

Article

Least-Squares Estimators of Drift Parameter for Discretely Observed Fractional Ornstein–Uhlenbeck Processes

Pavel Kríž *  and Leszek Szala

Department of Mathematics, Faculty of Chemical Engineering, University of Chemistry and Technology Prague, 16628 Prague, Czech Republic; szalal@vscht.cz

* Correspondence: krizp@vscht.cz; Tel.: +420-220-445-034

Received: 3 April 2020; Accepted: 22 April 2020; Published: 3 May 2020



Abstract: We introduce three new estimators of the drift parameter of a fractional Ornstein–Uhlenbeck process. These estimators are based on modifications of the least-squares procedure utilizing the explicit formula for the process and covariance structure of a fractional Brownian motion. We demonstrate their advantageous properties in the setting of discrete-time observations with fixed mesh size, where they outperform the existing estimators. Numerical experiments by Monte Carlo simulations are conducted to confirm and illustrate theoretical findings. New estimation techniques can improve calibration of models in the form of linear stochastic differential equations driven by a fractional Brownian motion, which are used in diverse fields such as biology, neuroscience, finance and many others.

Keywords: fractional Brownian motion; Ornstein–Uhlenbeck process; drift parameter estimation

MSC: 60G22; 62M09

1. Introduction

Stochastic models with fractional Brownian motion (fBm) as the noise source have attained increasing popularity recently. This is because fBm is a continuous Gaussian process, increments of which are positively, or negatively correlated if Hurst parameter $H > 1/2$, or $H < 1/2$, respectively. If $H = 1/2$ fBm coincides with classical Brownian motion and its increments are independent. The ability of fBm to include memory into the noise process makes it possible to build more realistic models in such diverse fields as biology, neuroscience, hydrology, climatology, finance and many others. The interested reader may check monographs [1,2], or more recent paper [3] and the references therein for more information.

Let $\{B_t^{(H)}\}_{t \in [0, \infty)}$ be a fractional Brownian motion with Hurst parameter H defined on an appropriate probability space $\{\Omega, \mathcal{A}, P\}$. Fractional Ornstein–Uhlenbeck process (fOU) is the unique solution to the following linear stochastic differential equation

$$dX_t = -\lambda X_t dt + \sigma dB_t^{(H)}, \quad X_0 = x_0 \in \mathbb{R}, \quad t \geq 0, \quad (1)$$

where $\lambda > 0$ is a drift parameter (we consider ergodic case only) and $\sigma > 0$ is a noise intensity (or volatility). Recall that solution to Equation (1) can be expressed by the exact analytical formula:

$$X_t = e^{-\lambda t} x_0 + \int_0^t e^{-\lambda(t-u)} \sigma dB_u^{(H)}. \quad (2)$$

A single realization of the random process $\{X_t(\omega)\}_{t \in [0, \infty)}$ for a particular $\omega \in \Omega$ is the model for the single real-valued trajectory, part of which is observed. Two examples of such trajectories are given in Figure 1. We assume $H > 1/2$ throughout this paper so that the fOU exhibits long-range dependence. For an example of application, see a neuronal model based on fOU described in the recent work [4].

The aim of this paper is to study the problem of estimating drift parameter λ based on an observation of a single trajectory of a fOU in discrete time instants $t = 0, h, \dots, Nh$ with fixed mesh size $h > 0$ and increasing time horizon $T = Nh \rightarrow \infty$ (long-span asymptotics). Estimating drift parameter of a fOU observed in continuous time has been considered in [5,6], where least-squares estimator (LSE) and ergodic-type estimator are studied. These have advantageous asymptotic properties, they are strongly consistent and, if $H \leq 3/4$, also asymptotically normal. Ergodic-type estimator is easy to implement, but it has greater asymptotic variance compared to LSE, requires a priori knowledge of H and σ and does not provide acceptable results for non-stationary processes with limited time horizon.

A straightforward discretization of the least-squares estimator for a fOU has been introduced and studied in [7] for $H > 1/2$ and in [8] for $H < 1/2$. For the precise formula, see (8). This estimator is consistent provided that both the time horizon $T = Nh \rightarrow \infty$ and the mesh size $h \rightarrow 0$ (mixed in-fill and long-span asymptotics). However, it is not consistent when h is fixed and $T \rightarrow \infty$. This has led us to construct and study LSE-type estimators that converge in this long-span setting.

An easy modification of the ergodic-type estimator to discrete-time setting with fixed time step was given in [9], see (10) for precise formula, and its strong consistency (assuming $H \geq 1/2$) and asymptotic normality (for $1/2 \leq H < 3/4$) when $N \rightarrow \infty$ were proved, but with possibly incorrect technique (as pointed out in [10]). Correct proofs of asymptotic normality for $0 < H \leq 3/4$ and strong consistency for $0 < H < 1$ of this estimator were provided (in more general setup) in [10]. Note that the use of this discrete ergodic estimator requires the knowledge of parameter σ (in contrast to the estimators of least-squares type introduced below). Other works related to estimating drift parameter for discretely observed fOU include [11–13], but this list is by no means complete.

This work contributes to the problem of estimating drift parameter of fOU by introducing three new LSE-type estimators: least-squares estimator from exact solution, asymptotic least-squares estimator and conditional least-squares estimator. These estimators are tailored to discrete-time observations with fixed time step. We provide proofs of their asymptotic properties and identify situations, in which these new estimators perform better than the already known ones. In particular, we eliminate the discretization error (the LSE from exact solution), construct strongly consistent estimators in the long-span regime without assuming in-fill condition (the asymptotic LSE and the conditional LSE), and eliminate the bias in the least-squares procedure caused by autocorrelation of the noise term (the conditional LSE). Especially the conditional LSE demonstrates outstanding performance in all studied scenarios. This suggests that the newly introduced (to our best knowledge) concept of conditioning in the least-squares procedure applied to the models with fractional noise provides a powerful framework for parameter estimation in this type of models. The proof of its strong consistency, presented within this paper, is rather non-trivial and may serve as a starting point for investigation of similar estimators in possibly different settings. A certain disadvantage of the conditional LSE is its complicated implementation (involving optimization procedure), which is in contrast to the other studied estimators.

Let us explain the strength of the conditional least-squares estimator in more detail. Comparison of the two trajectories in Figure 1 demonstrates the effect of different values of λ on trajectories of fOU. In particular, it affects the speed of exponential decay in initial non-stationary phase and the variability in stationary phase. As we illustrate below, the discretized least-squares estimator, cf. (8), utilizes information about λ from the exponential decay in initial phase, but is not capable to make use of the information contained in the variability in stationary phase. As a consequence, it is not consistent (in long-span setting). On the contrary, the ergodic-type estimator, cf. (10), is derived from the variance of the stationary distribution of the process. It works well for stationary processes (and is consistent), but leaves idle (and even worse, it is corrupted by) the observation of the process in its

initial non-stationary phase. In result, neither of these estimators can efficiently estimate drift from long trajectories with far-from-stationary initial values. This gap is best filled with the conditional least-squares estimator, cf. (25), which effectively utilizes both information stored in non-stationary phase and in stationary phase of the observed process. This unique property is demonstrated in Results and Discussion, where the conditional LSE (denoted by $\hat{\lambda}_5$) dominates the other estimators.

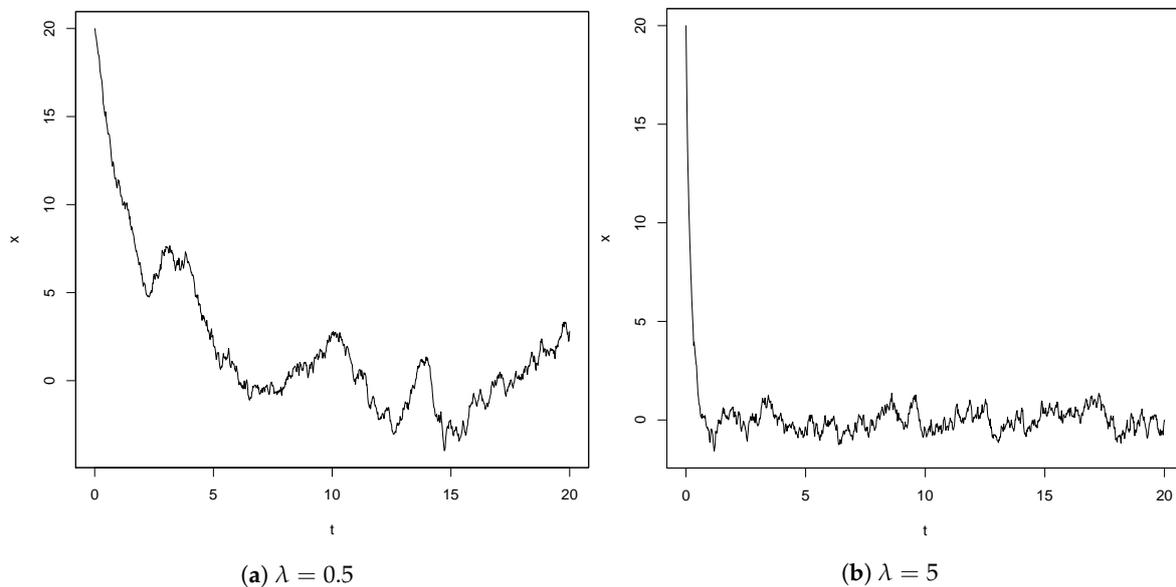


Figure 1. Two single trajectories of fOU with different values of λ , where $\sigma = 2$, $x_0 = 20$, $T = 20$, $H = 0.6$.

For the three newly introduced estimators the value of the Hurst parameter H is considered to be known a priori, whereas the knowledge of volatility parameter σ is not required, which is an advantage of these methods. If H is not known, it can be estimated in advance by some of many methods, such as methods based on quadratic variations (cf. [14]), sample quantiles or trimmed means (cf. [15]), or on a wavelet transform (cf. [16]), to name just a few. Another useful works in this direction include simultaneous estimation of σ and H using the powers of the second order variations (see [17], Chapter 3.3). The estimates of H (obtained independently from λ) can subsequently be used in the LSE-type estimators of lambda introduced below in a way similar to [18].

In Section 2, some elements of stochastic calculus with respect to fBm are recalled, stationary fOU is introduced and precise formulas for two existing drift estimators $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are provided. Section 3 is devoted to construction of a new LSE type estimator ($\hat{\lambda}_3$) based on exact formula for fOU. A certain modification of $\hat{\lambda}_3$ (denoted as $\hat{\lambda}_4$), which ensures long-span consistency, is introduced in Section 4. In Section 5, we rewrite the linear model using conditional expectations to overcome the bias in LSE caused by autocorrelation of the noise. Least-squares method, applied to the conditional model with explicit formulas for conditional expectations, results in the conditional least-squares estimator ($\hat{\lambda}_5$). We prove strong consistency of this estimator. The actual performance of the newly introduced estimators $\hat{\lambda}_3, \hat{\lambda}_4$ and $\hat{\lambda}_5$ as well as its comparison to the already-known $\hat{\lambda}_1$ and $\hat{\lambda}_2$, is studied by Monte Carlo simulations in various scenarios and reported in Section 6. The simulated trajectories have been obtained in software R with YUIMA package (see [19]). Section 7 summarizes key points of the article and provides possible future extensions.

2. Preliminaries

For reader’s convenience we briefly review the basic concepts from theory of stochastic models with fractional noise in this section, including definition of fBm, Wiener integral of deterministic functions w.r.t. fBm and stationary fOU. This exposition follows [2,20]. For further reading, see also

the monograph [1]. In the end of this section, we also recall formulas for discretized LSE and discrete ergodic estimator.

Fractional Brownian motion with Hurst parameter $H \in (0, 1)$ is a centered (zero-mean) continuous Gaussian process $\{B_t^{(H)}\}_{t \in [0, \infty)}$ starting from zero ($B_0^{(H)} = 0$) and having the following covariance structure

$$E \left(B_s^{(H)} B_t^{(H)} \right) = \frac{1}{2} \left(s^{2H} + t^{2H} - |s - t|^{2H} \right), \quad s, t \geq 0.$$

Note that for the purpose of construction of the stationary fOU, we need a two-sided fBm $\{B_t^{(H)}\}_{t \in \mathbb{R}}$ with t ranging over the whole \mathbb{R} . In this case we have

$$E \left(B_s^{(H)} B_t^{(H)} \right) = \frac{1}{2} \left(|s|^{2H} + |t|^{2H} - |s - t|^{2H} \right), \quad s, t \in \mathbb{R}.$$

As a consequence, the increments of fBm are negatively correlated for $H < 1/2$, independent for $H = 1/2$ and positively correlated for $H > 1/2$.

Consider a two-sided fBm with $H > 1/2$ and define Wiener integral of a deterministic step function with respect to the fBm by formula

$$\int_{\mathbb{R}} \sum_{i=1}^N \alpha_i I_{[t_i, t_{i+1}]}(s) dB_s^{(H)} = \sum_{i=1}^N \alpha_i \left(B_{t_{i+1}}^{(H)} - B_{t_i}^{(H)} \right)$$

for any positive integer N , real-valued coefficients $\alpha_1, \dots, \alpha_N \in \mathbb{R}$ and a partition $-\infty < t_1 \leq \dots \leq t_{N+1} < \infty$. This definition constitutes the following isometry for any pair of deterministic step functions f and g

$$E \left(\int_{\mathbb{R}} f(t) dB_t^{(H)} \int_{\mathbb{R}} g(s) dB_s^{(H)} \right) = \iint_{\mathbb{R}^2} f(t) g(s) \varphi(t, s) dt ds = \langle f, g \rangle_H, \tag{3}$$

where $\varphi(t, s) = H(2H - 1)|t - s|^{2H-2}$. Using this isometry, we can extend the definition of the Wiener integral w.r.t. fBm to all elements of the space $\mathcal{L}_H^2(\mathbb{R})$, defined as the completions of the space of deterministic step functions w.r.t. the scalar product $\langle \cdot, \cdot \rangle_H$ defined above. In result, the formula (3) holds true for any $f, g \in \mathcal{L}_H^2(\mathbb{R})$, see also [21]. We will frequently use this formula in what follows, mainly to calculate the covariances of Wiener integrals.

Let $\{B_t^{(H)}\}_{t \in \mathbb{R}}$ be again a two-sided fBm with $H > 1/2$. Define

$$Z_0 = \int_{-\infty}^0 e^{-\lambda(0-u)} \sigma dB_u^{(H)},$$

and denote by $\{Z_t\}_{t \in [0, \infty)}$ the solution to (1) with initial condition Z_0 in the sense that it satisfies

$$dZ_t = -\lambda Z_t dt + \sigma dB_t^{(H)}.$$

This process is referred to as the stationary fOU and it can be expressed as

$$Z_t = \int_{-\infty}^t e^{-\lambda(t-u)} \sigma dB_u^{(H)}, \quad t \geq 0.$$

Note that the stationary fOU is an ergodic stationary Gaussian process (its autocorrelation function vanishes at infinity).

Consider now a stationary fOU $\{Z_t\}_{t \in [0, \infty)}$ observed at discrete time instants $t = 0, h, 2h, \dots$. The ergodicity and the formula for the second moment of stationary fOU (see e.g., [20]) imply

$$\frac{1}{N} \sum_{n=0}^{N-1} Z_{nh}^2 \xrightarrow{N \rightarrow \infty} \mathbb{E}Z_0^2 = \frac{\sigma^2}{\lambda^{2H}} H\Gamma(2H) \quad \text{a.s.} \tag{4}$$

Analogously,

$$\frac{1}{N} \sum_{n=0}^{N-1} Z_{nh}Z_{(n+1)h} \xrightarrow{N \rightarrow \infty} \mathbb{E}Z_0Z_h \quad \text{a.s.,} \tag{5}$$

and the expectation can be calculated using (3) and the change-of-variables formula

$$\begin{aligned} \mathbb{E}Z_0Z_h &= \mathbb{E}Z_0\left(e^{-\lambda h}Z_0 + \int_0^h e^{-\lambda(h-u)}\sigma dB_u^{(H)}\right) \\ &= e^{-\lambda h} \frac{\sigma^2}{\lambda^{2H}} H\Gamma(2H) + \int_0^h \int_{-\infty}^0 e^{-\lambda(h-u)} e^{-\lambda(0-v)} \sigma^2 (u-v)^{2H-2} H(2H-1) dv du \\ &= e^{-\lambda h} \frac{\sigma^2}{\lambda^{2H}} H\Gamma(2H) \left[1 + \frac{2H-1}{\Gamma(2H)} \int_0^{\lambda h} \int_{-\infty}^0 e^{r+s} (r-s)^{2H-2} ds dr \right]. \end{aligned} \tag{6}$$

The rest of this section is devoted to the two popular estimators of the drift parameter of fOU observed at discrete time instants described in Introduction—the discretized LSE and the discrete ergodic estimator. Start with the former. Consider a straightforward discrete approximation of the Equation (1):

$$X_{n+h} - X_n \approx -\lambda X_n h + \sigma(B_{n+h}^{(H)} - B_n^{(H)}). \tag{7}$$

Application of the standard least-squares procedure to the linear approximation above provides the discretized LSE studied in [7,8], which takes the form

$$\hat{\lambda}_1 = -\frac{1}{h} (F_N - 1), \tag{8}$$

where h is the mesh size (time step) and

$$F_N = \frac{\sum_{n=0}^{N-1} X_{nh}X_{(n+1)h}}{\sum_{n=0}^{N-1} X_{nh}^2}, \tag{9}$$

with X_{nh} and $X_{(n+1)h}$ being the observations at adjacent time instants $t = nh$ and $t = (n + 1)h$ respectively, of the process X_t defined by (1) or (2). Note that having $\hat{\lambda}_1$ expressed in term of F_N simplifies its comparison with the estimators newly constructed in this paper. Recall that for consistency of $\hat{\lambda}_1$ mixed in-fill and long-span asymptotics is required due to the approximation error in (7).

The discrete ergodic estimator is derived from asymptotic behavior of the (stationary) fOU. Recall the convergence in (4). Rearranging the terms provides an asymptotic formula for drift parameter λ expressed in terms of the limit of the second sample moment of the stationary fOU. Substituting the stationary fOU by the observed fOU $X_h, X_{2h}, \dots, X_{Nh}$ in the asymptotic formula results in the discrete ergodic estimator:

$$\hat{\lambda}_2 = \left(\frac{1}{N\sigma^2 H\Gamma(2H)} \sum_{i=1}^N X_{ih} \right)^{-\frac{1}{2H}}, \tag{10}$$

which was studied in [9,10]. Recall that this estimator is strongly consistent in the long span regime (no in-fill condition needed), however, it heavily builds upon the asymptotic (stationary) behavior of the process and fails for processes with non-stationary initial phase (as illustrated by numerical experiments below).

3. Least-Squares Estimator from Exact Solution

Since the estimator $\hat{\lambda}_1$ obtained from naive discretization of (1) provides reasonable approximations only for non-stationary solutions with short time horizon and small time step $h > 0$ (as seen from numerical simulations below), we eliminate discretization error by considering exact analytical formula for X_t , see (2), and corresponding exact discrete formula for X_{t+h} ,

$$X_{t+h} = \beta X_t + \zeta_t, \tag{11}$$

where

$$\beta = e^{-\lambda h}, \quad \text{and} \quad \zeta_t = \int_t^{t+h} e^{-\lambda(t+h-u)} \sigma dB_u^{(H)}.$$

The least-squares estimator for β w.r.t. linear model (11) is given by $\hat{\beta} = F_N$, cf. (9), and the estimator for λ can be defined as

$$\hat{\lambda}_3 = -\frac{1}{h} \log F_N. \tag{12}$$

Numerical simulations show that $\hat{\lambda}_3$ works well for non-stationary solutions and short time horizon ($T = 10$ in simulations). The results for $H = 0.6$ are presented in Figure 2. Simulation results for $H \in \{0.75, 0.9\}$ are similar.

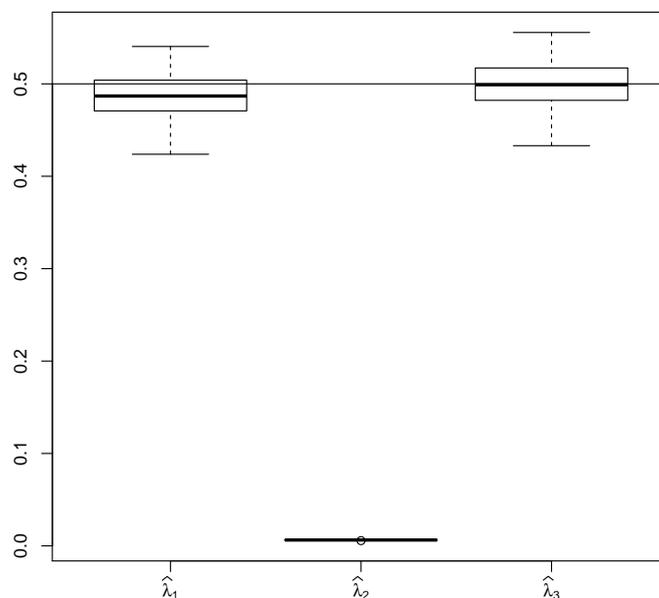


Figure 2. Comparison of $\hat{\lambda}_1$, $\hat{\lambda}_2$ and $\hat{\lambda}_3$ for 100 trajectories, where $H = 0.6$, $x_0 = 100$, $T = 10$, $h = 0.1$, $\sigma = 2$; the horizontal line shows the true value of the estimated parameter $\lambda = 0.5$.

On the other hand, estimator $\hat{\lambda}_3$ does not provide good results for observations with long time horizon ($T = 1000$ in simulations) since $\hat{\lambda}_3$ is not consistent if $N \rightarrow \infty$ and $h > 0$ is fixed. The reason is that ζ_t and X_t in (11) are correlated. In fact, we can calculate the almost sure limit of $\hat{\lambda}_3$ exactly.

The limit is provided in Theorem 1. Its proof uses the following simple lemma (see [22]) to show the diminishing effect of the initial condition on limiting behaviour of sample averages. It is later used in the proof of Lemma 3 as well.

Lemma 1. Consider real-valued sequences $(a_n)_{n=1}^\infty$ and $(b_n)_{n=1}^\infty$ such that

$$\frac{1}{N} \sum_{n=1}^N |b_n| \xrightarrow{N \rightarrow \infty} K < \infty, \quad \text{and } a_n \xrightarrow{n \rightarrow \infty} 0.$$

Then $\frac{1}{N} \sum_{n=1}^N a_n b_n \xrightarrow{N \rightarrow \infty} 0$.

Theorem 1. Let $h > 0$ be fixed and define $f: (0, \infty) \rightarrow \mathbb{R}$ by

$$f(x) = e^{-x} \left[1 + \frac{2H-1}{\Gamma(2H)} \int_0^x \left(\int_{-\infty}^0 e^s e^r (r-s)^{2H-2} ds \right) dr \right]. \tag{13}$$

Then

$$\lim_{N \rightarrow \infty} \widehat{\lambda}_3 = -\frac{1}{h} \log f(\lambda h) \quad \text{almost surely.} \tag{14}$$

In particular, $\lim_{N \rightarrow \infty} \widehat{\lambda}_3 < \lambda$.

Proof. Recall the asymptotic behavior of a stationary fOU described in formulas (4)–(6).

Since the effect of the initial condition vanishes at infinity, the limit behaviour of the non-stationary solution $(X_{nh})_{n=0}^\infty$ is same. Indeed,

$$\frac{1}{N} \sum_{n=0}^{N-1} X_{nh} X_{(n+1)h} - Z_{nh} Z_{(n+1)h} = \frac{1}{N} \sum_{n=0}^{N-1} (X_{nh} - Z_{nh}) X_{(n+1)h} + \frac{1}{N} \sum_{n=0}^{N-1} (X_{(n+1)h} - Z_{(n+1)h}) Z_{nh}.$$

The convergence of the first summand to zero follows from the facts that

$$(X_{nh} - Z_{nh}) \xrightarrow{n \rightarrow \infty} 0 \quad \text{a.s.,} \quad \frac{1}{N} \sum_{n=0}^{N-1} X_{(n+1)h} \xrightarrow{N \rightarrow \infty} 0 \quad \text{a.s.}$$

and Lemma 1. Similar argument guarantees convergence of the second summand to zero as well. The convergence

$$\frac{1}{N} \sum_{n=0}^{N-1} X_{nh}^2 - Z_{nh}^2 \xrightarrow{N \rightarrow \infty} 0$$

can be shown correspondingly. In result, we obtain the almost sure convergence

$$\frac{\sum_{n=0}^{N-1} X_{nh} X_{(n+1)h}}{\sum_{n=0}^{N-1} X_{nh}^2} \xrightarrow{N \rightarrow \infty} e^{-\lambda h} \left[1 + \frac{2H-1}{\Gamma(2H)} \int_0^{\lambda h} \int_{-\infty}^0 e^{r+s} (r-s)^{2H-2} ds dr \right] = f(\lambda h). \tag{15}$$

The claim follows immediately from definition of $\widehat{\lambda}_3$. \square

Remark 1. Note that the convergence in (14) holds true also for $H = 1/2$ (the double integral in f disappears) and for $H < 1/2$ (utilizing the fact that a relation analogous to (3) is true even for $H < 1/2$, if the two domains of integration are disjoint). Consequently

$$\lim_{N \rightarrow \infty} \hat{\lambda}_3 = -\frac{1}{h} \log f(\lambda h) \begin{cases} < \lambda, & H > 1/2, \\ = \lambda, & H = 1/2, \\ > \lambda, & H < 1/2. \end{cases} \quad a.s.$$

4. Asymptotic Least-Squares Estimator

Our goal in this section is modifying $\hat{\lambda}_3$ so that it converges to λ when $N \rightarrow \infty$ and $h > 0$ is fixed. Combination of (12) and (14) yields $F_N \rightarrow f(\lambda h)$ a.s. Thus, we can define the asymptotic least-squares estimator $\hat{\lambda}_4$ by relation $f(h\hat{\lambda}_4) = F_N$. Since f is one-to-one (see below), the explicit formula for $\hat{\lambda}_4$ reads

$$\hat{\lambda}_4 = \frac{1}{h} f^{-1}(F_N). \tag{16}$$

The following lemma justifies invertibility of f in the definition of $\hat{\lambda}_4$.

Lemma 2. In our setting $H > 1/2$ the function f defined by (13) is strictly decreasing on $[0, \infty)$.

Proof. Calculate the derivative

$$\begin{aligned} f'(x) &= -e^{-x} \left[1 + \frac{2H-1}{\Gamma(2H)} \int_0^x \left(\int_{-\infty}^0 e^s e^r (r-s)^{2H-2} ds \right) dr \right] + e^{-x} \left[\frac{2H-1}{\Gamma(2H)} \left(\int_{-\infty}^0 e^s e^x (x-s)^{2H-2} ds \right) \right] \\ &= e^{-x} \frac{2H-1}{\Gamma(2H)} \left[\int_{-\infty}^0 \left(e^s e^x (x-s)^{2H-2} - \int_0^x e^s e^r (r-s)^{2H-2} dr \right) ds - \frac{\Gamma(2H)}{2H-1} \right]. \end{aligned}$$

Continue with

$$e^s e^x (x-s)^{2H-2} - \int_0^x e^s e^r (r-s)^{2H-2} dr < (x-s)^{2H-2} e^s \left(e^x - \int_0^x e^r dr \right) = (x-s)^{2H-2} e^s < (-s)^{2H-2} e^s.$$

Plug this estimate into the formula for $f'(x)$ to see

$$f'(x) < e^{-x} \frac{2H-1}{\Gamma(2H)} \left[\int_{-\infty}^0 (-s)^{2H-2} e^s ds - \frac{\Gamma(2H)}{2H-1} \right] = e^{-x} \frac{2H-1}{\Gamma(2H)} [\Gamma(2H-1) - \Gamma(2H-1)] = 0. \tag{17}$$

□

Remark 2. Note that f is monotonous also for $H = 1/2$, but it is not monotonous if $H < 1/2$, which rules out the possibility to use estimator $\hat{\lambda}_4$ in this singular case.

Theorem 2. The asymptotic least-squares estimator $\hat{\lambda}_4$ is strongly consistent, i.e.,

$$\lim_{N \rightarrow \infty} \hat{\lambda}_4 = \lambda \quad a.s.$$

Proof. Recall the definition of $\hat{\lambda}_4$ in (16) and the limit in (15):

$$F_N \xrightarrow{N \rightarrow \infty} f(\lambda h) \quad \text{a.s.}$$

Further recall that $f : (0, \infty) \rightarrow (0, \infty)$ is differentiable with strictly negative derivative (see (17)). Thus, f^{-1} is also differentiable with strictly negative derivative and this implies

$$\lim_{N \rightarrow \infty} \hat{\lambda}_4 = \lim_{N \rightarrow \infty} \frac{1}{h} f^{-1}(F_N) = \frac{1}{h} f^{-1}(f(\lambda h)) = \lambda \quad \text{a.s.}$$

□

The strongly consistent estimator $\hat{\lambda}_4$ works well for stationary solutions or observations with long time horizon (see Figure 3). Moreover it does not require explicit knowledge of σ (in contrast to $\hat{\lambda}_2$, which is also strictly consistent). On the other hand it does not provide adequate results for non-stationary solutions and short time horizon, since the correction function f reflects stationary behavior of the process (see Figure 4).

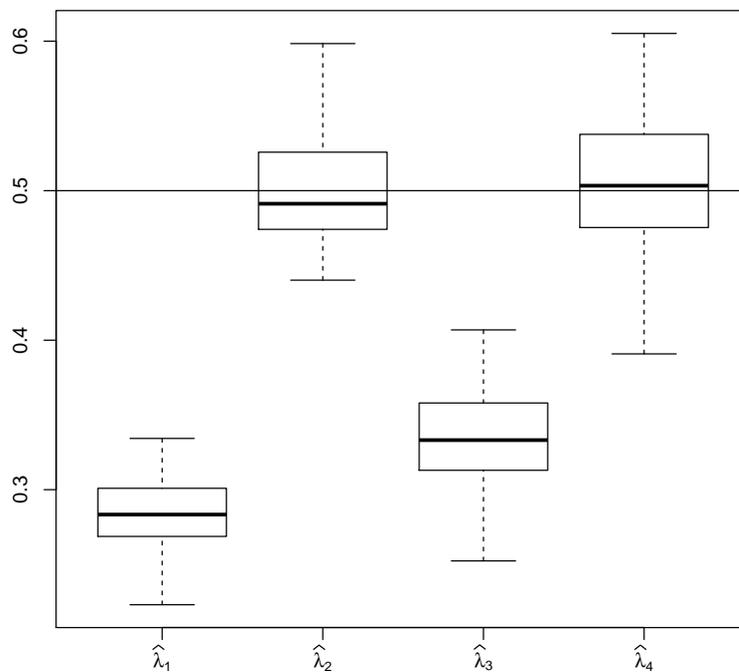


Figure 3. Comparison of $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{\lambda}_3$ and $\hat{\lambda}_4$ for 100 trajectories, where $H = 0.6$, $x_0 = 0$, $T = 1000$, $h = 1$, $\sigma = 2$; the horizontal line shows the true value of the estimated parameter $\lambda = 0.5$.

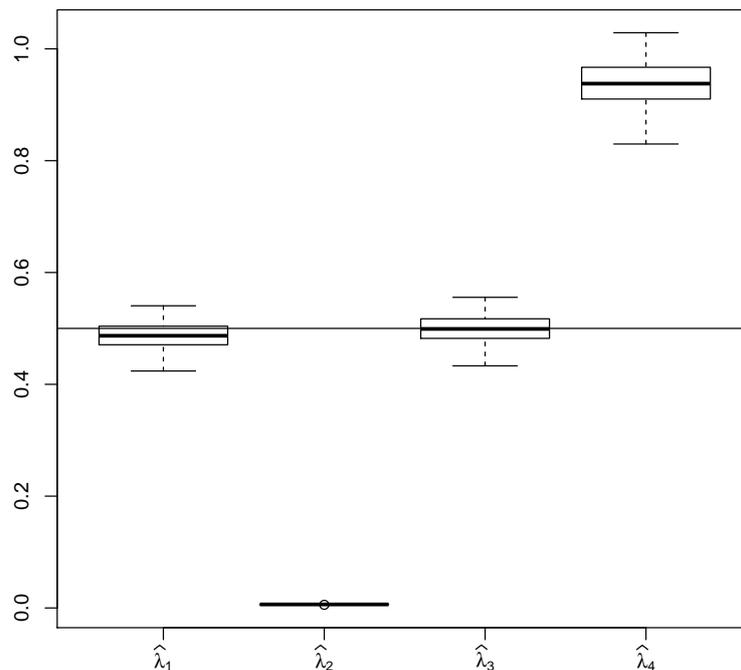


Figure 4. Comparison of $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3$ and $\hat{\lambda}_4$ for 100 trajectories, where $H = 0.6, x_0 = 100, T = 10, h = 0.1, \sigma = 2$; the horizontal line shows the true value of the estimated parameter $\lambda = 0.5$.

5. Conditional Least-Squares Estimator

Non-stationary trajectories with long time horizon contain a lot of information about λ , which is encoded mainly in two aspects: speed of decay in initial non-stationary phase and variance in stationary phase (see Figure 1). However, neither of the estimators $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3$ or $\hat{\lambda}_4$ can utilize all the information effectively. This motivates us to introduce another estimator. Recall that $\hat{\lambda}_3$ fails to be consistent because of bias in LSE caused by the correlation between X_t and ζ_t in Equation (11). To eliminate the correlation between explanatory variable and noise term in the linear model, we switch to conditional expectations. Start from the following equation, which defines η_t :

$$X_{t+h} = E[X_{t+h}|X_t] + \eta_t = E_{\Lambda=\lambda}[X_{t+h}|X_t] + \eta_t, \tag{18}$$

where λ is the true value of the unknown drift parameter and E_{Λ} the (conditional) expectation with respect to the measure generated by the fOU $\{X_t\}_{t \in [0, \infty)}$ with drift value Λ and initial condition x_0 . (Λ stands for an unknown throughout this section). In other words, $E_{\Lambda}[X_{t+h}|X_t]$ means the conditional expectation of X_{t+h} , conditioned by X_t , where the process X is given by (2) with drift $\lambda = \Lambda$. Hence, E_{Λ} has the same meaning as E in previous sections.

Obviously $E_{\Lambda}[\eta_t|X_t] = 0$ and, consequently, $c_t(X_t, \Lambda) = E_{\Lambda}[X_{t+h}|X_t]$ and η_t are uncorrelated. Indeed,

$$\begin{aligned} E_{\Lambda} \left((c_t(X_t, \Lambda) - E_{\Lambda}[c_t(X_t, \Lambda)])\eta_t \right) &= E_{\Lambda} \left(E_{\Lambda} \left[(c_t(X_t, \Lambda) - E_{\Lambda}[c_t(X_t, \Lambda)])\eta_t | X_t \right] \right) \\ &= E_{\Lambda} \left((c_t(X_t, \Lambda) - E_{\Lambda}[c_t(X_t, \Lambda)])E_{\Lambda} \left[\eta_t | X_t \right] \right) = 0. \end{aligned}$$

In result, we apply the least-squares technique to Equation (18), where λ is to be estimated, i.e., we would like to minimize

$$\min_{\Lambda} \sum_{n=0}^{N-1} \left(X_{(n+1)h} - E_{\Lambda}[X_{(n+1)h}|X_{nh}] \right)^2.$$

To calculate $c_t(X_t, \Lambda) = E_\Lambda[X_{t+h}|X_t]$ explicitly, use (11) and obtain

$$E_\Lambda[X_{t+h}|X_t] = e^{-\Lambda h} X_t + E_\Lambda[\xi_t|X_t]. \tag{19}$$

Note that random vector (ξ_t, X_t) has 2-dimensional normal distribution (dependent on parameter Λ)

$$\begin{bmatrix} \xi_t \\ X_t \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ e^{-\Lambda t} x_0 \end{bmatrix}, \begin{bmatrix} \sigma_\xi^2(\Lambda) & \sigma_{\xi, X}(\Lambda) \\ \sigma_{\xi, X}(\Lambda) & \sigma_X^2(\Lambda) \end{bmatrix} \right),$$

and we can use explicit expression for its conditional expectation to write

$$E_\Lambda[\xi_t|X_t] = \left(X_t - e^{-\Lambda t} x_0 \right) \frac{\sigma_{\xi, X}(\Lambda)}{\sigma_X^2(\Lambda)}. \tag{20}$$

With respect to the exact formula for X_t given by (2) and relation (3) we get

$$\begin{aligned} \sigma_X^2(\Lambda) &= \sigma^2 H(2H - 1) \int_0^t \int_0^t e^{\Lambda(u-t)} e^{\Lambda(v-t)} |u - v|^{2H-2} dv du \\ &= \frac{\sigma^2 H(2H - 1)}{\Lambda^{2H}} \int_{-\Lambda t}^0 \int_{-\Lambda t}^0 e^r e^s |r - s|^{2H-2} ds dr, \end{aligned}$$

where we used change-of-variable formula in the last step. Analogously

$$\begin{aligned} \sigma_{\xi, X}(\Lambda) &= \sigma^2 H(2H - 1) \int_t^{t+h} \left(\int_0^t e^{\Lambda(u-t-h)} e^{\Lambda(v-t)} (u - v)^{2H-2} dv \right) du \\ &= \frac{\sigma^2 H(2H - 1)}{\Lambda^{2H}} e^{-\Lambda h} \int_0^{\Lambda h} \left(\int_{-\Lambda t}^0 e^r e^s (r - s)^{2H-2} ds \right) dr. \end{aligned}$$

Using the expressions for $\sigma_{\xi, X}(\Lambda)$ and $\sigma_X^2(\Lambda)$ in (20) we obtain

$$E_\Lambda[\xi_t|X_t] = \left(X_t - e^{-\Lambda t} x_0 \right) e^{-\Lambda h} \frac{\int_0^{\Lambda h} \left(\int_{-\Lambda t}^0 e^r e^s (r - s)^{2H-2} ds \right) dr}{\int_{-\Lambda t}^0 \int_{-\Lambda t}^0 e^r e^s |r - s|^{2H-2} ds dr}. \tag{21}$$

Combining formula (21) with (19) yields

$$E_\Lambda[X_{t+h}|X_t] = c_t(X_t, \Lambda) = X_t A_{\Lambda t, \Lambda h} - B_{\Lambda t, \Lambda h},$$

with

$$A_{\tau, x} = e^{-x} \left(1 + \frac{\int_0^x \left(\int_{-\tau}^0 e^r e^s (r - s)^{2H-2} ds \right) dr}{\int_{-\tau}^0 \int_{-\tau}^0 e^r e^s |r - s|^{2H-2} ds dr} \right), \tag{22}$$

and

$$B_{\tau,x} = e^{-\tau} e^{-x} x_0 \frac{\int_0^x \left(\int_{-\tau}^0 e^r e^s (r-s)^{2H-2} ds \right) dr}{\int_{-\tau}^0 \int_0^0 e^r e^s |r-s|^{2H-2} ds dr}. \tag{23}$$

We can thus reformulate the Equation (18) for the observed process X as the following model (linear in X_t , but non-linear in Λ):

$$X_{t+h} = \left[X_t A_{\Lambda t, \Lambda h} - B_{\Lambda t, \Lambda h} \right]_{\Lambda=\lambda} + \eta_t. \tag{24}$$

Now we aim to apply the least-squares method to the reformulated model to get the conditional least-squares estimator $\hat{\lambda}_5$. To ensure the existence of global minima, we choose a closed interval $[\Lambda_L, \Lambda_U] \subset (0, \infty)$ and define $\hat{\lambda}_5$ as the minimizer of sum-of-squares function on this interval:

$$S_N(\hat{\lambda}_5) = \min_{\Lambda \in [\Lambda_L, \Lambda_U]} \{S_N(\Lambda)\}, \tag{25}$$

with criterion function S_N defined as

$$S_N(\Lambda) = \sum_{n=0}^{N-1} \left[X_{(n+1)h} - (X_{nh} A_{\Lambda nh, \Lambda h} - B_{\Lambda nh, \Lambda h}) \right]^2, \tag{26}$$

where we used (24) with $t = nh$.

Note that $S_N(\Lambda)$ is continuous in Λ and therefore a minimum on the compact interval $[\Lambda_L, \Lambda_U]$ exists. Although model (24) is linear in X_t , the coefficients A and B depend on t and that complicates the numerical minimization of S_N .

Remark 3. Let $\{Z_t\}_{t \in [0, \infty)}$ be the stationary solution to (1). Then

$$E_{\Lambda}[Z_{t+h}|Z_t] = Z_t A_{\infty, \Lambda h} + 0 = Z_t f(\Lambda h),$$

where f is defined in (13) and $\Lambda > 0$ is arbitrary. Since the coefficient $f(\Lambda h)$ does not depend on t , it is possible to calculate LSE for $f(\lambda h)$ explicitly and to construct the estimator of λ by applying f^{-1} . Such estimator coincides with $\hat{\lambda}_4$ introduced in previous chapter. Thus $\hat{\lambda}_4$ can be understood as the special case of conditional LSE for the stationary solution.

In order to prove strong consistency of the estimator $\hat{\lambda}_5$ we need to verify uniform convergence of $(1/N)S_N(\Lambda)$ to a function $S_{\infty}(\Lambda)$ specified below. Let us start with the following proposition on uniform convergence of $A_{\tau,x}$ and $B_{\tau,x}$. This proposition will help us in the sequel to investigate limiting behaviour of the two terms $A_{\Lambda nh, \Lambda h}$ and $B_{\Lambda nh, \Lambda h}$ in the sum-of-squares function S_N .

Proposition 1. Consider $A_{\tau,x}$ and $B_{\tau,x}$ defined by (22) and (23), and f defined by (13). Fix arbitrary $0 < x_L < x_U < \infty$. Then

$$\lim_{\tau \rightarrow \infty} \sup_{x \in [x_L, x_U]} |A_{\tau,x} - f(x)| = 0, \tag{27}$$

and

$$\lim_{\tau \rightarrow \infty} \sup_{x \in [x_L, x_U]} |B_{\tau,x}| = 0. \tag{28}$$

Proof. In order to simplify the notation, denote

$$I(\tau) = \int_{-\tau}^0 \int_{-\tau}^0 e^r e^s |r - s|^{2H-2} ds dr, \quad J(\tau) = \int_{-\tau}^0 e^s (-s)^{2H-2} ds,$$

and notice that

$$I(\tau) \xrightarrow{\tau \rightarrow \infty} \frac{\Gamma(2H)}{2H-1} = \Gamma(2H-1), \quad J(\tau) \xrightarrow{\tau \rightarrow \infty} \Gamma(2H-1).$$

Begin with (27):

$$\begin{aligned} & \sup_{x \in [x_L, x_U]} |A_{\tau, x} - f(x)| \\ &= \sup_{x \in [x_L, x_U]} e^{-x} \left| \frac{\int_0^x \left(\int_{-\tau}^0 e^r e^s (r-s)^{2H-2} ds \right) dr - \int_0^x \left(\int_{-\infty}^0 e^s e^r (r-s)^{2H-2} ds \right) dr}{\int_{-\tau}^0 \int_{-\tau}^0 e^r e^s |r-s|^{2H-2} ds dr} - \frac{\int_0^x \left(\int_{-\infty}^0 e^s e^r (r-s)^{2H-2} ds \right) dr}{\Gamma(2H-1)} \right| \\ &\leq e^{-x_L} \sup_{x \in [x_L, x_U]} \left| \frac{\int_0^x \left(\int_{-\tau}^0 e^r e^s (r-s)^{2H-2} ds \right) dr - \int_0^x \left(\int_{-\infty}^0 e^s e^r (r-s)^{2H-2} ds \right) dr}{I(\tau)} \right| \\ &\quad + e^{-x_L} \sup_{x \in [x_L, x_U]} \left| \frac{[\Gamma(2H-1) - I(\tau)] \int_0^x \left(\int_{-\infty}^0 e^s e^r (r-s)^{2H-2} ds \right) dr}{I(\tau)\Gamma(2H-1)} \right| \\ &\leq \frac{e^{-x_L}}{I(\tau)} \int_0^{x_U} e^r \left(\int_{-\infty}^{-\tau} e^s (r-s)^{2H-2} ds \right) dr \\ &\quad + \frac{e^{-x_L} |\Gamma(2H-1) - I(\tau)|}{I(\tau)\Gamma(2H-1)} \int_0^{x_U} e^r \left(\int_{-\infty}^0 e^s (r-s)^{2H-2} ds \right) dr \\ &\leq \frac{e^{-x_L} (\Gamma(2H-1) - J(\tau))}{I(\tau)} \int_0^{x_U} e^r dr + \frac{e^{-x_L} |\Gamma(2H-1) - I(\tau)|}{I(\tau)} \int_0^{x_U} e^r dr \xrightarrow{\tau \rightarrow \infty} 0. \end{aligned}$$

Similarly

$$\sup_{x \in [x_L, x_U]} |B_{\tau, x}| \leq e^{-\tau} e^{-x_L} x_0 \frac{J(\tau)}{I(\tau)} \int_0^{x_U} e^r dr \xrightarrow{\tau \rightarrow \infty} 0.$$

□

Choose any $0 < \Lambda_L < \Lambda_U < \infty$ and recall that $h > 0$ is fixed. The uniform convergences in (27) and (28) imply the following convergences uniformly in $\Lambda \in [\Lambda_L, \Lambda_U]$:

$$\lim_{n \rightarrow \infty} \sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |A_{\Lambda n h, \Lambda h} - f(\Lambda h)| = 0, \tag{29}$$

and

$$\lim_{n \rightarrow \infty} \sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |B_{\Lambda n h, \Lambda h}| = 0, \tag{30}$$

respectively. Indeed, set $x_L = \Lambda_L h$ and $x_U = \Lambda_U h$ and fix any $\varepsilon > 0$. There is $\tau_0 > 0$ such that for any $\tau > \tau_0$,

$$\sup_{x \in [x_L, x_U]} |A_{\tau, x} - f(x)| < \varepsilon.$$

If $n > \frac{\tau_0}{\Lambda_L h}$, then $\Lambda n h > \tau_0$ for any $\Lambda \in [\Lambda_L, \Lambda_U]$. Consequently

$$\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |A_{\Lambda n h, \Lambda h} - f(\Lambda h)| < \varepsilon,$$

which proves (29). The convergence in (30) can be shown analogously. These uniform convergences will be helpful in the proof of the following Lemma, which provides uniform convergence of $\frac{1}{N} S_N(\Lambda)$ to a limiting function $S_\infty(\Lambda)$. This uniform convergence is the key ingredient for the convergence of the minimizers $\hat{\lambda}_5$.

Lemma 3. *Let f be defined by (13) and let $S_N(\Lambda)$ be defined by (26), where $\{X_t\}_{t \in [0, \infty)}$ is the observed process with drift value λ . Denote*

$$S_\infty(\Lambda) = \frac{\sigma^2}{\lambda^{2H}} H\Gamma(2H) \left(1 - 2f(\Lambda h)f(\lambda h) + f^2(\Lambda h) \right). \tag{31}$$

Then

$$\lim_{N \rightarrow \infty} \sup_{\Lambda \in [\Lambda_L, \Lambda_U]} \left| \frac{1}{N} S_N(\Lambda) - S_\infty(\Lambda) \right| = 0 \quad \text{a.s.} \tag{32}$$

Proof. First consider the stationary solution $\{Z_t\}_{t \in [0, \infty)}$ to (1) corresponding to drift value λ . Comparison of (6) with (13) yields

$$E_\lambda Z_0 Z_h = f(\lambda h) \frac{\sigma^2}{\lambda^{2H}} H\Gamma(2H).$$

It enables us to write

$$S_\infty(\Lambda) = E_\lambda Z_0^2 - 2f(\Lambda h)E_\lambda Z_0 Z_h + f^2(\Lambda h)E_\lambda Z_0^2$$

for any $\Lambda > 0$, and, consequently

$$\begin{aligned} \frac{1}{N} S_N(\Lambda) - S_\infty(\Lambda) &= \left(\frac{1}{N} \sum_{n=0}^{N-1} X_{(n+1)h}^2 - E_\lambda Z_0^2 \right) \\ &\quad - 2 \left(\frac{1}{N} \sum_{n=0}^{N-1} A_{\Lambda n h, \Lambda h} X_{(n+1)h} X_{nh} - f(\Lambda h) E_\lambda Z_0 Z_h \right) \\ &\quad + \left(\frac{1}{N} \sum_{n=0}^{N-1} X_{nh}^2 A_{\Lambda n h, \Lambda h}^2 - f^2(\Lambda h) E_\lambda Z_0^2 \right) \\ &\quad + \left(\frac{1}{N} \sum_{n=0}^{N-1} B_{\Lambda n h, \Lambda h} \left(2X_{(n+1)h} - 2A_{\Lambda n h, \Lambda h} X_{nh} + B_{\Lambda n h, \Lambda h} \right) \right). \end{aligned} \tag{33}$$

Recall that $\{Z_t\}_{t \in [0, \infty)}$ is ergodic and $|Z_t - X_t|$ vanishes at infinity. Using Lemma 1 in the same way as in the proof of Theorem 1 implies

$$\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} \left| \frac{1}{N} \sum_{n=0}^{N-1} X_{(n+1)h}^2 - E_\lambda Z_0^2 \right| \xrightarrow{N \rightarrow \infty} 0 \quad \text{a.s.}$$

For the second term, write

$$\begin{aligned} & \sup_{\Lambda \in [\Lambda_L, \Lambda_U]} \left| \frac{1}{N} \sum_{n=0}^{N-1} A_{\Lambda nh, \Lambda h} X_{(n+1)h} X_{nh} - f(\Lambda h) E_\lambda Z_0 Z_h \right| \\ & \leq \sup_{\Lambda \in [\Lambda_L, \Lambda_U]} \left| \frac{1}{N} \sum_{n=0}^{N-1} (A_{\Lambda nh, \Lambda h} - f(\Lambda h)) X_{(n+1)h} X_{nh} \right| \\ & \quad + \frac{1}{N} \sum_{n=0}^{N-1} f(\Lambda h) \left(X_{(n+1)h} X_{nh} - E_\lambda Z_0 Z_h \right) \Big| \\ & \leq \frac{1}{N} \sum_{n=0}^{N-1} \left(\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |A_{\Lambda nh, \Lambda h} - f(\Lambda h)| \right) |X_{(n+1)h} X_{nh}| \\ & \quad + \left(\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |f(\Lambda h)| \right) \left| \frac{1}{N} \sum_{n=0}^{N-1} X_{(n+1)h} X_{nh} - E_\lambda Z_0 Z_h \right|. \end{aligned}$$

Application of Lemma 1, the convergence in (29) and the continuity of f ensure the convergence with probability one of both summands to zero as $N \rightarrow \infty$.

The uniform convergence of the third term can be shown analogously:

$$\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} \left| \frac{1}{N} \sum_{n=0}^{N-1} X_{nh}^2 A_{\Lambda nh, \Lambda h}^2 - f^2(\Lambda h) E_\lambda Z_0^2 \right| \xrightarrow{N \rightarrow \infty} 0 \quad \text{a.s.},$$

where we use

$$\lim_{n \rightarrow \infty} \sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |A_{\Lambda nh, \Lambda h}^2 - f^2(\Lambda h)| = 0,$$

which follows directly from (29) and the continuity of f .

The last term in (33) can be treated similarly:

$$\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} \left| \frac{1}{N} \sum_{n=0}^{N-1} B_{\Lambda nh, \Lambda h} \left(2X_{(n+1)h} - 2A_{\Lambda nh, \Lambda h} X_{nh} + B_{\Lambda nh, \Lambda h} \right) \right| \leq \frac{1}{N} \sum_{n=0}^{N-1} \left(\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |B_{\Lambda nh, \Lambda h}| \right) C_n,$$

where

$$\begin{aligned} C_n &= 2|X_{(n+1)h}| + 2 \left(\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |A_{\Lambda nh, \Lambda h} - f(\Lambda h)| \right) |X_{nh}| \\ & \quad + 2 \left(\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |f(\Lambda h)| \right) |X_{nh}| + \left(\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |B_{\Lambda nh, \Lambda h}| \right). \end{aligned}$$

By (30)

$$\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |B_{\Lambda nh, \Lambda h}| \xrightarrow{n \rightarrow \infty} 0,$$

and

$$\frac{1}{N} \sum_{n=0}^{N-1} C_n \xrightarrow{N \rightarrow \infty} E_\lambda |Z_0| \left(1 + 2 \sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |f(\Lambda h)| \right) < \infty \quad \text{a.s.}$$

Lemma 1 concludes the proof:

$$\frac{1}{N} \sum_{n=0}^{N-1} \left(\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} |B_{\Lambda nh, \Lambda h}| \right) C_n \xrightarrow{N \rightarrow \infty} 0 \quad \text{a.s.}$$

□

Previous considerations lead to the convergence of $\hat{\lambda}_5$, being the minimizers of $\frac{1}{N} S_N(\Lambda)$ to the minimizer of $S_\infty(\Lambda)$. Next lemma ensures that this minimizer coincides with the true drift value λ .

Lemma 4. S_∞ defined by (31) is continuous on $(0, \infty)$ and λ is the unique minimizer of S_∞ , i.e.,

$$S_\infty(\lambda) < S_\infty(\Lambda) \quad \forall \Lambda > 0, \Lambda \neq \lambda.$$

Proof. By definition

$$\begin{aligned} S_\infty(\Lambda) &= \frac{\sigma^2}{\lambda^{2H}} H\Gamma(2H) \left(1 - 2f(\Lambda h)f(\lambda h) + f^2(\Lambda h) \right) \\ &= \frac{\sigma^2}{\lambda^{2H}} H\Gamma(2H) \left(1 - f^2(\lambda h) + f^2(\lambda h) - 2f(\Lambda h)f(\lambda h) + f^2(\Lambda h) \right) \\ &= \frac{\sigma^2}{\lambda^{2H}} H\Gamma(2H) \left(1 - f^2(\lambda h) + [f(\lambda h) - f(\Lambda h)]^2 \right). \end{aligned}$$

The claim follows immediately, because f is one-to-one (it is strictly decreasing). Continuity of S_∞ is a direct consequence of the continuity of f . □

Now we are in a position to prove the strong consistency of $\hat{\lambda}_5$.

Theorem 3. Consider bounds $0 < \Lambda_L < \Lambda_U < \infty$ so that they cover the true drift λ of the observed solution $\{X_t\}_{t \in [0, \infty)}$ to Equation (1), i.e., $\lambda \in (\Lambda_L, \Lambda_U)$. Then $\hat{\lambda}_5$ defined in (25) is strongly consistent, i.e.,

$$\hat{\lambda}_5 \xrightarrow{N \rightarrow \infty} \lambda \quad \text{a.s.}$$

Proof. The proof follows standard argumentation from nonlinear regression and utilizes Lemma 3 and Lemma 4. Choose $\varepsilon > 0$ sufficiently small so that $[\Lambda_L, \Lambda_U] \setminus (\lambda - \varepsilon, \lambda + \varepsilon) \neq \emptyset$ and set

$$\delta = \min \left\{ S_\infty(\Lambda) - S_\infty(\lambda) : \Lambda \in [\Lambda_L, \Lambda_U] \setminus (\lambda - \varepsilon, \lambda + \varepsilon) \right\} > 0.$$

Consider a set of full measure on which the uniform convergence (32) holds and take $N_0 > 0$ such that

$$\sup_{\Lambda \in [\Lambda_L, \Lambda_U]} \left| \frac{1}{N} S_N(\Lambda) - S_\infty(\Lambda) \right| < \frac{\delta}{3} \quad \forall N \geq N_0.$$

Fix any $N \geq N_0$. Then for arbitrary $\Lambda \in [\Lambda_L, \Lambda_U] \setminus (\lambda - \varepsilon, \lambda + \varepsilon)$ we get

$$\frac{1}{N} S_N(\lambda) < S_\infty(\lambda) + \frac{\delta}{3} < S_\infty(\Lambda) - \frac{\delta}{3} < \frac{1}{N} S_N(\Lambda).$$

As $\hat{\lambda}_5$ minimizes S_N , for all $N \geq N_0$ we have

$$|\hat{\lambda}_5 - \lambda| < \varepsilon.$$

Since ϵ was arbitrary (if small enough), we obtain the convergence

$$\widehat{\lambda}_5 \xrightarrow{N \rightarrow \infty} \lambda$$

on a set of full measure. \square

6. Results and Discussion

In Table 1 we present comparison of the root mean square errors (RMSE) of all considered estimators for $\lambda = 0.5$ and several combinations of x_0 , T and H . Estimators $\widehat{\lambda}_1$ and $\widehat{\lambda}_3$ demonstrate good performance in scenarios with far-from-zero initial ($x_0 = 100$) condition and short time horizon ($T = 10$) This illustrates the fact that these estimators reflect mainly the speed of convergence to zero of the observed process in its initial phase. Increasing time horizon to $T = 1000$ adds a stationary phase to the observed trajectories, which distorts the estimators $\widehat{\lambda}_1$ and $\widehat{\lambda}_3$.

Table 1. Root mean square errors of the studied estimators calculated using 100 numerical simulations with $\lambda = 1/2$.

x_0	T	H	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\lambda}_3$	$\widehat{\lambda}_4$	$\widehat{\lambda}_5$
0	1000	0.6	0.216161	0.0364895	0.167108	0.0464931	0.0462385
0	1000	0.75	0.350571	0.0484904	0.33802	0.0637274	0.0643273
0	1000	0.9	0.444239	0.120954	0.442423	0.174213	0.170646
100	1000	0.6	0.158561	0.237689	0.0840923	0.125846	0.0297993
100	1000	0.75	0.244412	0.160116	0.205152	0.358587	0.0497393
100	1000	0.9	0.326662	0.105573	0.309256	1.17871	0.0454533
0	10	0.6	0.333834	0.513897	0.347278	0.539418	0.488992
0	10	0.75	0.439545	0.590848	0.438978	0.608744	0.552812
0	10	0.9	0.528586	0.841335	0.527932	0.840721	0.635887
100	10	0.6	0.0259539	0.49369	0.024092	0.441152	0.0277489
100	10	0.75	0.031949	0.480362	0.029853	1.53454	0.0411947
100	10	0.9	0.0398104	0.457155	0.0376366	4.54265	0.0291719

Estimators $\widehat{\lambda}_2$ and $\widehat{\lambda}_4$ perform well in settings with stationary-like initial condition ($x_0 = 0$) and long time horizon ($T = 1000$). This is because they are constructed from the stationary behavior of the process. Taking far-from-zero initial condition ruins these estimators, unless trajectory is very long.

The conditional LSE, $\widehat{\lambda}_5$, shows reasonable performance in all studied scenarios and it significantly outperforms the other estimators in scenario with far-from-stationary initial condition ($x_0 = 100$) and long time horizon ($T = 1000$). This results from the unique ability of this estimator to reflect and utilize information about the drift from both non-stationary (decreasing) phase and stationary (oscillating) phase. This is also illustrated on Figure 5. On the other hand, evaluation of $\widehat{\lambda}_5$ is the most numerically demanding compared the other studied estimators.

If $x_0 = 0$ and $T = 10$, $\widehat{\lambda}_5$ shows greater RMSE than $\widehat{\lambda}_1$ and $\widehat{\lambda}_3$ in Table 1 due to $\lambda = 1/2$ being relatively close to zero. This causes that $\widehat{\lambda}_1$ and $\widehat{\lambda}_3$ have smaller variance (although greater bias) compared to $\widehat{\lambda}_5$ (see Figure 6). In order to present this effect we have calculated RMSE for simulations in same scenario but with $\lambda = 3/2$ (see Table 2). $\widehat{\lambda}_5$ provides smaller RMSE than the other estimators in this setting.

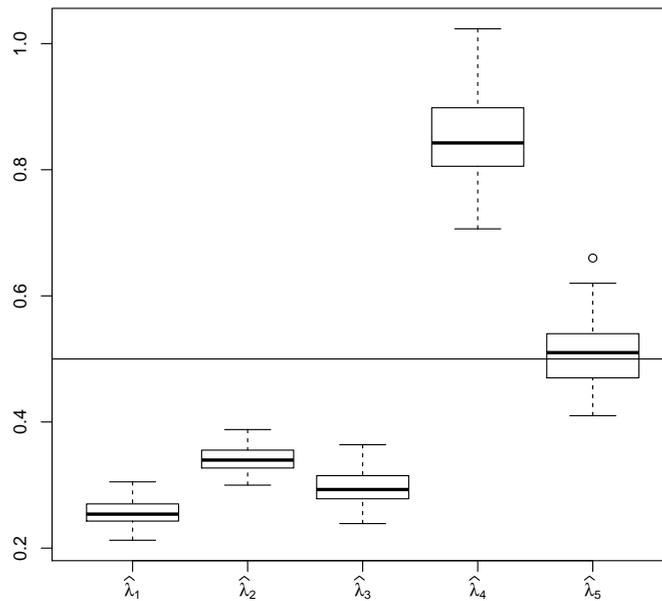


Figure 5. Comparison of $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\lambda}_4$ and $\hat{\lambda}_5$ for 100 trajectories, where $H = 0.75, x_0 = 100, T = 1000, h = 1, \sigma = 2$; the horizontal line shows the true value of the estimated parameter $\lambda = 0.5$.

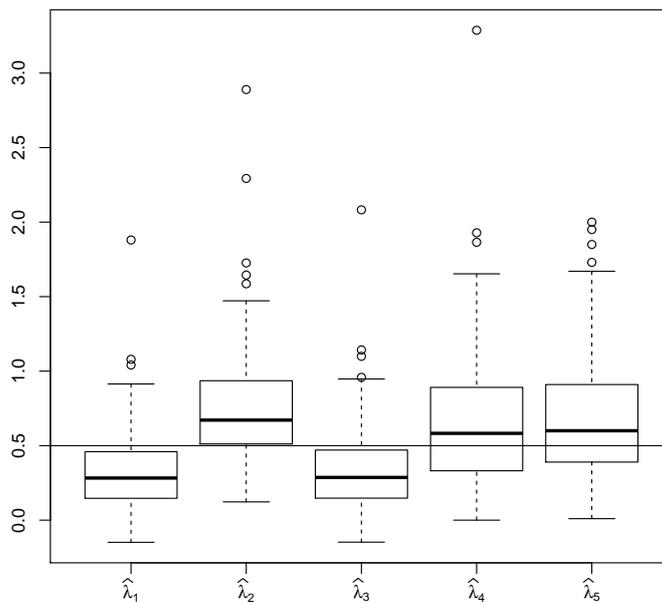


Figure 6. Comparison of $\hat{\lambda}_1, \hat{\lambda}_2, \hat{\lambda}_3, \hat{\lambda}_4$ and $\hat{\lambda}_5$ for 100 trajectories, where $H = 0.6, x_0 = 0, T = 10, h = 0.1, \sigma = 2$; the horizontal line shows the true value of the estimated parameter $\lambda = 0.5$.

Table 2. Root mean square errors of the studied estimators calculated using 100 numerical simulations with $\lambda = 3/2$.

x_0	T	H	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\lambda}_4$	$\hat{\lambda}_5$
0	10	0.6	0.68946	0.641613	0.686183	0.761677	0.399386
0	10	0.75	1.11793	0.761793	1.10949	0.877707	0.457095
0	10	0.9	1.38463	1.21928	1.38299	1.52812	0.504919

7. Conclusions

Three new estimators were defined and studied:

1. The least-squares estimator from exact solution ($\hat{\lambda}_3$), which improves the popular discretized LSE ($\hat{\lambda}_1$) by eliminating the discretization error. It is easy to implement, since it can be calculated by a closed formula. However, it fails to be strongly consistent in long-span regime.
2. The asymptotic least-squares estimator ($\hat{\lambda}_4$), which is a modification of $\hat{\lambda}_3$ with respect to its asymptotic behavior. In result, $\hat{\lambda}_4$ is strongly consistent in the long-span regime and behaves similarly to the well-established discrete ergodic estimator ($\hat{\lambda}_2$). The advantage of $\hat{\lambda}_4$ is that it does not require a priori knowledge of the volatility σ . On the other hand, its implementation includes a root-finding numerical procedure.
3. The conditional least-squares estimator ($\hat{\lambda}_5$), which eliminates the bias in the least-squares procedure by considering the conditional expectation of the response as the explanatory variable. The possibility to express the conditional expectation explicitly makes this approach feasible. This conditioning idea (which is new in the context of the models with fractional noise, to our best knowledge) provides exceptionally reliable estimator, which outperforms all the other studied estimators. We proved the strong consistency (in long-span regime) of this estimator. The implementation comprises solving an optimization problem.

These new estimating procedures can help practitioners or scientists from various fields to improve the calibration of their models based on available data with autocorrelated noise (these are typically observed/measured in discrete time instants) and, consequently, obtain more reliable conclusions from the calibrated models.

An interesting future extension would certainly be to explore the potential of the promising idea of the conditioning within least-squares procedure to more general models and settings (including d-dimensional fOU, fOU with $H < 1/2$, non-linear drift, multiplicative noise, etc.).

Author Contributions: Conceptualization, P.K.; methodology, P.K.; software, L.S.; writing—original draft, P.K. and L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the grant LTAIN19007 Development of Advanced Computational Algorithms for Evaluating Post-surgery Rehabilitation.

Acknowledgments: We are grateful to four anonymous reviewers for their valuable comments, which helped improve this paper significantly.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mishura, Y. *Stochastic Calculus for Fractional Brownian Motion and Related Processes*; Springer: Berlin/Heidelberg, Germany, 2008.
2. Biagini, F.; Hu, Y.; Øksendal, B.; Zhang, T. *Stochastic Calculus for Fractional Brownian Motion and Applications*; Springer: London, UK, 2008.
3. Abundo, M.; Pirozzi, E. On the Integral of the Fractional Brownian Motion and Some Pseudo-Fractional Gaussian Processes. *Mathematics* **2019**, *7*, 991. [[CrossRef](#)]
4. Ascione, G.; Mishura, Y.; Pirozzi, E. Fractional Ornstein–Uhlenbeck Process with Stochastic Forcing, and its Applications. *Methodol. Comput. Appl. Probab.* **2019**. [[CrossRef](#)]
5. Hu, Y.; Nualart, D. Parameter estimation for fractional Ornstein–Uhlenbeck processes. *Stat. Probab. Lett.* **2010**, *80*, 1030–1038. [[CrossRef](#)]
6. Hu, Y.; Nualart, D.; Zhou, H. Parameter estimation for fractional Ornstein–Uhlenbeck processes of general Hurst parameter. *Stat. Inference Stoch. Process.* **2019**, *22*, 111–142. [[CrossRef](#)]
7. Es-Sebaiy, K. Berry-Esseen bounds for the least squares estimator for discretely observed fractional Ornstein–Uhlenbeck processes. *Stat. Probab. Lett.* **2013**, *83*, 2372–2385. [[CrossRef](#)]

8. Kubilius, K.; Mishura, Y.; Ralchenko, K.; Seleznev, O. Consistency of the drift parameter estimator for the discretized fractional Ornstein–Uhlenbeck process with Hurst index H is an element of $(0,1/2)$. *Electron. J. Stat.* **2015**, *9*, 1799–1825. [[CrossRef](#)]
9. Hu, Y.; Song, J. Parameter estimation for fractional Ornstein–Uhlenbeck processes with discrete observations. In *Malliavin Calculus and Stochastic Analysis*; Springer: Boston, MA, USA, 2013; Volume 34, pp. 427–442. [[CrossRef](#)]
10. Es-Sebaiy, K.; Viens, F. Optimal rates for parameter estimation of stationary Gaussian processes. *Stoch. Process. Their. Appl.* **2019**, *129*, 3018–3054. [[CrossRef](#)]
11. Azmoodeh, E.; Viitasaari, L. Parameter estimation based on discrete observations of fractional Ornstein–Uhlenbeck process of the second kind. *Stat. Inference Stoch. Process.* **2015**, *18*, 205–227. [[CrossRef](#)]
12. Neuenkirch, A.; Tindel, S. A least square-type procedure for parameter estimation in stochastic differential equations with additive fractional noise. *Stat. Inference Stoch. Process.* **2014**, *17*, 99–120. [[CrossRef](#)]
13. Xiao, W.; Zhang, W.; Xu, W. Parameter estimation for fractional Ornstein–Uhlenbeck processes at discrete observation. *Appl. Math. Model.* **2011**, *35*, 4196–4207. [[CrossRef](#)]
14. Istas, J.; Lang, G. Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Annales de l'I.H.P. Probabilités et Statistiques* **1997**, *33*, 407–436. [[CrossRef](#)]
15. Coeurjolly, J. Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles. *Ann. Stat.* **2008**, *36*, 1404–1434. [[CrossRef](#)]
16. Rosenbaum, M. Estimation of the volatility persistence in a discretely observed diffusion model. *Stoch. Process. Their. Appl.* **2008**, *118*, 1434–1462. [[CrossRef](#)]
17. Berzin, C.; Latour, A.; León, J. *Inference on the Hurst Parameter and Variance of Diffusions Driven by Fractional Brownian Motion*; Springer International Publishing: Cham, Switzerland, 2014.
18. Brouste, A.; Iacus, S. Parameter estimation for the discretely observed fractional Ornstein–Uhlenbeck process and the Yuima R package. *Comput. Stat.* **2013**, *28*, 1529–1547. [[CrossRef](#)]
19. Brouste, A.; Fukasawa, M.; Hino, H.; Iacus, S.; Kamatani, K.; Koike, Y.; Masuda, H.; Nomura, R.; Ogihara, T.; Shimuzu, Y.; et al. The YUIMA Project: A Computational Framework for Simulation and Inference of Stochastic Differential Equations. *J. Stat. Softw.* **2014**, *4*, 1–51. [[CrossRef](#)]
20. Kubilius, K.; Mishura, Y.; Ralchenko, K. *Parameter Estimation in Fractional Diffusion Models*; Springer International Publishing AG: Cham, Switzerland, 2017.
21. Pipiras, V.; Taqqu, M. Integration questions related to fractional Brownian motion. *Probab. Theory Relat. Fields* **2000**, *118*, 251–291. [[CrossRef](#)]
22. Kříž, P.; Maslowski, B. Central limit theorems and minimum-contrast estimators for linear stochastic evolution equations. *Stochastics* **2019**, *91*, 1109–1140. [[CrossRef](#)]

