ISSN 1392-5113 Nonlinear Analysis: Modelling and Control, 2018, Vol. 23, No. 1, 120–140
https://doi.org/10.15388/NA.2018.1.9

Maximum likelihood estimation for Gaussian process with nonlinear drift

Yuliya Mishura, Kostiantyn Ralchenko, Sergiy Shklyar

Department of Probability Theory, Statistics and Actuarial Mathematics,
Taras Shevchenko National University of Kyiv,
Volodymyrska str. 64, 01601 Kyiv, Ukraine
myus@univ.kiev.ua; k.ralchenko@gmail.com; shklyar@univ.kiev.ua

Received: July 10, 2017 / **Revised:** November 11, 2017 / **Published online:** December 14, 2017

Abstract. We investigate the regression model $X_t = \theta G(t) + B_t$, where θ is an unknown parameter, G is a known nonrandom function, and B is a centered Gaussian process. We construct the maximum likelihood estimators of the drift parameter θ based on discrete and continuous observations of the process X and prove their strong consistency. The results obtained generalize the paper [13] in two directions: the drift may be nonlinear, and the noise may have nonstationary increments. As an example, the model with subfractional Brownian motion is considered.

Keywords: Gaussian process, discrete observations, continuous observations, maximum likelihood estimator, strong consistency.

1 Introduction

Let $B = \{B_t,\, t \geqslant 0\}$ be a centered Gaussian process with known covariance function, $B_0 = 0$. We assume that all finite-dimensional distributions of the process $\{B_t,\, t > 0\}$ are multivariate normal distributions with nonsingular covariance matrices. Now, let the process X_t have a drift $\theta G(t)$, that is,

$$X_t = \theta G(t) + B_t, \tag{1}$$

where $G(t) = \int_0^t g(s)\,\mathrm{d}s$, and $g \in L_1[0,t]$ for any $t > 0$.

The paper is devoted to the estimation of the parameter θ by observations of the process X. We construct the maximum likelihood estimators (MLEs) for discrete and continuous schemes of observations. We establish the strong consistency of both estimators. Moreover, we prove the a.s. convergence of the discrete estimator to the continuous one. This paper generalizes the results of [13], where model (1) with $G(t) = t$ was considered. Moreover, contrary to [13], we do not assume the stationarity of increments of the driving process B. This substantially extends the class of possible models. As an example, we consider the model where B is the subfractional Brownian motion.

Note that the problem of drift estimation for Gaussian processes is important for many applied areas, where an observed process can be decomposed as the sum of a useful signal and a random noise, which is usually modeled by a centered Gaussian process, see, e.g., [10, Chap. VII]. In particular, such processes arise in telecommunication and on financial markets. For example, Samuelson's model (see [18]), which is popular in finance, is of the form (1).

Mention also that similar problems for the model with linear drift driven by fractional Brownian motion were studied in [3,9,11,15]. The mixed Brownian-fractional Brownian model was treated in [7, 13]. Another approach to the drift parameter estimation in the model with two fractional Brownian motions was proposed in [12, 14]. In [2, 16], the nonparametric functional estimation of the drift of a Gaussian processes was considered (such estimators for fractional and subfractional Brownian motions were studied in [8] and [19], respectively).

The paper is organized as follows. In Section 2, we study the case of discrete observations and prove the strong consistency of MLE. In Section 3, we consider the estimator constructed by continuous observations and establish the relations between discrete and continuous estimators. Then we prove the strong consistency of the estimator in the continuous scheme. In Section 4, these results are applied to the models with fractional and subfractional Brownian motions. Auxiliary results for nonrandom functions and integral equations are collected in the Appendix.

2 The case of discrete-time observations

Let the process X_t be observed at the points $0 < t_1 < t_2 < \cdots < t_N$. Then the vector of increments $\Delta X^{(N)} = (X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_N} - X_{t_{N-1}})^\top$ is a one-to-one function of the observations. We assume in this section that the inequality $G(t_k) \neq 0$ holds at least for one k.

2.1 The likelihood function and MLE

Evidently, vector $\Delta X^{(N)}$ has Gaussian distribution $\mathcal{N}(\theta\Delta G^{(N)}, \Gamma^{(N)})$, where $\Delta G^{(N)} = (G(t_1), G(t_2) - G(t_1), \ldots, G(t_N) - G(t_{N-1}))^\top$. Let $\Gamma^{(N)}$ be the covariance matrix of the vector $\Delta B^{(N)} = (B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_N} - B_{t_{N-1}})^\top$. The density of the distribution of $\Delta X^{(N)}$ w.r.t. the Lebesgue measure is

$$\operatorname{pdf}_{\Delta X^{(N)}}(x) = \frac{(2\pi)^{-N/2}}{\sqrt{\det \Gamma^{(N)}}} \exp\left\{-\frac{1}{2}\big(x - \theta\Delta G^{(N)}\big)^\top \big(\Gamma^{(N)}\big)^{-1}\big(x - \theta\Delta G^{(N)}\big)\right\}.$$

Then one can take the density of the distribution of the vector $\Delta X^{(N)}$ for a given θ w.r.t. the density for $\theta = 0$ as a likelihood function:

$$L^{(N)}(\theta) = \exp\left\{\theta\big(\Delta G^{(N)}\big)^\top \big(\Gamma^{(N)}\big)^{-1}\Delta X^{(N)} - \frac{\theta^2}{2}\big(\Delta G^{(N)}\big)^\top \big(\Gamma^{(N)}\big)^{-1}\Delta G^{(N)}\right\}. \quad (2)$$

Then the corresponding MLE equals

$$\hat{\theta}^{(N)} = \frac{(\Delta G^{(N)})^\top (\Gamma^{(N)})^{-1} \Delta X^{(N)}}{(\Delta G^{(N)})^\top (\Gamma^{(N)})^{-1} \Delta G^{(N)}}. \tag{3}$$

Since the observed process X_t is Gaussian and the MLE $\hat{\theta}^{(N)}$ is a linear functional of the values of the process, we have that this estimator has a normal distribution. Taking into account that $\Delta X^{(N)} = \Delta B^{(N)} + \theta \Delta G^{(N)}$, the estimator $\hat{\theta}^{(N)}$ can be represented in the following form:

$$\hat{\theta}^{(N)} = \theta + \frac{(\Delta G^{(N)})^\top (\Gamma^{(N)})^{-1} \Delta B^{(N)}}{(\Delta G^{(N)})^\top (\Gamma^{(N)})^{-1} \Delta G^{(N)}}. \tag{4}$$

Hence, the estimator is unbiased and $\operatorname{var} \hat{\theta}^{(N)} = 1/((\Delta G^{(N)})^\top (\Gamma^{(N)})^{-1} \Delta G^{(N)})$.

2.2 The behaviour of the MLE for the increasing number of points

Let $N_1 \leqslant N_2$, and a set of points $\{t_1^{(1)}, \dots, t_{N_1}^{(1)}\}$ be a subset of $\{t_1^{(2)}, \dots, t_{N_2}^{(2)}\}$. Then there exists a matrix M, relating the increments w.r.t. these two sets of points: $\Delta X^{(N_1)} = M \Delta X^{(N_2)}$, $\Delta B^{(N_1)} = M \Delta B^{(N_2)}$, $\Delta G^{(N_1)} = M \Delta G^{(N_2)}$. Evidently, M is $N_1 \times N_2$-matrix, and it consists of zeros and ones. It has the following form:

$$M = \begin{pmatrix} 1 & \dots & 1 & 0 & \dots & 0 & \dots \\ 0 & \dots & 0 & 1 & \dots & 1 & \dots \\ \multicolumn{7}{c}{\dots\dots\dots\dots\dots\dots\dots\dots\dots} \end{pmatrix}.$$

If $t_{N_1}^{(1)} = t_{N_2}^{(2)}$, then each column of the matrix M contains exactly one 1. If $t_{N_1}^{(1)} = t_k^{(2)}$, $k < N_2$, then each of the first k columns of the matrix M contains exactly one 1, other $N_2 - k$ columns consist of zeros.

Lemma 1. *If $N_1 \leqslant N_2$ and the time-points of the process X_t used for the estimator $\hat{\theta}^{(N_1)}$ make a subset of the time-points used for the estimator $\theta^{(N_2)}$, then the increment $\hat{\theta}^{(N_2)} - \hat{\theta}^{(N_1)}$ is independent of the value $\hat{\theta}^{(N_2)}$.*

Proof. Let us compute the covariance between the estimators constructed by samples of different sizes:

$$\begin{aligned}
&\operatorname{cov}\bigl(\hat{\theta}^{(N_1)}, \hat{\theta}^{(N_2)}\bigr) \\
&\quad= \frac{\operatorname{cov}((\Delta G^{(N_1)})^\top (\Gamma^{(N_1)})^{-1} \Delta X^{(N_1)}, (\Delta G^{(N_2)})^\top (\Gamma^{(N_2)})^{-1} \Delta X^{(N_2)})}{(\Delta G^{(N_1)})^\top (\Gamma^{(N_1)})^{-1} \Delta G^{(N_1)} (\Delta G^{(N_2)})^\top (\Gamma^{(N_2)})^{-1} \Delta G^{(N_2)}} \\
&\quad= \frac{(\Delta G^{(N_1)})^\top (\Gamma^{(N_1)})^{-1} \mathbf{E} \Delta B^{(N_1)} (\Delta B^{(N_2)})^\top (\Gamma^{(N_2)})^{-1} \Delta G^{(N_2)}}{(\Delta G^{(N_1)})^\top (\Gamma^{(N_1)})^{-1} \Delta G^{(N_1)} (\Delta G^{(N_2)})^\top (\Gamma^{(N_2)})^{-1} \Delta G^{(N_2)}} \\
&\quad= \frac{(\Delta G^{(N_1)})^\top (\Gamma^{(N_1)})^{-1} \mathbf{E} M \Delta B^{(N_2)} (\Delta B^{(N_2)})^\top (\Gamma^{(N_2)})^{-1} \Delta G^{(N_2)}}{(\Delta G^{(N_1)})^\top (\Gamma^{(N_1)})^{-1} \Delta G^{(N_1)} (\Delta G^{(N_2)})^\top (\Gamma^{(N_2)})^{-1} \Delta G^{(N_2)}}
\end{aligned}$$

$$= \frac{(\Delta G^{(N_1)})^\top (\Gamma^{(N_1)})^{-1} M \Gamma^{(N_2)} (\Gamma^{(N_2)})^{-1} \Delta G^{(N_2)}}{(\Delta G^{(N_1)})^\top (\Gamma^{(N_1)})^{-1} \Delta G^{(N_1)} (\Delta G^{(N_2)})^\top (\Gamma^{(N_2)})^{-1} \Delta G^{(N_2)}}$$
$$= \frac{(\Delta G^{(N_1)})^\top (\Gamma^{(N_1)})^{-1} M \Delta G^{(N_2)}}{(\Delta G^{(N_1)})^\top (\Gamma^{(N_1)})^{-1} \Delta G^{(N_1)} (\Delta G^{(N_2)})^\top (\Gamma^{(N_2)})^{-1} \Delta G^{(N_2)}}$$
$$= \frac{(\Delta G^{(N_1)})^\top (\Gamma^{(N_1)})^{-1} \Delta G^{(N_1)}}{(\Delta G^{(N_1)})^\top (\Gamma^{(N_1)})^{-1} \Delta G^{(N_1)} (\Delta G^{(N_2)})^\top (\Gamma^{(N_2)})^{-1} \Delta G^{(N_2)}}$$
$$= \frac{1}{(\Delta G^{(N_2)})^\top (\Gamma^{(N_2)})^{-1} \Delta G^{(N_2)}} = \operatorname{var} \hat\theta^{(N_2)}.$$

Therefore, $\operatorname{cov}(\hat\theta^{(N_2)} - \hat\theta^{(N_1)}, \hat\theta^{(N_2)}) = \operatorname{var}(\hat\theta^{(N_2)}) - \operatorname{cov}(\hat\theta^{(N_1)}, \hat\theta^{(N_2)}) = 0$. The random variables $\hat\theta^{(N_1)} - \hat\theta^{(N_2)}$ and $\hat\theta^{(N_2)}$ are linear functionals of a Gaussian process. Therefore, their joint distribution is Gaussian. Hence, their uncorrelatedness implies independence. □

Corollary 1. *Under the assumptions of Lemma* 1*,* $\operatorname{var} \hat\theta^{(N_1)} \geqslant \operatorname{var} \hat\theta^{(N_2)}$.

Proof. By the independence of $\hat\theta^{(N_2)} - \hat\theta^{(N_1)}$ and $\hat\theta^{(N_2)}$, we have

$$\operatorname{var} \hat\theta^{(N_1)} = \operatorname{var} \hat\theta^{(N_2)} + \operatorname{var}\bigl(\hat\theta^{(N_2)} - \hat\theta^{(N_1)}\bigr) \geqslant \operatorname{var} \hat\theta^{(N_2)}.$$ □

2.3 Consistency of MLE

Theorem 1. *Let the following assumption hold*:

$$\frac{\operatorname{var} B_t}{(G(t))^2} \to 0 \quad \text{as } t \to \infty. \tag{5}$$

If $t_N \to \infty$ *as* $N \to \infty$*, then the discrete-time MLE* $\hat\theta^{(N)}$ *is* L_2*-consistent.*

Proof. The estimator is unbiased: $\mathbf{E}\hat\theta^{(N)} = \theta$. The estimator constructed by single observation X_{t_N},

$$\hat\theta^{(1)}_{t_N} = \frac{G(t_N)(\operatorname{var} B_{t_N})^{-1} X_{t_N}}{G(t_N)(\operatorname{var} B_{t_N})^{-1} G(t_N)} = \frac{X_{t_N}}{G(t_N)}, \tag{6}$$

has the variance

$$\operatorname{var} \hat\theta^{(1)}_{t_N} = \frac{\operatorname{var} B_{t_N}}{(G(t_N))^2}. \tag{7}$$

The estimator constructed by N observations has smaller variance according to Corollary 1. Therefore,

$$\mathbf{E}\bigl(\hat\theta^{(N)} - \theta\bigr)^2 = \operatorname{var} \hat\theta^{(N)} \leqslant \operatorname{var} \hat\theta^{(1)}_{t_N} = \frac{\operatorname{var} B_{t_N}}{(G(t_N))^2} \to 0.$$ □

The following statement follows from the proof of [13, Thm. 2.7]. (Note that it can be generalized: the mean-square convergence condition can be replaced with convergence in probability, see [5, Thm. 3.2.1].)

Lemma 2. *Assume that $\{\xi_i,\, i \geqslant 1\}$ is a sequence of random variables such that its elements $\xi_2, \xi_3, \dots$ (not including ξ_1) are mutually independent. If the series $\sum_{k=1}^{\infty} \xi_k$ converges in the mean square sense to a random variable ζ, that is,*

$$\lim_{n\to\infty} \mathbf{E}\left(\sum_{k=1}^{n} \xi_k - \zeta\right)^2 = 0,$$

then it converges a.s. to the same limit as well.

Theorem 2. *Under the assumptions of Theorem 1, the estimator $\hat\theta^{(N)}$ is strongly consistent.*

Proof. Let us show that the increments of the process $\{\hat\theta^{(N)},\, N \in \mathbb{N}\}$ are uncorrelated. For $2 \leqslant k_1 < k_2$,

$$\begin{aligned}
&\operatorname{cov}\bigl(\hat\theta^{(k_1)} - \hat\theta^{(k_1-1)},\, \hat\theta^{(k_2)} - \hat\theta^{(k_2-1)}\bigr)\\
&\quad = \operatorname{cov}\bigl(\hat\theta^{(k_2)} - \hat\theta^{(k_1-1)},\, \hat\theta^{(k_2)}\bigr) - \operatorname{cov}\bigl(\hat\theta^{(k_2)} - \hat\theta^{(k_1)},\, \hat\theta^{(k_2)}\bigr)\\
&\qquad + \operatorname{cov}\bigl(\hat\theta^{(k_2-1)} - \hat\theta^{(k_1)},\, \hat\theta^{(k_2-1)}\bigr) - \operatorname{cov}\bigl(\hat\theta^{(k_2-1)} - \hat\theta^{(k_1-1)},\, \hat\theta^{(k_2-1)}\bigr)\\
&\quad = 0
\end{aligned}$$

by Lemma 1. The estimators $\hat\theta^{(N)}$ and their increments $\hat\theta^{(k)} - \hat\theta^{(k-1)}$ are linear functionals of values of the Gaussian process X_t. Therefore, uncorrelatedness implies mutual independence. By Theorem 1, $\hat\theta^{(N)}$ converges to θ in mean square. Hence, by Lemma 2, $\hat\theta^{(N)}$ converges to θ a.s. □

3 The case of continuous-time observations

In this section, we suppose that the process X_t is observed on the whole interval $[0, T]$. We investigate MLE for the parameter θ based on these observations.

3.1 Assumptions on function G and process B

Evidently, B and X are Gaussian processes with the same covariance function, but, generally speaking, with different means since G is not zero identically. Our additional assumptions are:

(A) There exists a linear self-adjoint operator $\varGamma : L_2[0,T] \to L_2[0,T]$ such that

$$\operatorname{cov}(X_s, X_t) = \mathbf{E} B_s B_t = \int_0^t \varGamma\, \mathbf{1}_{[0,s]}(u)\, \mathrm{d}u = \langle \varGamma\, \mathbf{1}_{[0,s]},\, \mathbf{1}_{[0,t]} \rangle, \tag{8}$$

where $\langle f, g\rangle = \int_0^T f(t) g(t)\, \mathrm{d}t$.

(B) The drift function G is not zero identically, and in the representation $G(t) = \int_0^t g(s)\, \mathrm{d}s$, function $g \in L_2[0,T]$.

Note that, under assumption (A), the covariance between integrals of deterministic functions $f \in L_2[0,T]$ and $g \in L_2[0,T]$ w.r.t. the process B equals

$$\mathbf{E}\int_0^T f(s)\,\mathrm{d}B_s \int_0^T g(t)\,\mathrm{d}B_t = \langle \Gamma f, g\rangle.$$

3.2 Likelihood function

Now we establish the form of the likelihood function. In this order, introduce the notation

$$\mathcal{F}_{t_1,\dots,t_N} = \sigma(X_{t_1},\dots,X_{t_N}) = \sigma(B_{t_1},\dots,B_{t_N}),$$

the σ-field generated by the observations $X_{t_1},\dots,X_{t_N}$.

Theorem 3. *Let T be fixed, assumptions* (A), (B) *and additional assumption*

(C) *there exists a function $h_T \in L_2[0,T]$ such that $g = \Gamma h_T$*

hold. Then one can choose

$$L(\theta) = \exp\left\{\theta\int_0^T h_T(s)\,\mathrm{d}X_s - \frac{\theta^2}{2}\int_0^T g(s)h_T(s)\,\mathrm{d}s\right\} \tag{9}$$

as a likelihood function.

Proof. Let us show that the function $L(\theta)$ defined in (9) is a density function for the distribution of the process $\{X_t,\, t \in (0,T]\}$ for a given θ w.r.t. the distribution of the process $\{B_t,\, t \in (0,T]\}$, which coincides with $\{X_t,\, t \in (0,T]\}$ when $\theta = 0$. In other words, we need to prove that

$$\mathrm{d}P_\theta = L(\theta)\,\mathrm{d}P_0 \quad \text{for all } \theta \in \mathbb{R},$$

where P_θ is the probability measure that corresponds to the value of the parameter θ. For that reason, let $\vartheta \in \mathbb{R}$ be fixed and prove

$$\mathrm{d}P_\vartheta = L(\vartheta)\,\mathrm{d}P_0.$$

It is enough to verify that for all N, for all $t_1,\dots,t_N$, $0 < t_1 < \dots < t_N \leqslant T$, for all random events $A \in \mathcal{F}_{t_1,\dots,t_N}$,

$$\int_A \mathrm{d}P_\vartheta = \int_A L(\vartheta)\,\mathrm{d}P_0.$$

Under assumption (B), there exists $t_{\mathrm{nz}} \in (0,T]$ such that $G(t_{\mathrm{nz}}) \neq 0$. We can always assume that for at least one of the observations X_{t_k}, the inequality $G(t_k) \neq 0$ holds (otherwise, due to the fact that $A \in \mathcal{F}_{t_1,\dots,t_{\mathrm{nz}},\dots,t_N}$, we can insert t_{nz} into the set $t_1,\dots,t_N$

and repeat what follows). For such A,

$$\int_A \mathrm{d}P_\vartheta = \int_A L^{(N)}(\vartheta)\,\mathrm{d}P_0,$$

$$\int_A L(\vartheta)\,\mathrm{d}P_0 = \int_A \mathbf{E}_{\theta=0}\bigl[L(\vartheta) \bigm| X_{t_1},\dots,X_{t_N}\bigr]\,\mathrm{d}P_0,$$

where $L^{(N)}(\vartheta)$ is a likelihood function (2) for discrete-time model. Thus, it is enough to prove

$$\int_A L^{(N)}(\vartheta)\,\mathrm{d}P_0 = \int_A \mathbf{E}_{\theta=0}\bigl[L(\vartheta) \bigm| X_{t_1},\dots,X_{t_N}\bigr]\,\mathrm{d}P_0,$$

$$L^{(N)}(\vartheta) = \mathbf{E}_{\theta=0}\bigl[L(\vartheta) \bigm| X_{t_1},\dots,X_{t_N}\bigr]. \tag{10}$$

Let us evaluate

$$\mathbf{E}_{\theta=0}\bigl[L(\vartheta) \bigm| X_{t_1},\dots,X_{t_N}\bigr] = \mathbf{E}_{\theta=0}\bigl[L(\vartheta) \bigm| \Delta X^{(N)}\bigr] = \mathbf{E}_{\theta=0}\bigl[L(\vartheta) \bigm| \Delta B^{(N)}\bigr].$$

Note that $X_t = B_t$ if $\theta = 0$. The random vector $v = (\int_0^T h_T(t)\,\mathrm{d}B_t,\, (\Delta B^{(N)})^\top)^\top$ has multivariate Gaussian distribution because all its elements are linear functions in B_t. $\mathbf{E}v = 0$; evaluate its covariance matrix. The $(k+1,\,1)$ element of the matrix $\mathbf{E}vv^\top$ is equal to

$$\mathbf{E}\Biggl[\int_0^T h_T(t)\,\mathrm{d}B_t\,(B_{t_k} - B_{t_{k-1}})\Biggr]$$
$$= \langle \Gamma h_T, \mathbf{1}_{(t_k,t_{k-1})}\rangle = \langle g, \mathbf{1}_{(t_k,t_{k-1})}\rangle = G(t_k) - G(t_{k-1}),$$

which is the kth element of the vector $\Delta G^{(N)}$ (here $t_0 = 0$); thus, the lower-left block of the matrix $\mathbf{E}vv^\top$ is equal to $\mathbf{E}[\int_0^T h_T(t)\,\mathrm{d}B_t\,\Delta B^{(N)}] = \Delta G^{(N)}$. Other blocks are $\operatorname{var}[\int_0^T h_T(t)\,\mathrm{d}B_t] = \langle \Gamma h_T, h_T\rangle = \langle g, h_T\rangle$ and $\operatorname{var}(\Delta B^{(N)}) = \Gamma^{(N)}$. Thus, the covariance matrix of vector v is equal to

$$\mathbf{E}vv^\top = \begin{pmatrix} \langle g, h_T\rangle & (\Delta G^{(N)})^\top \\ \Delta G^{(N)} & \Gamma^{(N)} \end{pmatrix}. \tag{11}$$

By [1, Thm. 2.5.1], the conditional distribution of $\int_0^T h_T(t)\,\mathrm{d}B_t$ is Gaussian with

$$\mathbf{E}\Biggl[\int_0^T h_T(t)\,\mathrm{d}B_t \Bigm| \Delta B^{(N)}\Biggr] = \bigl(\Delta G^{(N)}\bigr)^\top \bigl(\Gamma^{(N)}\bigr)^{-1} \Delta B^{(N)},$$

$$\operatorname{var}\Biggl[\int_0^T h_T(t)\,\mathrm{d}B_t \Bigm| \Delta B^{(N)}\Biggr] = \langle g, h_T\rangle - \bigl(\Delta G^{(N)}\bigr)^\top \bigl(\Gamma^{(N)}\bigr)^{-1} \Delta G^{(N)}. \tag{12}$$

Finally,

$$\begin{aligned}
&\mathbf{E}_{\theta=0}\bigl[L(\vartheta) \bigm| \Delta B^{(N)}\bigr] \\
&\quad= \mathbf{E}\Biggl[\exp\Biggl\{\vartheta \int_0^T h_T(s)\,\mathrm{d}B_s - \frac{\vartheta^2}{2}\int_0^T g(s)h_T(s)\,\mathrm{d}s\Biggr\} \Bigm| \Delta B^{(N)}\Biggr] \\
&\quad= \exp\Biggl\{\bigl(\Delta G^{(N)}\bigr)^\top \bigl(\Gamma^{(N)}\bigr)^{-1}\Delta B^{(N)} \\
&\qquad+ \frac{\vartheta^2}{2}\bigl(\langle g, h_T\rangle - \bigl(\Delta G^{(N)}\bigr)^\top \bigl(\Gamma^{(N)}\bigr)^{-1}\Delta G^{(N)}\bigr) - \frac{\vartheta^2}{2}\langle g, h_T\rangle\Biggr\} \\
&\quad= L^{(N)}(\vartheta).
\end{aligned}$$

Thus, (10) is proved. From the fact that (10) holds true for all sets of $t_1, \dots, t_N$ and for all $\vartheta \in \mathbb{R}$ it follows that $L(\theta)$ is a likelihood function. □

3.3 MLE and its properties

The MLE maximizes the function $L(\theta)$. It equals

$$\hat{\theta}_T = \frac{\int_0^T h_T(s)\,\mathrm{d}X_s}{\int_0^T g(s)h_T(s)\,\mathrm{d}s}. \tag{13}$$

Since $X_t = B_t + \int_0^t g(s)\,\mathrm{d}s$, we have the following representation:

$$\hat{\theta}_T = \theta + \frac{\int_0^T h_T(s)\,\mathrm{d}B_s}{\int_0^T g(s)h_T(s)\,\mathrm{d}s}. \tag{14}$$

We see that the estimator is normally distributed and unbiased. Its variance equals

$$\operatorname{var}\hat{\theta}_T = \frac{\operatorname{var}[\int_0^T h_T(s)\,\mathrm{d}B_s]}{(\int_0^T g(s)h_T(s)\,\mathrm{d}s)^2} = \frac{\langle \Gamma h_T, h_T\rangle}{\langle g, h_T\rangle^2} = \frac{\langle g, h_T\rangle}{\langle g, h_T\rangle^2} = \frac{1}{\int_0^T g(s)h_T(s)\,\mathrm{d}s}.$$

Remark 1. Under the assumptions of Theorem 3, the denominator in the definition of the estimator (13) is positive. Indeed, by (12),

$$\begin{aligned}
\int_0^T g(s)h_T(s)\,\mathrm{d}s &= \operatorname{var}\Biggl[\int_0^T h_T(t)\,\mathrm{d}B^{(N)} \Bigm| \Delta B^{(N)}\Biggr] + \bigl(\Delta G^{(N)}\bigr)^\top \bigl(\Gamma^{(N)}\bigr)^{-1}\Delta G^{(N)} \\
&\geqslant \bigl(\Delta G^{(N)}\bigr)^\top \bigl(\Gamma^{(N)}\bigr)^{-1}\Delta G^{(N)} > 0.
\end{aligned}$$

The last inequality holds if $G(t_k) \neq 0$ at least for one k because, in this case, $\Delta G^{(N)} \neq 0$, and the matrix $(\Gamma^{(N)})^{-1}$ is positively definite.

3.4 Relations between discrete and continuous estimators

Let $0 = t_0 < t_1 < t_2 < \cdots < t_N \leqslant T$. Consider the discrete estimator $\hat{\theta}^{(N)}$ and the continuous estimator $\hat{\theta}_T$. Using (4) and (14), we can write

$$\begin{pmatrix} \hat{\theta}_T \\ \hat{\theta}^{(N)} \end{pmatrix} = \begin{pmatrix} \theta \\ \theta \end{pmatrix} + \begin{pmatrix} \frac{1}{\langle g, h_T \rangle} & 0 \\ 0 & ((\Delta G^{(N)})^\top (\Gamma^{(N)})^{-1} \Delta G^{(N)})^{-1} (\Delta G^{(N)})^\top (\Gamma^{(N)})^{-1} \end{pmatrix} v,$$

where the vector v is defined in the proof of the Theorem 3. By (11), we get

$$\mathbf{E} \begin{pmatrix} \hat{\theta}_T - \theta \\ \hat{\theta}^{(N)} - \theta \end{pmatrix} \begin{pmatrix} \hat{\theta}_T - \theta \\ \hat{\theta}^{(N)} - \theta \end{pmatrix}^\top = \begin{pmatrix} \frac{1}{\langle g, h_T \rangle} & \frac{1}{\langle g, h_T \rangle} \\ \frac{1}{\langle g, h_T \rangle} & \frac{1}{(\Delta G^{(N)})^\top (\Gamma^{(N)})^{-1} \Delta G^{(N)}} \end{pmatrix},$$

whence

$$\mathbf{E} \begin{pmatrix} \hat{\theta}_T - \theta \\ \hat{\theta}^{(N)} - \hat{\theta}_T \end{pmatrix} \begin{pmatrix} \hat{\theta}_T - \theta \\ \hat{\theta}^{(N)} - \hat{\theta}_T \end{pmatrix}^\top = \begin{pmatrix} \frac{1}{\langle g, h_T \rangle} & 0 \\ 0 & \frac{1}{(\Delta G^{(N)})^\top (\Gamma^{(N)})^{-1} \Delta G^{(N)}} - \frac{1}{\langle g, h_T \rangle} \end{pmatrix}.$$

The random variables $\hat{\theta}_T - \theta$ and $\hat{\theta}^{(N)} - \hat{\theta}_T$ are linear functions of the Gaussian process B. Therefore, they have bivariate normal distribution. Hence, their uncorrelatedness implies their independence. Consequently, $\hat{\theta}_T$ and $\hat{\theta}^{(N)} - \hat{\theta}_T$ are independent. Then

$$\operatorname{var} \hat{\theta}^{(N)} = \operatorname{var}\bigl(\hat{\theta}^{(N)} - \hat{\theta}_T\bigr) + \operatorname{var} \hat{\theta}_T \geqslant \operatorname{var} \hat{\theta}_T. \tag{15}$$

Note that the function $G(t)$ is continuous. Therefore, under assumptions of Theorem 3, there exists N_0 such that for all $N > N_0$, $G(k/N) \neq 0$ for some $1 \leqslant k \leqslant N$.

Theorem 4. *Let the assumptions of Theorem* 3 *hold. Construct the estimator* $\hat{\theta}^{(N)}$ *from* (3) *by observations* $X_{Tk/N}$, $k = 1, \ldots, N$. *Then* $\hat{\theta}^{(N)}$ *converges to* $\hat{\theta}_T$ *in mean square.*

Proof. By Lemma A.2, there exists a sequence of piecewise constant functions $g : [0, T] \to \mathbb{R}$ (constant on the intervals $((k-1)T/N,\, kT/N))$ such that $f_N \to h_T$ in $L_2[0, T]$, and $\int_0^T f_N(s) g(s)\, \mathrm{d}s = \int_0^T h_T(s) g(s)\, \mathrm{d}s$ for sufficiently large N. The function $f_N(t)$ can always be chosen in the form

$$f_N(t) = \sum_{k=1}^N a_{N,k}\, \mathbf{1}_{((k-1)T/N,\, kT/N]}(t).$$

Denote $a_N = (a_{N,1},\, a_{N,2},\, \ldots,\, a_{N,N})^\top$. Then

$$\langle \Gamma f_N, f_N \rangle \to \langle \Gamma h_T, h_T \rangle \quad \text{as } N \to \infty. \tag{16}$$

We have

$$\langle \Gamma f_N, f_N \rangle = \sum_{k=1}^N \sum_{l=1}^N a_k a_l \langle \Gamma\, \mathbf{1}_{((k-1)T/N,\, kT/N]}, \mathbf{1}_{((l-1)T/N,\, lT/N]} \rangle.$$

It follows from assumption (8) that

$$\begin{aligned}&\langle \Gamma\,\mathbf{1}_{((k-1)T/N,\,kT/N]}, \mathbf{1}_{((l-1)T/N,\,lT/N]}\rangle\\ &\quad= \mathbf{E}(B_{(k-1)T/N} - B_{kT/N})(B_{(l-1)T/N} - B_{lT/N})\end{aligned}$$

is (k,l)th element of the matrix $\Gamma^{(N)}$. Thus, $\langle \Gamma f_N, f_N\rangle = a_N^\top \Gamma^{(N)} a_N$. Recall also that $\langle \Gamma h_T, h_T\rangle = \langle g, h_T\rangle$, $\operatorname{var}\hat\theta_T = 1/\langle g, h_T\rangle$. Hence, convergence (16) can be written in the form

$$a_N^\top \Gamma^{(N)} a_N \to \langle g, h_T\rangle. \tag{17}$$

For sufficiently large N, we have

$$\begin{aligned}a_N^\top \Delta G^{(N)} &= \sum_{k=1}^N a_{N,k}\biggl(G\biggl(\frac{kT}{N}\biggr) - G\biggl(\frac{(k-1)T}{N}\biggr)\biggr)\\ &= \sum_{k=1}^N a_{N,k}\int_{(k-1)T/N}^{kT/N} g(t)\,\mathrm{d}t = \sum_{k=1}^N \int_{(k-1)T/N}^{kT/N} f_N(t)g(t)\,\mathrm{d}t\\ &= \int_0^T f_N(t)g(t)\,\mathrm{d}t = \int_0^T h_T(t)g(t)\,\mathrm{d}t = \langle g, h_T\rangle.\end{aligned}$$

Taking into account Lemma A.3 and convergence (17), we get

$$\operatorname{var}\hat\theta^{(N)} = \frac{1}{(\Delta G^{(N)})^\top \bigl(\Gamma^{(N)}\bigr)^{-1}\Delta G^{(N)}} \leqslant \frac{a_N^\top \Gamma^{(N)} a_N}{(a_N^\top \Delta G^{(N)})^2} \to \frac{1}{\langle g, h_T\rangle} = \operatorname{var}\hat\theta_T.$$

By (15),

$$\begin{aligned}\mathbf{E}\bigl(\hat\theta^{(N)} - \hat\theta_T\bigr)^2 &= \operatorname{var}\bigl(\hat\theta^{(N)} - \hat\theta_T\bigr) = \operatorname{var}\hat\theta^{(N)} - \operatorname{var}\hat\theta_T\\ &\leqslant \frac{a_N^\top \Gamma^{(N)} a_N}{(a_N^\top \Delta G^{(N)})^2} - \operatorname{var}\hat\theta_T \to 0 \quad \text{as } N\to\infty,\end{aligned}$$

whence the proof follows. □

Theorem 5. *Let the assumptions of Theorem* 3 *hold. The estimator* $\hat\theta^{(2^n)}$ *constructed by the observations* $X_{Tk/2^n}$, $k = 1,\dots,2^n$, *converges to* $\hat\theta_T$ *a.s.*

The proof repeats that of Theorem 2, where the reference to Theorem 1 is replaced by the reference to Theorem 4.

3.5 Consistency of the estimator

Theorem 6. *Assume that for all* $T > 0$, *there exists a function* $h_T \in L_2[0,T]$ *such that* $g|_{[0,T]} = \Gamma_T h_T$ *(*Γ_T *denotes the dependence of the operator* Γ *on* T*). If*

$$\liminf_{t\to\infty} \frac{\operatorname{var} B_t}{G(t)^2} = 0, \tag{18}$$

then the estimator $\hat\theta_T$ is consistent in mean square, that is,

$$\mathbf{E}(\hat\theta_T-\theta)^2\to 0 \quad \textit{as } T\to\infty.$$

Proof. By (18), there exists an increasing sequence of positive numbers $\{t_k,\ k\in\mathbb{N}\}$ such that $\lim_{k\to\infty} t_k=+\infty$, for all k, the inequality $G(t_k)\neq 0$ holds, and

$$\lim_{k\to\infty}\frac{\operatorname{var} B_{t_k}}{G(t_k)^2}=0.$$

Denote by $t(T)$ the largest t_k that does not exceed T. Then

$$\lim_{T\to+\infty}\frac{\operatorname{var} B_{t(T)}}{G(t(T))^2}=0.$$

The estimator $\hat\theta_T$ is unbiased. Compare its variance with the variance of the estimator $\hat\theta^{(1)}_{t(T)}$ constructed by single observation $X_{t(T)}$ (see (6) and (7) for the estimator and its variance). According to inequality (15),

$$\mathbf{E}(\hat\theta_T-\theta)=\operatorname{var}\hat\theta_T\leqslant \operatorname{var}\hat\theta^{(1)}_{t(T)}=\frac{\operatorname{var} B_{t(T)}}{G(T(t))^2}\to 0 \quad \text{as } T\to+\infty. \qquad \square$$

Lemma 3. *The stochastic process $\hat\theta_T$ (defined for all T such that $\int_0^T h_T(s)\,\mathrm{d}s\neq 0$) is a process with independent increments.*

Proof. Let us calculate the covariance between estimators with $T_1<T_3$.

$$\begin{aligned}
&\operatorname{cov}(\hat\theta_{T_1},\hat\theta_{T_3})\\
&\quad=\frac{\mathbf{E}\int_0^{T_1} h_{T_1}(s)\,\mathrm{d}B_s\int_0^{T_3} h_{T_3}(t)\,\mathrm{d}B_t}{\int_0^{T_1} g(s)h_{T_1}(s)\,\mathrm{d}s\int_0^{T_3} g(t)h_{T_3}(t)\,\mathrm{d}t}=\frac{\int_0^{T_1}\varGamma_{T_3}h_{T_3}(u)\,h_{T_1}(u)\,\mathrm{d}u}{\int_0^{T_1} g(s)h_{T_1}(s)\,\mathrm{d}s\int_0^{T_3} g(t)h_{T_3}(t)\,\mathrm{d}t}\\
&\quad=\frac{\int_0^{T_1} g(u)h_{T_1}(u)\,\mathrm{d}u}{\int_0^{T_1} g(s)h_{T_1}(s)\,\mathrm{d}s\int_0^{T_3} g(t)h_{T_3}(t)\,\mathrm{d}t}=\frac{1}{\int_0^{T_3} g(t)h_{T_3}(t)\,\mathrm{d}t}=\operatorname{var}\hat\theta_{T_3}
\end{aligned}$$

Therefore, for $0<T_1<T_2\leqslant T_3<T_4$,

$$\begin{aligned}
&\operatorname{cov}(\hat\theta_{T_2}-\hat\theta_{T_1},\,\hat\theta_{T_4}-\hat\theta_{T_3})\\
&\quad=\operatorname{cov}(\hat\theta_{T_2},\hat\theta_{T_4})-\operatorname{cov}(\hat\theta_{T_2},\hat\theta_{T_3})-\operatorname{cov}(\hat\theta_{T_1},\hat\theta_{T_4})+\operatorname{cov}(\hat\theta_{T_1},\hat\theta_{T_3})\\
&\quad=\operatorname{var}\hat\theta_{T_4}-\operatorname{var}\hat\theta_{T_3}-\operatorname{var}\hat\theta_{T_4}+\operatorname{var}\hat\theta_{T_3}=0.
\end{aligned}$$

The values of $\hat\theta_T$ are linear functions of Gaussian process X_t. Hence they have joint Gaussian distribution, and uncorrelatedness of the increments $\hat\theta_{T_2}-\hat\theta_{T_1}$ and $\hat\theta_{T_4}-\hat\theta_{T_3}$ implies their independence. $\square$

Theorem 7. *Under the assumptions of Theorem 6, the estimator $\hat\theta_T$ is strongly consistent.*

The proof repeats the proof of [13, Thm. 3.9].

4 Examples

4.1 Model with subfractional Brownian motion

The subfractional Brownian motion $\widetilde{B}^H$ with Hurst parameter $H \in (0,1)$ is a centered Gaussian random process with covariance function

$$\operatorname{cov}\big(\widetilde{B}^H_s, \widetilde{B}^H_t\big) = \frac{2\,|t|^{2H} + 2\,|s|^{2H} - |t-s|^{2H} - |t+s|^{2H}}{2}, \tag{19}$$

see [4, 20] for its properties. Obviously, neither $\widetilde{B}^H$, nor its increments are stationary. If $\{B^H_t,\ t \in \mathbb{R}\}$ is a fractional Brownian motion, i.e., a centered Gaussian process with covariance function $\operatorname{cov}(B^H_s, B^H_t) = (|t|^{2H} + |s|^{2H} - |t-s|^{2H})/2$, then the random process $(B^H_t + B^H_{-t})/\sqrt{2}$ is a subfractional Brownian motion. Evidently, mixed derivative of the covariance function (19) equals

$$K_H(s,t) := \frac{\partial^2 \operatorname{cov}(\widetilde{B}^H_s, \widetilde{B}^H_t)}{\partial t \partial s} = H(2H-1)\big(|t-s|^{2H-2} - |t+s|^{2H-2}\big).$$

If $H \in (1/2, 1)$, then the operator $\Gamma = \Gamma^H_T$ that satisfies (8) for $\widetilde{B}^H$ equals

$$\Gamma^H_T f(t) = \int_0^T K_H(s,t) f(s)\,\mathrm{d}s.$$

Consider model (1) for $G(t) = t$ and $B = \widetilde{B}^H$:

$$X_t = \theta t + \widetilde{B}^H_t. \tag{20}$$

Let us construct the estimators $\hat{\theta}^{(N)}$ and $\hat{\theta}_T$ from (3) and (13), respectively, and establish their properties. In particular, Proposition 1 allows to define finite-sample estimator $\hat{\theta}^{(N)}$.

Proposition 1. *The linear equation $\Gamma^H_T f = 0$ has only trivial solution in $L_2[0,T]$. As a consequence, the finite slice $(\widetilde{B}^H_{t_1}, \dots, \widetilde{B}^H_{t_N})$ with $0 < t_1 < \cdots < t_N$ has a multivariate normal distribution with nonsingular covariance matrix.*

Proof. Let $f \in L_2[0,T]$ be a solution to equation $\Gamma^H_T f = 0$. Then $f_1 \in L_2[0,2T]$,

$$\int_0^{2T} \frac{f_1(s)\,\mathrm{d}s}{|t-s|^{2-2H}} = 0 \quad \text{for almost all } t \in (0, 2T),$$

where $f_1(s) = -f(T-s)$ for $0 \leqslant s < T$, and $f_1(s) = f(s-T)$ for $T \leqslant s \leqslant 2T$. Then both $y = f_1 \in L_1[0,2T]$ and $y = 0 \in L_1[0,2T]$ are solutions to the equation

$$\int_0^{2T} \frac{y(s)\,\mathrm{d}s}{|t-s|^{2-2H}} = 0 \quad \text{for almost all } t \in (0, 2T).$$

By statement (ii) of Lemma A.4, $f_1(s) = 0$ almost everywhere on $(0, 2T)$, whence $f(s) = 0$ almost everywhere on $(0, T)$. Thus, the operator Γ_T^H is self-adjoint, compact, and positive definite. It admits the spectral representation $\Gamma_T^H f = \sum_{i=1}^{\infty} \lambda_i \langle f, e_i\rangle e_i$ with $\{e_i,\, i \in \mathbb{N}\}$ an orthonormal basis in $L_2[0,T]$ and $\lambda_i > 0$ for all i. Let $0 = t_0 < t_1 < \cdots < t_N \leqslant T$ be a partition of $[0,T]$. For the matrix $\Gamma_H^{(N)}$ and vector $v = (v_1, \dots, v_N)^\top \in \mathbb{R}^N$,

$$v^\top \Gamma_H^{(N)} v = \langle \Gamma_T^H f_v, f_v \rangle = \sum_{i=1}^{\infty} \lambda_i \langle f_v, e_i\rangle^2 \geqslant 0, \tag{21}$$

where f_v is piecewise constant function, $f_v(s) = v_k$ for $t_{k-1} < s < t_k$, and $f_v(s) = 0$ for $t_N < s < T$ (if $t_N < T$). The equality in (21) is attained if only $\langle f_v, e_i\rangle = 0$ for all i. That is only possible if $f_v = 0$ almost everywhere on $[0,T]$ and $v = 0$. □

Corollary 2. *Since* $\operatorname{var}(\widetilde{B}_t^H) = (2 - 2^{2H-1})t^{2H}$, *the random process* $\widetilde{B}^H$ *satisfies Theorems* 1 *and* 2. *Hence, under condition* $t_N \to +\infty$ *as* $N \to \infty$, *the estimator* $\hat\theta^{(N)}$ *is* L_2*-consistent and strongly consistent.*

In order to define the maximum likelihood estimator (13), we have to solve an integral equation. The following statement guarantees the existence of the solution:

Proposition 2. *If* $1/2 < H < 3/4$, *then the integral equation* $\Gamma_T^H h = \mathbf{1}_{[0,T]}$, *that is,*

$$\int_0^T K_H(s,t) h(s)\,\mathrm{d}s = 1 \quad \textit{for almost all } t \in (0,T) \tag{22}$$

has a unique solution $h \in L_2[0,T]$.

Proof. By Lemma A.5, the integral equation

$$\int_0^{2T} \frac{y(s)\,\mathrm{d}s}{|t-s|^{2-2H}} = \mathbf{1}_{[T,\,2T]}(t) \quad \text{for almost all } t \in (0,\,2T)$$

has a solution $y \in L_2[0,2T]$. Then

$$\int_0^T \big(y(T+s) - y(T-s)\big)\left(\frac{1}{|t-s|^{2-2H}} - \frac{1}{|t+s|^{2-2H}}\right)\mathrm{d}s = 1$$

for almost all $t \in (0,T)$, and $h(s) = (y(T+s) - y(T-s))/(H(2H-1))$ is a solution to integral equation (22). Note that $h \in L_2[0,T]$, and this finishes the proof. □

Corollary 3. *If* $1/2 < H < 3/4$, *then the random process* $\widetilde{B}^H$ *satisfies Theorems* 3–7. *As the result,* $L(\theta)$ *defined in* (9) *is the likelihood function in model* (20), *and* $\hat\theta_T$ *defined in* (13) *is the maximum likelihood estimator. The estimator is* L_2*-consistent and strongly consistent.*

4.2 Model with fractional Brownian motion and power drift

Let $1/2 < H < 1$ and $\alpha > -1/2$. Consider the process

$$X_t = \theta t^{\alpha+1} + B_t^H, \tag{23}$$

where X_t is a stochastic process observed on interval $[0, T]$, B_t^H is an unobserved fractional Brownian motion with Hurst index H, and θ is a parameter of interest. This is a particular case of model (1) with $g(t) = (\alpha+1)t^\alpha$.

Now verify the conditions of the theorems. Due to Proposition 1, any finite slice of the stochastic process $\{B_t^H,\ t > 0\}$ has a multivariate normal distribution with nonsingular covariance matrix; the process satisfies condition (A) with injective operator Γ. Condition (5) holds true if and only if $\alpha > H - 1$.

Corollary 4. *If* $\alpha > H - 1$, *model* (23) *satisfies the conditions of Theorems* 1 *and* 2. *The estimator* $\hat\theta^{(N)}$ *is* L_2*-consistent and strongly consistent* (*provided that* $\lim_{N\to\infty} t_N = +\infty$).

Condition (B): $g \in L_2[0, T]$ holds true if and only if $\alpha > -1/2$. The integral equation $\Gamma h = g$ is rewritten as

$$\int_0^T \frac{h(s)\,\mathrm{d}s}{|t-s|^{2H-2}} = \frac{\alpha+1}{(2H-1)H} t^\alpha.$$

If $\alpha > 2H - 2$, then the solution is

$$h(t) = \mathrm{const}\cdot\left(\frac{T^\alpha}{t^{H-1/2}(T-t)^{H-1/2}} - \alpha t^{\alpha+1-2H} W\left(\frac{T}{t},\ \alpha,\ H-\frac{1}{2}\right)\right), \tag{24}$$

where $W(T/t,\ \alpha,\ H-1/2) = \int_0^{T/t-1} (v+1)^{\alpha-1} v^{1/2-H}\,\mathrm{d}v$. The asymptotic behaviour of the function $W(T/t,\ \alpha,\ H-1/2)$ as $t \to 0+$ is

$$W\left(\frac{T}{t},\ \alpha,\ H-\frac{1}{2}\right) \sim \begin{cases} \mathrm{B}(\frac{3}{2}-H,\ H-\frac{1}{2}-\alpha) & \text{if } \alpha < H-\frac{1}{2}, \\ \ln\frac{T}{t} & \text{if } \alpha = H-\frac{1}{2}, \\ \frac{2}{2\alpha+1-2H}\frac{T^{\alpha-H+1/2}}{t^{\alpha-H+1/2}} & \text{if } \alpha > H-\frac{1}{2}. \end{cases}$$

Therefore, the function $h(t)$ defined in (24) is square integrable if $\alpha+1-2H-\max(0,\ \alpha-H+1/2) > -1/2$, which holds if $\alpha > 2H - 3/2$. Note that if $\alpha > 2H - 3/2$, then the following inequalities hold true: $\alpha > 2H - 2$ (whence h defined in (24) is indeed a solution to the integral equation $\Gamma h = g$), $\alpha > H - 1$ (whence conditions (5) and so (18) are satisfied), and $\alpha > -1/2$ (whence condition (B) is satisfied).

Corollary 5. *If* $\alpha > 2H - 3/2$, *the conditions of Theorems* 3–7 *are satisfied. The estimator* $\hat\theta_T$ *is consistent,* L_2*-consistent and strongly consistent. For fixed* T, *it can be approximated by discrete-sample estimator in mean-square sense.*

Simulations

Tables 1 and 2 contain the results of numerical simulations for model (23) with $\alpha = 1$ and $\alpha = 2$, respectively. For $T = 1$ and $T = 10$ and various values of H, we find h_T directly by (24). For $\theta = 2$, we simulate 1000 realizations of the process for each H and compute the values of $\hat{\theta}_T$ by (13). The means and standard deviations of these estimates are reported. We see that these simulation studies confirm the consistency of $\hat{\theta}_T$. The results are quite similar for different values of H. Moreover, the increase of α increases the rate of convergence.

Table 1. $X_t = \theta t^2 + B_t^H$, $\theta = 2$.

	$H = 0.6$		$H = 0.7$		$H = 0.8$		$H = 0.9$	
	$T = 1$	$T = 10$	$T = 1$	$T = 10$	$T = 1$	$T = 10$	$T = 1$	$T = 10$
Mean	1.9690	2.0046	1.9687	2.0098	1.9931	2.0203	1.9613	2.0354
Std. dev.	0.8361	0.0328	0.8033	0.0409	0.6977	0.0433	0.5668	0.0501

Table 2. $X_t = \theta t^3 + B_t^H$, $\theta = 2$.

	$H = 0.6$		$H = 0.7$		$H = 0.8$		$H = 0.9$	
	$T = 1$	$T = 10$	$T = 1$	$T = 10$	$T = 1$	$T = 10$	$T = 1$	$T = 10$
Mean	1.9820	2.0001	1.9847	2.0002	1.9512	1.9994	1.9827	1.9964
Std. dev.	0.7009	0.0027	0.6046	0.0032	0.5153	0.0033	0.3625	0.0033

Acknowledgment. The authors are grateful to Alexei Konstantinov for advice on the short proof of Lemma A.1. We would like to thank three anonymous referees for their useful comments and suggesting a number of improvements.

Appendix

Lemma A.1. *If $f \in L_2[0,T]$, then*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \int\limits_{(k-1)T/n}^{kT/n} \left(f(t) - \frac{n}{T} \int\limits_{(k-1)T/n}^{kT/n} f(s)\,\mathrm{d}s \right)^2 \mathrm{d}t = 0. \tag{A.1}$$

Proof. The lemma follows from the criterion of strong convergence of linear operators, since (A.1) holds true if the function $f(x)$ is continuous, the continuous functions make a dense set in $L_2[0,T]$, and the sequence of linear operators

$$A_n(f) = f - \sum_{k=1}^{n} \frac{n}{T} \int\limits_{(k-1)T/n}^{kT/n} f(s)\,\mathrm{d}s\, \mathbf{1}_{((k-1)T/N,\, kT/N]}, \quad f \in L_2[0,T],$$

is bounded in the space of linear bounded operators $L_2[0,T] \to L_2[0,T]$. □

Lemma A.2. *If $f \in L_2[0,T]$, $g \in L_2[0,T]$, then one can choose a sequence of piecewise constant functions $\{f_n(s),\ n \geqslant n_0\}$ (the function $f_n(s)$ is constant on the intervals $((k-1)T/n,\ kT/n)$) such that $\lim_{n\to\infty} \int_0^T (f(s) - f_n(s))^2 \,\mathrm{d}s = 0$, and, for sufficiently large n, $\int_0^T f_n(t)g(t)\,\mathrm{d}t = \int_0^T f(t)g(t)\,\mathrm{d}t$.*

Proof. If $g(t) = 0$ almost everywhere on $[0,T]$, then the statement of the lemma follows from Lemma A.1. In what follows, we assume that $g(t)$ is not zero everywhere. This is equivalent to $\int_0^T |g(s)|\,\mathrm{d}s > 0$. Put

$$s_n(t) = \operatorname{sign}\Biggl(\int_{(k-1)T/n}^{kT/n} g(s)\,\mathrm{d}s\Biggr), \quad \frac{(k-1)T}{n} < t < \frac{kT}{n};$$

$$V_n = \int_0^T s_n(t)g(t)\,\mathrm{d}t = \sum_{k=1}^n \Biggl|\int_{(k-1)T/n}^{kT/n} g(s)\Biggr|\,\mathrm{d}s.$$

Then

$$\lim_{n\to\infty} V_n = \operatorname*{Var}_{[0,T]} \int_0^t g(s)\,\mathrm{d}s = \int_0^T \bigl|g(s)\bigr|\,\mathrm{d}s > 0. \tag{A.2}$$

In particular, $V_n > 0$ for sufficiently large n. Put

$$\phi_n(t) = \frac{n}{T}\int_{(k-1)T/n}^{kT/n} f(s)\,\mathrm{d}s, \quad \frac{(k-1)T}{n} < t < \frac{kT}{n};$$

$$f_n(t) = \phi_n(t) + \frac{\int_0^T (f(s) - \phi_n(s))g(s)\,\mathrm{d}s}{V_n} s_n(t). \tag{A.3}$$

Then $\phi_n \to f$ in $L_2[0,T]$. Therefore, the numerator in (A.3) tends to 0. It follows from the uniform boundedness of $s_n(t)$ and (A.2) that $f_n \to f$ in $L_2[0,T]$. Moreover,

$$\begin{aligned}
&\int_0^T f_n(t)g(t)\,\mathrm{d}t \\
&\quad= \int_0^T \phi_n(t)g(t)\,\mathrm{d}t + \frac{\int_0^T (f(s) - \phi_n(s))g(s)\,\mathrm{d}s}{V_n}\int_0^T s_n(t)g(t)\,\mathrm{d}t \\
&\quad= \int_0^T \phi_n(t)g(t)\,\mathrm{d}t + \int_0^T \bigl(f(s) - \phi_n(s)\bigr)g(s)\,\mathrm{d}s = \int_0^T f(t)g(t)\,\mathrm{d}t
\end{aligned}$$

for all n such that $V_n > 0$. □

Lemma A.3. *Let M be a positively definite $n \times n$-matrix, a and b be n-dimensional vectors. Then $a^\top M a b^\top M^{-1} b \geqslant (a^\top b)^2$.*

Proof. The statement of the lemma holds if $b = 0$. Otherwise, M^{-1} is also positively definite, therefore, $b^\top M^{-1}b > 0$. Then

$$\begin{aligned} 0 &\leqslant \left(a - \frac{(a^\top b)}{b^\top M^{-1}b}M^{-1}b\right)^\top M\left(a - \frac{(a^\top b)}{b^\top M^{-1}b}M^{-1}b\right) \\ &= a^\top Ma - 2\frac{(a^\top b)}{b^\top M^{-1}b}b^\top M^{-1}Ma + \frac{(a^\top b)^2}{(b^\top M^{-1}b)^2}b^\top M^{-1}MM^{-1}b \\ &= a^\top Ma - \frac{(a^\top b)^2}{b^\top M^{-1}b}. \end{aligned}$$

□

Lemma A.4. *Let $0 < p < 1$ and $b > 0$.*

(i) *If $y \in L_1[0,b]$ is a solution to integral equation*

$$\int_0^b \frac{y(s)\,\mathrm{d}s}{|t-s|^p} = f(t) \tag{A.4}$$

for almost all $t \in (0,b)$, then $y(x)$ satisfies

$$y(x) = \frac{\Gamma(p)\cos\frac{\pi p}{2}}{\pi x^{(1-p)/2}}\mathcal{D}_{b-}^{(1-p)/2}\left(x^{1-p}\mathcal{D}_{0+}^{(1-p)/2}\left(\frac{f(x)}{x^{(1-p)/2}}\right)\right) \tag{A.5}$$

almost everywhere on $[0,b]$, where $\mathcal{D}_{a+}^\alpha$ and $\mathcal{D}_{b-}^\alpha$ are fractional derivatives,

$$\begin{aligned} \mathcal{D}_{a+}^\alpha f(x) &= \frac{1}{\Gamma(1-\alpha)}\frac{\mathrm{d}}{\mathrm{d}x}\left(\int_a^x \frac{f(t)}{(x-t)^\alpha}\,\mathrm{d}t\right), \\ \mathcal{D}_{b-}^\alpha f(x) &= \frac{-1}{\Gamma(1-\alpha)}\frac{\mathrm{d}}{\mathrm{d}x}\left(\int_x^b \frac{f(t)}{(t-x)^\alpha}\,\mathrm{d}t\right). \end{aligned}$$

(ii) *If $y_1 \in L_1[0,b]$ and $y_2 \in L_1[0,b]$ are two solutions to integral equation* (A.4)*, then $y_1(x) = y_2(x)$ almost everywhere on $[0,b]$.*

(iii) *If $y \in L_1[0,b]$ satisfies* (A.5) *almost everywhere on $[0,b]$ and the fractional derivatives are solutions to respective Abel integral equations, that is,*

$$\frac{1}{\Gamma(\frac{1-p}{2})}\int_0^t \frac{\mathcal{D}_{0+}^{(1-p)/2}(f(x)x^{(p-1)/2})}{(t-x)^{(p+1)/2}}\,\mathrm{d}x = \frac{f(t)}{t^{(1-p)/2}} \tag{A.6}$$

for almost all $t \in (0,b)$ and

$$\frac{1}{\Gamma(\frac{1-p}{2})}\int_x^b \frac{\pi y(s)s^{(1-p)/2}}{\Gamma(p)\cos\frac{\pi p}{2}}\frac{\mathrm{d}s}{(s-x)^{(p+1)/2}} = x^{1-p}\mathcal{D}_{0+}^{(1-p)/2}\left(\frac{f(x)}{x^{(1-p)/2}}\right) \tag{A.7}$$

for almost all $x \in (0,b)$, then $y(s)$ is a solution to integral equation (A.4).

Sketch of proof. The integral equation is solved in [6, 11]. In [6, Sect. 2.3], the equation is rewritten as

$$\frac{t^{(1-p)/2}}{\mathrm{B}(p, \frac{1-p}{2})} \int_0^t \frac{1}{(t-\tau)^{(p+1)/2}\tau^{1-p}} \int_\tau^b \frac{s^{(1-p)/2} y(s)\,\mathrm{d}s}{(s-\tau)^{(p+1)/2}}\,\mathrm{d}\tau = f(t).$$

For the new equation, the statement of the theorem can readily be obtained. □

Lemma A.5. *Let* $1/2 < p < 1$ *and* $0 < \xi < b$. *Then the integral equation*

$$\int_0^b \frac{y(s)\,\mathrm{d}s}{|s-t|^p} = \mathbf{1}_{[\xi,b]}(t) \quad \textit{for almost all } t \in [0,b] \tag{A.8}$$

has a unique solution $y \in L_2[0,b]$.

Proof. By Lemma A.4, if the solution to (A.8) exists within $L_1[0,b]$, it is equal to

$$y(x) = \frac{\Gamma(p)\cos\frac{\pi p}{2}}{\pi\Gamma(\frac{p+1}{2})x^{(1-p)/2}} \mathcal{D}_{b-}^{(1-p)/2}\left(x^{1-p}\frac{\mathrm{d}}{\mathrm{d}x}\left[\int_0^x \frac{\mathbf{1}_{[\xi,b]}(s)\,\mathrm{d}s}{s^{(1-p)/2}(x-s)^{(1-p)/2}}\right]\right).$$

Note that the function

$$\int_0^x \frac{\mathbf{1}_{[\xi,b]}(s)\,\mathrm{d}s}{s^{(1-p)/2}(x-s)^{(1-p)/2}} = \begin{cases} 0 & \text{for } 0 \leqslant x \leqslant \xi, \\ x^p \mathrm{B}(1-\frac{\xi}{x}; \frac{p+1}{2}, \frac{p+1}{2}) & \text{for } \xi \leqslant x \leqslant b \end{cases} \tag{A.9}$$

is absolutely continuous on $[0,b]$ and is equal to 0 at the neighbourhood of 0 (here $\mathrm{B}(x;\alpha,\beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t$ is the incomplete beta function). Then, by [17, Thm. 2.1], its derivative is a solution to Abel integral equation

$$\frac{1}{\Gamma(\frac{1-p}{2})}\int_0^x \frac{1}{\Gamma(\frac{p+1}{2})}\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_0^t \frac{\mathbf{1}_{[\xi,b]}(s)\,\mathrm{d}s}{s^{(1-p)/2}(t-s)^{(1-p)/2}}\right)\frac{\mathrm{d}t}{(x-t)^{(p+1)/2}} = \frac{\mathbf{1}_{[\xi,b]}(x)}{x^{(1-p)/2}}.$$

Thus, condition (A.6) holds true for $f(x) = \mathbf{1}_{[\xi,b]}(x)$.

The derivative of the right-hand side of (A.9) for $x \geqslant \xi$ is equal to

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\int_\xi^x \frac{\mathrm{d}s}{s^{(1-p)/2}(x-s)^{(1-p)/2}}\right) = \frac{p}{x^{1-p}}\mathrm{B}\left(1-\frac{\xi}{x}; \frac{p+1}{2}, \frac{p+1}{2}\right) + \frac{\xi^{(p+1)/2}}{x(x-\xi)^{(1-p)/2}}.$$

Thus,

$$y(x) = \frac{\Gamma(p)\cos\frac{\pi p}{2}}{\pi\,\Gamma(\frac{p+1}{2})x^{(1-p)/2}}\big(\mathcal{D}_{b-}^{(1-p)/2}G_1(x) + \mathcal{D}_{b-}^{(1-p)/2}G_2(x)\big) \tag{A.10}$$

with $G_1(x) = p\mathrm{B}(1-\xi/x;\, (p+1)/2,\, (p+1)/2)\,\mathbf{1}_{(\xi,+\infty)}(x)$, $G_2(x) = \xi^{(p+1)/2}/(x^p(x-\xi)^{(1-p)/2})\,\mathbf{1}_{(\xi,+\infty)}(x)$.

The function $G_1(x)$ is continuous, therefore $G_1 \in L_2[0,T]$. Also, it is piecewise differentiable and Hölder with exponent $(1+p)/2$. Consequently, its fractional derivative $\mathcal{D}_{b-}^{(1-p)/2}G_1$ is bounded on any interval $[0,b_1]$, $0<b_1<b$, see [17, Thm. 13.6]. It makes a solution to Abel integral equation: $I_{b-}^{(1-p)/2}\mathcal{D}_{b-}^{(1-p)/2}G_1 = G_1$ almost everywhere on $(0,b)$. At the neighbourhood of b, it has the asymptotic behaviour: $\mathcal{D}_{b-}^{(1-p)/2}G_1(x) \sim G_1(b)/(\Gamma((p+1)/2)(b-x)^{(1-p)/2})$, $x\to b-$. Thus, $\mathcal{D}_{b-}^{(1-p)/2}G_1 \in L_2[0,b]$.

The function $G_2(x)$ is piecewise continuous, unbounded as $x\to\xi+$, and $G_2 \in L_2[0,b]$. In order to prove that $\mathcal{D}_{b-}^{(1-p)/2}G_2 \in L_2[0,T]$, we use Theorem 13.2 from [17], according to which, $I_{b-}^{(1-p)/2}\mathcal{D}_{b-}^{(1-p)/2}G_2 = G_2$ and $\mathcal{D}_{b-}^{(1-p)/2}G_2 \in L_2[0,b]$ if and only if two conditions hold: (i) $G_2\in L_2[0,b]$; and (ii) ψ_ϵ converges in $L_2[0,b]$ as $\epsilon\to 0+$, where

$$\psi_\epsilon(x) = \begin{cases} \int_{x+\epsilon}^b \frac{G_2(t)-G_2(x)}{(t-x)^{(3-p)/2}}\,\mathrm{d}t, & 0\leqslant x\leqslant b-\epsilon,\\ \int_{x+\epsilon}^b \frac{-G_2(x)\,\mathrm{d}t}{(t-x)^{(3-p)/2}} = \frac{-2G_2(x)}{1-p}\bigl(\frac{1}{\epsilon^{(1-p)/2}} - \frac{1}{(b-x)^{(1-p)/2}}\bigr), & b-\epsilon\leqslant x<b,\end{cases}$$

for $b-\epsilon\leqslant x<b$. Condition (i) holds true. Condition (ii) is an immediate consequence of the following assertions:

- The integrand, which equals either $(G_2(t)-G_2(x))(t-x)^{(p-3)/2}$ if $x+\epsilon\leqslant t\leqslant b$ or $-G_2(t)(t-x)^{(p-3)/2}$ if $b\leqslant t\leqslant x+\epsilon$, $t>x$, is positive or zero if $0\leqslant x\leqslant\xi$ and negative if $\xi<x<b$. Therefore, for fixed x, $\psi_\epsilon(x)$ is monotonous in ϵ.
- At $\epsilon=b$,

$$\psi_b(x) = \begin{cases} 0, & 0\leqslant x\leqslant\xi,\\ \frac{2}{1-p}\frac{\xi^{(p+1)/2}}{x^p(x-\xi)^{(1-p)/2}}\bigl(\frac{1}{(b-x)^{(1-p)/2}} - \frac{1}{b^{(1-p)/2}}\bigr), & \xi<x<b,\end{cases}$$

  belongs to $L_2[0,T]$.
- At $\epsilon=0$,

$$\psi_0(x) = \int_x^b \frac{G_2(t)-G_2(x)}{(t-x)^{(3-p)/2}}\,\mathrm{d}t,$$

  and

$$\bigl|\psi_0(x)\bigr| < \frac{\mathrm{const}}{|x-\xi|^{1-p}}$$

  with

$$\mathrm{const} = \xi^{(1-p)/2}\frac{1+p}{1-p}\int_0^\infty \frac{\mathrm{d}u}{u^{(1-p)/2}(1+u)^{(3-p)/2}} < \frac{\xi^{(1-p)/2}(3-p)}{(1-p)^2}.$$

  Therefore, $\psi_0\in L_2[0,b]$.

- $\lim_{\epsilon\to 0+}\psi_\epsilon(x) = \psi_0(x)$ for all x, $0 \leqslant x < b$, because the Lebesgue integral is continuous-on-the-right with respect to the lower limit. The limit is finite at the points where $\psi_0(x)$ is finite, i.e., almost everywhere on $(0,b)$; however $\lim_{\epsilon\to 0+}\psi_\epsilon(\xi) = +\infty = \psi_0(\xi)$.

Thus, $\mathcal{D}_{b-}^{(1-p)/2}G_2 \in L_2[0,b]$. Moreover, for $x<\xi$, the function

$$\mathcal{D}_{b-}^{(1-p)/2}G_2(x) = \frac{1}{\Gamma(\frac{p-1}{2})}\int_\xi^b \frac{G_2(t)\,\mathrm{d}t}{(t-x)^{(3-p)/2}}$$

is bounded in the neighbourhood of 0.

Since the function $\mathcal{D}_{b-}^{(1-p)/2}G_1(x) + \mathcal{D}_{b-}^{(1-p)/2}G_2(x)$ is square integrable on $(0,b)$ and bounded in the neighbourhood of 0, multiplying it by $x^{(p-1)/2}$ does not drive it out of $L_2[0,b]$. Thus, $y \in L_2[0,b]$, see (A.10). From (A.10) and equalities

$$I_{b-}^{(1-p)/2}\mathcal{D}_{b-}^{(1-p)/2}G_1 = G_1 \quad \text{and} \quad I_{b-}^{(1-p)/2}\mathcal{D}_{b-}^{(1-p)/2}G_2 = G_2$$

it follows that

$$I_{b-}^{(1-p)/2}\left(\frac{\pi\Gamma(p)y(x)x^{(1-p)/2}}{\Gamma(p)\cos\frac{\pi p}{2}}\right) = G_1(x) + G_2(x) = x^{1-p}\mathcal{D}_{0+}^{(1-p)/2}\left(\frac{\mathbf{1}_{[\xi,b]}(x)}{x^{(1-p)/2}}\right).$$

Condition (A.7) holds true. By statement (iii) of Lemma A.4, $y(x)$ is indeed a solution to integral equation (A.8). Uniqueness of the solution to (A.8) follows from statement (ii) of Lemma A.4 and from the fact that $L_2[0,b] \subset L_1[0,b]$. □

References

1. T.V. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, Hoboken NJ, 2003.
2. J. Berger, R. Wolpert, Estimating the mean function of a Gaussian process and the Stein effect, *J. Multivariate Anal.*, **13**(3):401–424, 1983.
3. K. Bertin, S. Torres, C. A. Tudor, Maximum-likelihood estimators and random walks in long memory models, *Statistics*, **45**(4):361–374, 2011.
4. T. Bojdecki, L.G. Gorostiza, A. Talarczyk, Sub-fractional Brownian motion and its relation to occupation times, *Stat. Probab. Lett.*, **69**(4):405–419, 2004.
5. V. Buldygin, *Convergence of Random Elements in Topological Spaces*, Naukova Dumka, Kyiv, 1980 (in Russian).
6. C. Cai, P. Chigansky, M. Kleptsyna, The maximum likelihood drift estimator for mixed fractional Brownian motion, 1st version, arXiv:1208.6253v1.
7. C. Cai, P. Chigansky, M. Kleptsyna, Mixed Gaussian processes: A filtering approach, *Ann. Probab.*, **44**(4):3032–3075, 2016.
8. K. Es-Sebaiy, I. Ouassou, Y. Ouknine, Estimation of the drift of fractional Brownian motion, *Stat. Probab. Lett.*, **79**(14):1647–1653, 2009.

9. Y. Hu, D. Nualart, W. Xiao, W. Zhang, Exact maximum likelihood estimator for drift fractional Brownian motion at discrete observation, *Acta Math. Sci., Ser. B, Engl. Ed.*, **31**(5):1851–1859, 2011.
10. I.A. Ibragimov, Y.A. Rozanov, *Gaussian Random Processes*, Appl. Math. (N. Y.), Vol. 9, Springer, New York, Berlin, 1978.
11. A. Le Breton, Filtering and parameter estimation in a simple linear system driven by a fractional Brownian motion, *Stat. Probab. Lett.*, **38**(3):263–274, 1998.
12. Yu. Mishura, Maximum likelihood drift estimation for the mixing of two fractional Brownian motions, in *Stochastic and Infinite Dimensional Analysis*, Birkhäuser, Basel, 2016, pp. 263–280.
13. Yu. Mishura, K. Ralchenko, S. Shklyar, Maximum likelihood drift estimation for Gaussian process with stationary increments, *Austrian J. Stat.*, **46**(3–4):67–78, 2017.
14. Yu. Mishura, I. Voronov, Construction of maximum likelihood estimator in the mixed fractional-fractional Brownian motion model with double long-range dependence, *Mod. Stoch., Theory Appl.*, **2**(2):147–164, 2015.
15. I. Norros, E. Valkeila, J. Virtamo, An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions, *Bernoulli*, **5**(4):571–587, 1999.
16. N. Privault, A. Réveillac, Stein estimation for the drift of Gaussian processes using the Malliavin calculus, *Ann. Stat.*, **36**(5):2531–2550, 2008.
17. S. Samko, A. Kilbas, O. Marichev, *Fractional Integrals and Derivatives*, Gordon and Breach, Yverdon, 1993.
18. P.A. Samuelson, Rational theory of warrant pricing, *Ind. Manag. Rev.*, **6**(2):13–32, 1965.
19. G. Shen, L. Yan, Estimators for the drift of subfractional Brownian motion, *Commun. Stat., Theory Methods*, **43**(8):1601–1612, 2014.
20. C. Tudor, Some properties of the sub-fractional Brownian motion, *Stochastics*, **79**(5):431–448, 2007.