2\alpha^{-1}c_\theta^2)$$

where $c_\theta = \sigma^2 \frac{H(2H-1)}{\lambda} = \frac{3\sigma^2}{8\lambda}$.

(ii) If $H > \frac{3}{4}$, $(\hat{\mathbf{g}}_N(\theta))_0$ does not converge to a normal law, or even a second-chaos law. However,

$$E\left[|N^{2-2H}(\hat{\mathbf{g}}_N(\theta))_0|^2\right] = O(1).$$

Proof. First note that since $SA_{i,0} = A_i$ for all $i \in \{0, \ldots, N\}$ (see equations (2.5) and (2.9)), the variance of $\hat{g}_N(\theta) := (\hat{\mathbf{g}}_N(\theta))_0$ is:

$$E[|\hat{g}_N(\theta)|^2] = \frac{2}{N^2} \sum_{i,j=1}^{N} \langle A_i \otimes A_i, A_j \otimes A_j \rangle_{\mathcal{H}}$$

$$= \frac{2}{N^2} \sum_{i,j=1}^{N} \langle A_i, A_j \rangle_{\mathcal{H}}^2 = \frac{2}{N^2} \sum_{i,j=1}^{N} [\rho_\theta(|t_i - t_j|)]^2$$

$$= \frac{2\rho_\theta(0)^2}{N} + \frac{2}{N^2} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} [\rho_{\theta_0}(|t_i - t_j|)]^2$$

$$= \frac{2}{N} \underbrace{\left[\rho_\theta(0)^2 + \rho_\theta(1)^2 \right]}_{\beta_\theta} + \frac{2}{N^2} \sum_{\substack{i,j=1 \\ |i-j| \geq 2}}^{N} [\rho_{\theta_0}(|t_i - t_j|)]^2$$

where $\mathcal{H} := L^2\left[\left(-\frac{\pi}{\alpha}, \frac{\pi}{\alpha} \right) \right]$ and its inner product is given by:

$$\langle A_j, A_{j'} \rangle_{\mathcal{H}} = \rho_\theta(|t_j - t_{j'}|) = E[X_{t_j} X_{t_{j'}}] = E[I_1(A_j(\cdot)) I_1(A_{j'}(\cdot))].$$

Using the expansion (3.3) we have:

$$E[|\hat{g}_N(\theta)|^2] = \frac{2\beta_\theta}{N} + \frac{2}{N^2} \sum_{\substack{i,j=1 \\ |i-j| \geq 2}}^{N} \left[c_\theta^2 |t_i - t_j|^{4H-4} + O(|t_i - t_j|^{4H-6}) \right] \qquad \text{(C.1)}$$

$$= \frac{2\beta_\theta}{N} + \frac{2}{N} \left[\alpha^{4H-4} c_\theta^2 \sum_{k=2}^{N} \left(\frac{N-k}{N} \right) k^{4H-4} + \alpha^{4H-6} \sum_{k=2}^{N} \left(\frac{N-k}{N} \right) O(k^{4H-6}) \right]$$

where $c_\theta := \sigma^2 \frac{H(2H-1)}{\lambda}$. Note that for $H = \frac{3}{4}$, as $N \to \infty$:

$$E\left[\frac{N}{\log N} |\hat{g}_N(\theta)|^2 \right] \longrightarrow 2\alpha^{-1} c_\theta^2$$

We can write (C.1) when $H > \frac{3}{4}$ as:

$$E[|\hat{g}_N(\theta)|^2] = \frac{2\beta_\theta}{N} + \frac{2}{N^2} \sum_{\substack{i,j=1 \\ |i-j| \geq 2}}^{N} \left[c_\theta^2 |\alpha(i-j)|^{4H-4} + O(|\alpha(i-j)|^{4H-6}) \right]$$

$$= \frac{2\beta_\theta}{N} + 2\alpha^{4H-4} c_\theta^2 N^{4H-4} \sum_{\substack{i,j=1 \\ |i-j| \geq 2}}^{N} \frac{|i-j|^{4H-4}}{N^{4H-4}} \frac{1}{N^2}$$

$$+ \frac{2}{N} \alpha^{4H-6} \sum_{k=2}^{N} \left(\frac{N-k}{N} \right) O(k^{4H-6})$$

therefore:

$$E[|N^{2-2H} \hat{g}_N(\theta)|^2] \longrightarrow 2\alpha^{4H-4} c_\theta^2 \iint_{[0,1]^2} |x-y|^{4H-4} dx\, dy = \frac{2\alpha^{4H-4} c_\theta^2}{(2H-1)(4H-3)}$$

Define:

$$\tilde{D}_N(\theta) := \begin{cases} \sqrt{\frac{N}{\tilde{c}_1 \log N}} & \text{if } H = \frac{3}{4} \\ \sqrt{\frac{N^{4-4H}}{\tilde{c}_2}} & \text{if } H > \frac{3}{4} \end{cases}$$

where $\tilde{c}_1 = 2\alpha^{-1} c_\theta^2$ and $\tilde{c}_2 = \frac{2\alpha^{4H-4} c_\theta^2}{(2H-1)(4H-3)}$ and denote $F_N := \tilde{D}_N(\theta) \hat{g}_N(\theta)$. For both cases we have that $E[|F_N|^2] \to 1$ and hence the main assumption of Theorem 4 in [39] is satisfied. Now we only need to prove that:

$$\|DF_N\|_{\mathcal{H}}^2 \xrightarrow{L^2(\Omega)} 2$$

in order to prove the asymptotic convergence of F_N towards a normal law. We can compute the Malliavin derivative of $\hat{g}_N(\theta)$ as follows:

$$D_r \hat{g}_N(\theta) = \frac{1}{N} \sum_{i=1}^{N} 2 X_{t_i} D_r X_{t_i} = \frac{1}{N} \sum_{i=1}^{N} I_1(A_i(\cdot)) A_i(r)$$

and hence it turns out that:

$$E[\|D\hat{g}_N(\theta)\|_{\mathcal{H}}^2] = \frac{4}{N^2} \sum_{i,j=1}^{N} E[I_1(A_i(\cdot)) I_1(A_j(\cdot))] \langle A_i, A_j \rangle_{\mathcal{H}}$$

$$= \frac{4}{N^2} \sum_{i,j=1}^{N} \langle A_i, A_j \rangle_{\mathcal{H}}^2 = 2E[\|\hat{g}_N(\theta)\|_{\mathcal{H}}^2]$$

and this implies that:

$$E[\|DF_N\|_{\mathcal{H}}^2] = 2E[\|F_N\|_{\mathcal{H}}^2] \longrightarrow 2.$$

Therefore, we only need to prove that:

$$\|DF_N\|_{\mathcal{H}}^2 - E[\|DF_N\|_{\mathcal{H}}^2] \xrightarrow{L^2(\Omega)} 0$$

First note that by using the multiplication rule of Weiner integrals:

$$\|DF_N\|_{\mathcal{H}}^2 = \frac{4}{N^2} \tilde{D}_N(\theta)^2 \sum_{i,j=1}^{N} I_1(A_i(\cdot)) I_1(A_j(\cdot)) \langle A_i, A_j \rangle_{\mathcal{H}}$$

$$= \frac{4}{N^2} \tilde{D}_N(\theta)^2 \sum_{i,j=1}^{N} [I_2(A_i(\cdot) \otimes A_j(\cdot)) + \langle A_i, A_j \rangle_{\mathcal{H}}] \langle A_i, A_j \rangle_{\mathcal{H}}$$

$$= \frac{4}{N^2} \tilde{D}_N(\theta)^2 \sum_{i,j=1}^{N} I_2(A_i(\cdot) \otimes A_j(\cdot)) + E[\|DF_N\|_{\mathcal{H}}^2]$$

and hence using the Lemma C.1 (i) and (ii) we have that:

$$E\left[\|DF_N\|_{\mathcal{H}}^2 - E[\|DF_N\|_{\mathcal{H}}^2]\right]$$

$$= \frac{16}{N^4} \tilde{D}_N(\theta)^4 \sum_{i,j,i',j'=1}^{N} E[I_2(A_i(\cdot) \otimes A_j(\cdot)) I_2(A_{i'}(\cdot) \otimes A_{j'}(\cdot))] \langle A_i, A_j \rangle_{\mathcal{H}} \langle A_{i'}, A_{j'} \rangle_{\mathcal{H}}$$

$$= \frac{32}{N^4} \tilde{D}_N(\theta)^4 \sum_{i,j,i',j'=1}^{N} \langle A_i \tilde{\otimes} A_j, A_{i'} \tilde{\otimes} A_{j'} \rangle_{\mathcal{H}^2} \langle A_i, A_j \rangle_{\mathcal{H}} \langle A_{i'}, A_{j'} \rangle_{\mathcal{H}}$$

and therefore, using Lemma C.2, we conclude that:

$$E\left[\|DF_N\|_{\mathcal{H}}^2 - E[\|DF_N\|_{\mathcal{H}}^2]\right] \longrightarrow 0$$

only if $H = \frac{3}{4}$. Then for $H = \frac{3}{4}$ we can apply the Theorem 4 in [39] and deduce that:

$$\sqrt{\frac{N}{\tilde{c}_1 \log N}} \hat{g}_N(\theta) \xrightarrow{d} N(0,1).$$

For $H > \frac{3}{4}$, Lemma C.2 allows us to conclude that the limit of

$$\sqrt{\frac{N^{4-4H}}{\tilde{c}_2}} \hat{g}_N(\theta)$$

does not belong to the first Weiner chaos, and hence the GMM estimator is not normally distributed. Define the following function:

$$\mathcal{I}_N(r,s) := \frac{N^{1-2H}}{\sqrt{\tilde{c}_2}} \sum_{j=1}^{N} (A_j \otimes A_j)(r,s) \tag{C.2}$$

for any $(r,s) \in \mathbb{R}^2$. By Lemma C.3, this function is not a Cauchy sequence in $\mathcal{H}^2$. Hence we cannot find a second-Wiener integral $I_2(g(\cdot,\cdot))$ such that:

$$E[|F_N - I_2(g(\cdot,\cdot))|^2] = 2\|\mathcal{I}_N(\cdot,\cdot) - g(\cdot,\cdot)\|_{\mathcal{H}^2}^2 \longrightarrow 0$$

when $H > \frac{3}{4}$ (see the proof of Theorem 2 in [43]). Then the limit in $L^2(\Omega)$ of $\hat{g}_N(\theta)$ is not even a second-chaos random variable. □

Lemma C.1. *If $A_k(u) := A_k(u|\theta)$ is defined as (2.5) for $u > 0$, then:*

(i) The tensor product $(A_k \otimes A_j)(u,v)$ is symmetric if and only if $k = j$.
(ii)

$$\langle A_{i_1} \otimes \cdots \otimes A_{i_n}, A_{j_1} \otimes \cdots \otimes A_{j_n} \rangle_{\mathcal{H}^n} = \prod_{k=1}^{n} \langle A_{i_k}, A_{j_k} \rangle_{\mathcal{H}}$$

for any $H \in (0,1)$.

Proof. (i) It is trivial to deduce that for $u, v > 0$:

$$(A_k \otimes A_j)(u,v) = \exp(it_k u)\bar{f}_{\theta,\alpha}(u)^{1/2} \exp(it_j v)\bar{f}_{\theta,\alpha}(v)^{1/2}$$
$$= \exp(it_k v)\bar{f}_{\theta,\alpha}(v)^{1/2} \exp(it_j u)\bar{f}_{\theta,\alpha}(u)^{1/2} = (A_i \otimes A_j)(v,u)$$

if and only if $\exp(it_k(u-v)) = \exp(it_j(u-v))$, which is equivalent to have that $k = j$.

(ii)

$$\langle A_{i_1} \otimes \cdots \otimes A_{i_n}, A_{j_1} \otimes \cdots \otimes A_{j_n} \rangle_{\mathcal{H}^n} =$$
$$= \int_{(-\pi/\alpha,\pi/\alpha)^n} A_{i_1}(s_1) \cdots A_{i_n}(s_n) \overline{A_{j_1}(s_1) \cdots A_{j_n}(s_n)} ds_1 \cdots ds_n$$
$$= \left[\int_{-\pi/\alpha}^{\pi/\alpha} A_{i_1}(s_1)\bar{A}_{j_1}(s_1)ds_1 \right] \cdots \left[\int_{-\pi/\alpha}^{\pi/\alpha} A_{i_n}(s_n)\bar{A}_{j_n}(s_n)ds_n \right]$$
$$= \prod_{k=1}^{n} \langle A_{i_k}, A_{j_k} \rangle_{\mathcal{H}} \qquad\qquad □$$

Lemma C.2. *Suppose that A_k is defined as (2.5). Define for $H \geq \frac{3}{4}$:*

$$\mathcal{P}_N := \frac{\tilde{D}_N(\theta)^4}{N^4} \sum_{i,j,i',j'=1}^{N} \langle A_i \tilde{\otimes} A_j, A_{i'} \tilde{\otimes} A_{j'} \rangle_{\mathcal{H}^2} \langle A_i, A_j \rangle_{\mathcal{H}} \langle A_{i'}, A_{j'} \rangle_{\mathcal{H}}$$

hence:

$$\mathcal{P}_N \longrightarrow \begin{cases} 0 & \text{if } H = \frac{3}{4} \\ k & \text{if } H > \frac{3}{4}. \end{cases}$$

for some $k \neq 0$.

Proof. First note that the inner product of the symmetric tensor products can be written as:

$$\langle A_i \tilde{\otimes} A_j, A_{i'} \tilde{\otimes} A_{j'} \rangle_{\mathcal{H}^2} = \frac{1}{4} \big[\langle A_i \otimes A_j, A_{i'} \otimes A_{j'} \rangle_{\mathcal{H}} + \langle A_i \otimes A_j, A_{j'} \otimes A_{i'} \rangle_{\mathcal{H}} +$$
$$\langle A_j \otimes A_i, A_{i'} \otimes A_{j'} \rangle_{\mathcal{H}} + \langle A_j \otimes A_i, A_{j'} \otimes A_{i'} \rangle_{\mathcal{H}} \big]$$
$$= \frac{1}{2} \langle A_i, A_{i'} \rangle_{\mathcal{H}} \langle A_j, A_{j'} \rangle_{\mathcal{H}} + \frac{1}{2} \langle A_i, A_{j'} \rangle_{\mathcal{H}} \langle A_j, A_{j'} \rangle_{\mathcal{H}}$$

then, by symmetry of the indices, we can write:

$$\mathcal{P}_N := \frac{\tilde{D}_N(\theta)^4}{N^4} \sum_{i,j,i',j'=1}^{N} \langle A_i, A_{i'} \rangle_{\mathcal{H}} \langle A_j, A_{j'} \rangle_{\mathcal{H}} \langle A_i, A_j \rangle_{\mathcal{H}} \langle A_{i'}, A_{j'} \rangle_{\mathcal{H}}$$

and we can study the limiting behavior of $\mathcal{P}_N$ through the folllowing cases:

Case I: $i = i' = j = j'$

In this case we analyzed the behavior of the diagonal, which is convergent for any $H \geq \frac{3}{4}$ as follows:

$$\mathcal{P}_N = \frac{\tilde{D}_N(\theta)^4}{N^4} N \rho_\theta(0)^4 = \begin{cases} \frac{\rho_\theta(0)^4}{2\tilde{c}_1^2 N \log N} & \text{if } H = \frac{3}{4} \\ \frac{\rho_\theta(0)^4}{\tilde{c}_2^2 N^{8H-5}} & \text{if } H > \frac{3}{4} \end{cases}$$
$$\longrightarrow 0.$$

Case II: $i = i'$, $j = j'$ and $i \neq j$

In this particular case we observe:

$$\mathcal{P}_N = \frac{\tilde{D}_N(\theta)^4}{N^4} \rho_\theta(0)^2 \sum_{\substack{i,j=1 \\ i \neq j}}^{N} \langle A_i, A_j \rangle_{\mathcal{H}}^2 = \frac{\tilde{D}_N(\theta)^4 \rho_\theta(0)^2}{N^4} \Big[2(N-1)\rho_\theta(1)^2$$
$$+ \sum_{|i-j| \geq 2} c_\theta^2 \alpha^{4H-4} |i-j|^{4H-4} + c_\theta \alpha^{4H-6} N \sum_{k=2}^{N} \left(\frac{N-k}{N} \right) O(k^{4H-6}) \Big]$$

Note that if $H = \frac{3}{4}$:

$$\mathcal{P}_N = \frac{\tilde{D}_N(\theta)^4 \rho_\theta(0)^2}{N^4}\left[2(N-1)\rho_\theta(1)^2 + c_\theta^2 \alpha^{4H-4}\sum_{k=2}^N \left(\frac{N-k}{N}\right)k^{4H-4}+\right.$$

$$\left. c_\theta \alpha^{4H-6}N\sum_{k=2}^N \left(\frac{N-k}{N}\right)O(k^{4H-6})\right]$$

$$\asymp \frac{1}{N\log N} + \frac{1}{N} \longrightarrow 0.$$

and if $H > \frac{3}{4}$:

$$\mathcal{P}_N = \frac{\tilde{D}_N(\theta)^4 \rho_\theta(0)^2}{N^4}\left[2(N-1)\rho_\theta(1)^2 + c_\theta^2 \alpha^{4H-4}\sum_{|i-j|\geq 2} \frac{|i-j|^{4H-4}}{N^{4H-4}}\frac{1}{N^2}+\right.$$

$$\left. c_\theta \alpha^{4H-6}N\sum_{k=2}^N \left(\frac{N-k}{N}\right)O(k^{4H-6})\right]$$

$$\asymp N^{5-8H} + N^{2-4H}\iint_{[0,1]^2} |x-y|^{4H-4}dxdy \longrightarrow 0$$

since $\iint_{[0,1]^2} |x-y|^{4H-4}dxdy < \infty$ if $H > \frac{3}{4}$.

Case III: $i = i'$, $j \neq j'$ and $i \neq j$

Without loss of generality, we can assume that $|i-j| \geq 2$ and $|j-j'|$. If $|i-j| = 1$ or $|j-j'| = 1$ then it can be proved that the rates of convergence are the same as Case I and II. We can write $\mathcal{P}_N$ as:

$$\mathcal{P}_N = \frac{\tilde{D}_N(\theta)^4}{N^4}\rho_\theta(0)\left[\overbrace{c_\theta^3\alpha^{6H-6}\sum_{i,j,j'=1}^N |i-j|^{2H-2}|i-j'|^{2H-2}|j-j'|^{2H-2}}^{\mathcal{A}_N}\right.$$

$$\left. + 3c_\theta^2\alpha^{6H-8}\underbrace{\sum_{i,j,j'=1}^N |i-j|^{2H-4}|i-j'|^{2H-2}|j-j'|^{2H-2}O(1)}_{\mathcal{B}_N}\right]$$

the first factor behaves as:

$$\mathcal{A}_N = c_\theta^3\alpha^{6H-3}N^{6H-3}\sum_{i,j,j'=1}^N \left|\frac{i-j}{N}\right|^{2H-2}\left|\frac{i-j'}{N}\right|^{2H-2}\left|\frac{j-j'}{N}\right|^{2H-2}\frac{1}{N^3}$$

$$\asymp N^{6H-3}\iiint_{[0,1]^3} |x-y|^{2H-2}|x-y'|^{2H-2}|y-y'|^{2H-2}dxdydy'$$

and the second:

$$\mathcal{B}_N = c_\theta^3\alpha^{6H-3}N^{6H-5}\sum_{i,j,j'=1}^N \left|\frac{i-j}{N}\right|^{2H-2}\left|\frac{i-j'}{N}\right|^{2H-4}\left|\frac{j-j'}{N}\right|^{2H-2}\frac{1}{N^3}$$

$$\leq c_\theta^3 \alpha^{6H-3} 2^{-2} N^{6H-3} \sum_{i,j,j'=1}^{N} \left|\frac{i-j}{N}\right|^{2H-2} \left|\frac{i-j'}{N}\right|^{2H-2} \left|\frac{j-j'}{N}\right|^{2H-2} \frac{1}{N^3}$$

$$\asymp N^{6H-3} \iiint_{[0,1]^3} |x-y|^{2H-2}|x-y'|^{2H-2}|y-y'|^{2H-2} dxdydy'$$

and the previous integral is finite for any $H > \frac{1}{2}$. Hence, $\mathcal{P}_N$ is bounded asymptotically, for $H = \frac{3}{4}$, by:

$$\frac{N^2 \cdot N^{6H-3}}{N^4 \log N} = \frac{1}{N^{5-6H} \log N} \longrightarrow 0$$

and if $H > \frac{3}{4}$, it is bounded by:

$$\frac{N^{8-8H} \cdot N^{6H-3}}{N^4} = N^{1-2H} \longrightarrow 0$$

Case IV: $i \neq i'$, $j \neq j'$ and $i \neq j$

As we did in the previous case, the rates of convergence when at least two indices are far apart by less than two units are the same as the ones calculated in Cases I-II-III. So let us assume that $|i - i'| \geq 2$, $|j - j'| \geq 2$ and $|i - j| \geq 2$.

$$\mathcal{P}_N = \frac{\tilde{D}_N(\theta)^4}{N^4} \overbrace{\left[c_\theta^4 \alpha^{8H-8} \sum_{i,j,i',j'=1}^{N} |i-j|^{2H-2}|i'-j'|^{2H-2}|i-j'|^{2H-2}|i'-j|^{2H-2} \right.}^{\mathcal{C}_N}$$

$$\underbrace{\left. + 4c_\theta^3 \alpha^{8H-10} \sum_{i,j,i',j'=1}^{N} |i-j|^{2H-2}|i'-j'|^{2H-2}|i-j'|^{2H-2}|i'-j|^{2H-4} \right]}_{\mathcal{D}_N}$$

and note that the first summand behaves asymptotically as:

$$\mathcal{C}_N = c_\theta^4 \alpha^{8H-8} N^{8H-4} \sum_{i,j,i',j'=1}^{N} \left|\frac{i-j}{N}\right|^{2H-2} \left|\frac{i'-j'}{N}\right|^{2H-2} \left|\frac{i-j'}{N}\right|^{2H-2}$$

$$\times \left|\frac{i'-j}{N}\right|^{2H-2} \frac{1}{N^4}$$

$$\asymp N^{8H-4} \int_{[0,1]^4} (x-y)^{2H-2}(x'-y)^{2H-2}$$

$$\times (y-y')^{2H-2}(x-x')^{2H-2} dxdydx'dy'$$

and the second summand can be asymptotically bounded by:

$$c_\theta^3 \alpha^{8H-10} N^{8H-4} \int_{[0,1]^4} (x-y)^{2H-2}(x'-y)^{2H-2}$$

$$\times (y-y')^{2H-2}(x-x')^{2H-2} dxdydx'dy'$$

and this quadruple integral is bounded for any $H > \frac{1}{2}$. Hence if $H = \frac{3}{4}$, $\mathcal{P}_N$ is asymptotically bounded by:

$$\frac{N^2}{N^4 \log N} N^{8H-4} = \frac{1}{\log N} \longrightarrow 0.$$

However, if $H > \frac{3}{4}$, the first summand of $\mathcal{P}_N$ is asymptotically equal to:

$$\frac{N^{8-8H} N^{8H-4}}{N^4} = 1$$

then the limit of $\mathcal{P}_N$ is not 0. Moreover, since $\mathcal{D}_N$ is asymptotically bounded by N^{8H-4}, there exists $k \neq 0$ such that:

$$\mathcal{P}_N \longrightarrow k. \qquad \square$$

Lemma C.3. *If $\mathcal{I}_N$ is defined as (C.2) and $H > \frac{3}{4}$, then it is not a Cauchy sequence in $\mathcal{H}^2$.*

Proof. We will prove the statement by contradiction. Assume that $\mathcal{I}_N$ is Cauchy. First of all, note that $\mathcal{I}_N$ is bounded in $\mathcal{H}^2 = L^2((-\pi, \pi)^2)$:

$$\|\mathcal{I}_N(\cdot, \cdot)\|_{\mathcal{H}^2}^2$$

$$= \frac{N^{2-4H}}{\tilde{c}_2} \sum_{i,j=1}^N \langle A_i \otimes A_i, A_j \otimes A_j \rangle_{\mathcal{H}^2} = \frac{N^{2-4H}}{\tilde{c}_2} \sum_{i,j=1}^N \langle A_i, A_j \rangle_{\mathcal{H}}^2$$

$$= \frac{N^{4-4H}}{\tilde{c}_2} \frac{1}{N^2} \left[N\rho_\theta(0) + 2(N-1)\rho_\theta(1) + \sum_{|i-j|\geq 2} [\rho_\theta(t_i - t_j)]^2 \right]$$

$$\leq M_1 N^{3-4H} + \frac{N^{2-4H}}{\tilde{c}_2} \sum_{|i-j|\geq 2} [c_\theta[\alpha(i-j)^{2H-2}] + O(\alpha(i-j)^{2H-4})]^2$$

$$= M_1 N^{3-4H} + \frac{1}{\tilde{c}_2} c_\theta^2 \alpha^{4H-4} \sum_{|i-j|\geq 2} \left| \frac{i-j}{N} \right|^{4H-4} \frac{1}{N^2}$$

$$+ \alpha^{4H-6} N^{3-4H} \sum_{k=2}^N \left(\frac{N-k}{N} \right) k^{4H-6}$$

$$< M_1 + \frac{1}{2} + \alpha^{4H-6} M_2 < \infty$$

where $M_1 := \frac{\rho_\theta(0) + 2\rho_\theta(1)}{\tilde{c}_2}$ and $M_2 := \sum_{k=1}^\infty k^{4H-6}$. This implies that $\mathcal{I}_N$ is dominated in $L^2((-\pi/\alpha, \pi/\alpha)^2)$.

Also note that the pointwise limit of $\mathcal{I}_N$ is (if $H > \frac{1}{2}$):

$$|\mathcal{I}_N| = \left| \frac{N^{1-2H}}{\sqrt{\tilde{c}_2}} \sum_{j=1}^N \exp(irt_j) \exp(ist_j) \bar{f}_{\theta,\alpha}(r) \bar{f}_{\theta,\alpha}(r) \right|$$

$$= \frac{N^{1-2H}}{\sqrt{\tilde{c}_2}} \bar{f}_{\theta,\alpha}(r)\bar{f}_{\theta,\alpha}(r) \left| \sum_{j=1}^{N} \exp(i\alpha rj)\exp(i\alpha sj) \right| \asymp N^{1-2H} \longrightarrow 0$$

since:

$$\left| \sum_{j=1}^{N} \exp(i\alpha rj)\exp(i\alpha sj) \right| < \left| \sum_{j=0}^{\infty} \exp(i\alpha rj)\exp(i\alpha sj) - 1 \right|$$
$$\leq \left| \frac{\exp[i\alpha(r+s)]}{1 - \exp[i\alpha(r+s)]} \right| < \infty.$$

Then we can apply the dominated convergence theorem to conclude that:

$$E\left[\left| \frac{N^{2-2H}}{\sqrt{\tilde{c}_2}} \hat{g}_N(\theta) \right|^2\right] = E\left[|I_1(\mathcal{I}_N(\cdot,\cdot))|^2\right] = 2\|\mathcal{I}_N(\cdot,\cdot)\|_{\mathcal{H}^2}^2 \longrightarrow 0$$

but this contradicts the fact that:

$$E\left[\left| \frac{N^{2-2H}}{\sqrt{\tilde{c}_2}} \hat{g}_N(\theta) \right|^2\right] \longrightarrow 1$$

which was already evidenced in the proof of Theorem 3.2. Hence, by contradiction, $\mathcal{I}_N$ is not Cauchy. $\qquad\square$

References

[1] Luis Barboza, Bo Li, Martin P. Tingley, and Frederi G. Viens, *Reconstructing Past Temperatures from Natural Proxies and Estimated Climate Forcings using Short- and Long-Memory Models*, The Annals of Applied Statistics **8** (2014), no. 4, 1966–2001. MR3292486

[2] J. M. Bardet, G. Lang, E. Moulines, and P. Soulier, *Wavelet Estimator of Long-Range Dependent Processes*, Statistical Inference for Stochastic Processes **3** (2000), no. 1–2, 85–99 (en).

[3] Jan Beran, *Statistics for long-memory processes*, Monographs on Statistics and Applied Probability, 61, Chapman & Hall, 1994.

[4] H. Biermé and F. Richard, *Estimation of anisotropic Gaussian fields through Radon transform*, ESAIM: Probability and Statistics **12** (2008), no. 1, 30–50. MR2367992

[5] Hermine Biermé, Aline Bonami, and José R. Leon, *Central Limit Theorems and Quadratic Variations in Terms of Spectral Density*, Electronic Journal of Probability **16** (2011), 362–395. MR2774094

[6] Péter Breuer and Péter Major, *Central limit theorems for non-linear functionals of Gaussian fields*, Journal of Multivariate Analysis **13** (1983), no. 3, 425–441. MR0716933

[7] Peter J. Brockwell and Richard A. Davis, *Time Series: Theory and Methods (Springer Series in Statistics)*, second ed., Springer, 2009. MR1093459

[8] Alexandre Brouste and Stefano M. Iacus, *Parameter estimation for the discretely observed fractional Ornstein–Uhlenbeck process and the Yuima R package*, Computational Statistics **28** (2012), no. 4, 1529–1547. MR3120827

[9] Richard H. Byrdt, Peihuang Lut, Jorge Nocedalt, Ciyou Zhu, R. H. Byrd, P. Lu, J. Nocedal, and Ciyou Zhu, *A limited memory algorithm for bound constrained optimization*, SIAM Journal on Scientific Computing **16** (1995), no. 5, 1190–1208.

[10] Patrick Cheridito, Hideyuki Kawaguchi, and Makoto Maejima, *Fractional Ornstein-Uhlenbeck processes*, Electronic Journal of Probability **8** (2003), no. 3, 1–14.

[11] A. Chronopoulou, F. G. Viens, and C. A. Tudor, *Variations and Hurst index estimation for a Rosenblatt process using longer filters*, Electronic Journal of Statistics **3** (2009), 1393–1435.

[12] Alexandra Chronopoulou, Ciprian A. Tudor, and Frederi G. Viens, *Application of Malliavin calculus to long-memory parameter estimation for non-Gaussian processes*, Comptes Rendus Mathematique **347** (2009), no. 11–12, 663–666.

[13] Alexandra Chronopoulou and Frederi G. Viens, *Stochastic Volatility and Option Pricing with Long-Memory in Discrete and Continuous Time*, Quantitative Finance **12** (2012), no. 4, 635–649. MR2909603

[14] Jean-Francois Coeurjolly, *Estimating the Parameters of a Fractional Brownian Motion by Discrete Variations of its Sample Paths*, Statistical Inference for Stochastic Processes (2001), no. 4, 199–227.

[15] Rainer Dahlhaus, *Efficient Parameter Estimation for Self-Similar Processes*, The Annals of Statistics **17** (1989), no. 4, 1749–1766.

[16] B. El Onsy, Khalifa Es-Sebaiy, and Frederi G. Viens, *Parameter Estimation for Ornstein-Uhlenbeck driven by fractional Ornstein-Uhlenbeck processes*, Preprint, in revision for Stochastics. (2014).

[17] Khalifa Es-Sebaiy and Frederi G. Viens, *Parameter estimation for SDEs related to stationary Gaussian processes*, Preprint (2015).

[18] P. Flandrin, *Wavelet analysis and synthesis of fractional Brownian motion*, IEEE Transactions on Information Theory **38** (1992), no. 2, 910–917.

[19] Robert Fox and Murad S. Taqqu, *Large-Sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series*, The Annals of Statistics **14** (1986), no. 2, 517–532 (EN).

[20] David Gale and Hukukane Nikaido, *The Jacobian matrix and global univalence of mappings*, Mathematische Annalen **159** (1965), no. 2, 81–93.

[21] Ulf Grenander, *Stochastic processes and statistical inference*, Arkiv för Matematik **1** (1950), no. 3, 195–277.

[22] Alastair R. Hall, *Generalized Method of Moments (Advanced Texts in Econometrics)*, Oxford University Press, USA, 2005.

[23] L. P. Hansen, J. Heaton, and A. Yaron, *Finite-sample properties of some alternative GMM estimators*, Journal of Business & Economic Statistics **14** (1996), no. 3, 262–280.

[24] Lars P. Hansen, *Large Sample Properties of Generalized Method of Moments Estimators*, Econometrica **50** (1982), no. 4, 1029–1054.

[25] Yaozhong Hu and David Nualart, *Parameter estimation for fractional Ornstein-Uhlenbeck processes*, Statistics & Probability Letters **80** (2010), no. 11–12, 1–16. MR2638974

[26] Yaozhong Hu and Jian Song, *Parameter Estimation for Fractional Ornstein-Uhlenbeck Processes with Discrete Observations*, Malliavin Calculus and Stochastic Analysis, Springer Proceedings in Mathematics & Statistics, Springer US, 2013, pp. 427–442.

[27] J. Istas and G. Lang, *Quadratic variations and estimation of the local Hölder index of a Gaussian process*, Annales de l'Institut Henri Poincare (B) Probability and Statistics **33** (1997), no. 4, 407–436.

[28] Svante Janson, *Gaussian Hilbert Spaces*, Cambridge Tracts in Mathematics, vol. 129, Cambridge University Press, 1997. MR1474726

[29] M. L. Kleptsyna and A. Le Breton, *Statistical Analysis of the Fractional Ornstein-Uhlenbeck Type Process*, Statistical Inference for Stochastic Processes **5** (2002), no. 3, 229–248.

[30] T. H. Koornwinder, R. Wong, R. Koekoek, and R. F. Swarttouw, *Orthogonal Polynomials*, NIST Handbook of Mathematical Functions (Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, and Charles W. Clark, eds.), Cambridge University Press, 2010.

[31] Guy Lebanon, *Probability: The Analysis of Data*, vol. 1, CreateSpace Independent Publishing Platform, 2012.

[32] Offer Lieberman, Roy Rosemarin, and Judith Rousseau, *Asymptotic Theory for Maximum Likelihood Estimation of the Memory Parameter in Stationary Gaussian Processes*, Econometric Theory **28** (2011), no. 02, 457–470 (English).

[33] Mikhail Lifshits, *Lectures on Gaussian Processes*, SpringerBriefs in Mathematics, Springer, 2012.

[34] Robert S. Liptser and Albert N. Shiryaev, *Statistics of Random Processes II*, second ed., Springer, 2000.

[35] E. Moulines, F. Roueff, and Murad S. Taqqu, *Central limit theorem for the log-regression wavelet estimation of the memory parameter in the gaussian semi-parametric context*, Fractals **15** (2007), no. 04, 301–313 (en).

[36] Whitney K. Newey and Daniel McFadden, *Large sample estimation and hypothesis testing*, Handbook of Econometrics (Robert F. Engle and Daniel L. McFadden, eds.), vol. 4, Elsevier, 1994, pp. 2111–2245.

[37] Jorge Nocedal and Stephen Wright, *Numerical Optimization*, second ed., Springer Series in Operations Research and Financial Engineering, Springer, 2006.

[38] Ivan Nourdin and Giovanni Peccati, *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality*, Cambridge Tracts in Mathematics, Cambridge University Press, 2012.

[39] D. Nualart and S. Ortiz-Latorre, *Central limit theorems for multiple stochastic integrals and Malliavin calculus*, Stochastic Processes and their Applications **118** (2008), no. 4, 614–628.

[40] David Nualart, *The Malliavin Calculus and Related Topics (Probability and Its Applications)*, second ed., Springer, 2010.

[41] Pham-Dinh-Tuan, *Estimation of parameters of a continuous time Gaussian stationary process with rational spectral density*, Biometrika **64** (1977), no. 2, 385–399.

[42] Mark J. Schervish, *Theory of Statistics*, Springer Series in Statistics, Springer, 1997. MR1354146

[43] Ciprian Tudor and Frederi Viens, *Variations of the fractional Brownian motion via Malliavin calculus*, To appear in Australian Journal of Mathematics (2008).

[44] Ciprian A. Tudor and Frederi G. Viens, *Statistical Aspects of the Fractional Stochastic Calculus*, The Annals of Statistics **35** (2007), no. 3, 1183–1212.

[45] Ciprian A. Tudor and Frederi G. Viens, *Variations and estimators for self-similarity parameters via Malliavin calculus*, Annals of Probability **37** (2009), no. 6, 2093–2134. MR2573552

[46] Weilin Xiao, Weiguo Zhang, and Weidong Xu, *Parameter estimation for fractional Ornstein–Uhlenbeck processes at discrete observation*, Applied Mathematical Modelling **35** (2011), no. 9, 4196–4207.

[47] Pu Zhang, Wei Lin Xiao, Xi Li Zhang, and Pan Qiang Niu, *Parameter identification for fractional Ornstein-Uhlenbeck processes based on discrete observation*, Economic Modelling **36** (2014), 198–203.