mathematics

MDPI

Article

The Combined Estimator for Stochastic Equations on Graphs with Fractional Noise

Pavel Kříž * and Leszek Szała

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.: +420-220443160

Received: 10 September 2020; Accepted: 6 October 2020; Published: 13 October 2020

Abstract: In the present paper, we study the problem of estimating a drift parameter in stochastic evolution equations on graphs. We focus on equations driven by fractional Brownian motions, which are particularly useful e.g., in biology or neuroscience. We derive a novel estimator (the combined estimator) and prove its strong consistency in the long-span asymptotic regime with a discrete-time sampling scheme. The promising performance of the combined estimator for finite samples is examined under various scenarios by Monte Carlo simulations.

Keywords: stochastic equations on graphs; fractional Brownian motion; parameter estimation

MSC: 2010 60G22; 62M09

1. Introduction

Evolution equations on graphs and networks have proven to be a useful tool for modeling dynamics of various phenomena in many diverse fields, such as biology, engineering, statistical physics, social sciences, economics etc., for several applications of such models (such as sensor networks, nanosystems, multiagent coordination, social networks etc.); see e.g., [1] and references therein. Many real-world phenomena are affected by random effects (random forcing). These effects can be included in an evolution equation as a stochastic noise term. Evolution equations on graphs with stochastic noise have been studied e.g., in [1]. A direct application in neurobiology motivated the study of a specific type of dynamical models on graphs with fractional noise (the noise generated by a fractional Brownian motion) presented in [2]. Fractional Brownian motion has become a popular tool to model random noise with memory (with auto-correlation), which is desirable in many diverse fields (see e.g., monographs [3,4] for discussion).

Another motivation for studying stochastic evolution equations on graphs comes from the discretization of parabolic stochastic partial differential equations (SPDEs) in space domain, where the selected spatial discretization method determines the evolution operator in the equation (cf. [5]).

In this paper, we consider the problem of estimating an unknown drift parameter in stochastic evolution equations (sEE) on graphs with fractional noise. Estimating a drift parameter in various stochastic differential equations has been considered by many authors. A great amount of literature has been devoted to equations with white noise (generated by a standard Brownian motion with uncorrelated increments), where the maximum-likelihood approach dominates; see for example the classical monograph [6] for 1-dimensional equations or the recent overview paper [7] for equations in ∞-dimensional spaces, to name just a few. Specific problems arise in equations with fractional noise (generated by a fractional Brownian motion with autocorrelated increments); see e.g., books [8,9] for discussion of various approaches to parameter estimation in this setting. We are not aware of any study on parameter estimation in the context of sEE on graphs with fractional noise, although it may be of a

great interest for both researchers and practitioners, as it corresponds to inferring certain properties of a random dynamical system on a graph based on observing its evolution in time. This paper is supposed to bridge this gap.

Our paper is organized as follows. In Section 2, we specify the basic setting of the estimation problem. In Section 3, we recall the definition of a fractional Brownian motion, its basic properties and calculus. Moreover, we provide a short overview of the concepts of mild solutions to stochastic evolution equations in both physical and frequency domain. Section 4 is devoted to the introduction, calculation and strong consistency of the combined estimator—the main contribution of this paper. Illustration of the actual performance of the combined estimator in several scenarios is presented in Section 5 with Monte Carlo simulations. Main findings and results of this paper are summarized in Section 6.

2. Problem Setup

Consider an undirected, connected, finite graph $G = (V, E)$ with a set of vertices $V = \{v_1, \ldots, v_N\}$ and a set of edges E. Denote by $\mathbb{R}^V$ a finite-dimensional Hilbert space of the real-valued functions on V with the canonical scalar product defined by $\langle x, y \rangle = \sum_{v \in V} x(v)y(v)$. Note that the indicator functions $e_v = 1_v$ for $v \in V$ form the canonical orthonormal basis of $\mathbb{R}^V$. Recall that the combinatorial graph Laplacian L on a graph G is a self-adjoint, positive semidefinite, linear operator on $\mathbb{R}^V$, characterized by the following matrix:

$$L = [l_{v,w}]_{v,w \in V}; \quad l_{v,w} = \begin{cases} \deg(v) & \text{if } v = w \\ -1 & \text{if } \{v, w\} \in E \\ 0 & \text{else,} \end{cases} \quad \forall v, w \in V, \tag{1}$$

where $\deg(v)$ is number of edges connected to v (the degree of v).

As a motivation (and a toy model) for further studies, consider a simple heat equation on the graph G with stochastic forcing,

$$\frac{\partial x}{\partial t}(t, v) = -\lambda(Lx(t))(v) + \eta(t, v), \quad \text{for } t \in \mathbb{R}_+ \text{ and } v \in V, \tag{2}$$

$$x(0, .) \in \mathbb{R}^V, \tag{3}$$

where $x(t, v)$ is a real-valued state variable at vertex v and time t, $\lambda \in \mathbb{R}_+$ is an unknown parameter (heat conductivity), L is the combinatorial graph Laplacian and $\eta(t, v)$ represents a noise (stochastic) term, being a formal time-derivative of a (fractional) Brownian motion.

Since the classical (strong) derivative of a (fractional) Brownian motion does not exist, we reformulate the above stochastic heat equation rigorously as a sEE on the finite-dimensional Hilbert space $\mathbb{R}^V$:

$$\begin{aligned} dX_t &= -\lambda L X_t dt + \Phi dB_t^H, \\ X_0 &= x(0, .) \in \mathbb{R}^V, \end{aligned} \tag{4}$$

where $\{X_t\}_{t \geq 0} = \{x(t, .)\}_{t \geq 0}$ is an $\mathbb{R}^V$-valued random process, $\lambda \in \mathbb{R}_+$ is an unknown drift parameter and L is the combinatorial graph Laplacian. For the noise (stochastic) part, B^H denotes the cylindrical fractional Brownian motion on $\mathbb{R}^V$ with Hurst index H and Φ, being a bounded operator on $\mathbb{R}^V$, determines the correlations between vertices.

Motivated by the above example, we consider the parameter estimation problem in an evolution equation on $\mathbb{R}^V$ of the following form in this paper:

$$\begin{aligned} dX_t &= -A(\lambda) X_t dt + \Phi dB_t^H, \\ X_0 &= x(0, .) \in \mathbb{R}^V, \end{aligned} \tag{5}$$

where the drift operator $A(\lambda)$ is a linear, self-adjoint, positive semidefinite operator (matrix) on $\mathbb{R}^V$. It depends on an unknown (potentially d-dimensional) parameter $\lambda \in C \subset \mathbb{R}^d$, with C being compact. In the context of graph theory, the operator A may represent (weighted) graph Laplacian, in-degree Laplacian (for oriented graphs), d-path Laplacian (for non-local interactions, see [10]) or other linear operators. B^H is again the cylindrical fractional Brownian motion on $\mathbb{R}^V$ with Hurst index $H \geq 1/2$ (the regular case with positively auto-correlated increments) and Φ a bounded linear operator on $\mathbb{R}^V$.

We address the problem of estimating the unknown drift parameter λ based on a single trajectory of a solution to evolution Equation (5) observed at discrete time instants with a time step $h > 0$, i.e., based on $\{X_{ih}\}_{i=0,\ldots n}$. We assume that the time step $h > 0$ is fixed and the time horizon goes to infinity, $T = nh \to \infty$, i.e., the discrete sampling scheme with long-span asymptotics. Literature on drift estimation for multidimensional sEE with the discrete sampling scheme and long-span asymptotics is rather sparse. In this respect, let us mention papers [11–13] considering SPDEs with white-in-time noise, and paper [14] focused on infinite-dimensional sEE with fractional noise. The so-called minimum-contrast estimator (MCE) studied in the latter paper is plausible in our setting, if the projection of a solution to the eigenspaces corresponding to strictly positive eigenvalues is considered (to ensure ergodicity). However, the comparison in the recent paper [15] shows that the ergodic estimator (the counterpart to the MCE in one-dimensional setting) provides less accurate estimates compared to the conditional least-squares estimator (introduced in [15]), which turned out to be a very powerful estimator of the unknown drift in models with autocorrelated noise. Hence, we generalize the idea of the conditional least-squares approach into the multi-dimensional framework. To enhance the efficiency of the estimator, we combine the conditional least-squares approach in the time domain (where the covariance structure is complicated) with a maximum-likelihood approach in the space domain (with rather simple covariance structure), which leads to a novel estimator—the combined estimator (CoE).

The value of the Hurst index H is considered to be known. If this is not the case, it can be estimated in advance (before an evaluation of the combined estimator) by some of many available methods—see for example [16–18] to name just a few. The estimated H can subsequently be used in the CoE of the drift parameter similarly to the approach applied in [19]. The simultaneous estimation of λ and H in sEE on graphs is outside the scope of this work and could be a potential direction for further research.

3. Preliminaries

3.1. Fractional Brownian Motion and Its Calculus

In this subsection, we briefly introduce the fractional Brownian motion (fBm) and specifics of the stochastic calculus w.r.t. the fBm. For further reading, we refer the interested reader to the monographs [3,4] and references therein.

The one-dimensional fractional Brownian motion $\{\beta^H(t)\}_{t\geq 0}$ is a centered continuous real-valued Gaussian process starting from zero and having the following covariance structure

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

The covariance depends on a parameter $H \in (0,1)$, which is called the Hurst index. If $H = 1/2$, fBm has independent increments and it coincides with a standard Brownian motion (the Wiener process). If $H > 1/2$, fBm has positively correlated increments and it exhibits a long-range dependence. It enjoys increasing popularity in many applications where long-range dependence is observed. It also has more regular trajectories than the Wiener process (in terms of the Hölder continuity). The case $H < 1/2$ (not considered in this work) leads to negatively correlated increments, less regular

trajectories and serious difficulties when integrating w.r.t. the fBm. A two-sided fBm is again a centered continuous Gaussian process $\{\beta^H(t)\}_{t\in\mathbb{R}}$ with t ranging over the whole $\mathbb{R}$, $\beta^H(0) = 0$ and

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

The N-dimensional fBm $\{B^H(t)\}_{t\geq 0} = \{(\beta_1^H(t), \ldots, \beta_N^H(t))^T\}_{t\geq 0}$ is an N-dimensional random process whose coordinates are mutually independent real-valued fBm with the same Hurst index H.

Consider a fBm $\{\beta^H(t)\}_{t\geq 0}$ with $H > 1/2$ and a deterministic step function $\psi(s) = \sum_{k=1}^{M} c_k I_{[t_k,t_{k+1}]}(s)$. The Wiener integral of ψ w.r.t. β^H is defined by formula

$$\int_0^\infty \psi(s)d\beta^H(s) = \sum_{k=1}^{M} c_k \left(\beta^H(t_{k+1}) - \beta^H(t_k)\right).$$

As a consequence, we obtain the following isometry for any pair of deterministic step functions ψ and ϕ:

$$\mathrm{E}\left(\int_0^\infty \psi(t)d\beta^H(t)\int_0^\infty \phi(s)d\beta^H(s)\right) = \int_0^\infty\int_0^\infty \psi(t)\phi(s)H(2H-1)|t-s|^{2H-2}\mathrm{d}t\mathrm{d}s = \langle\psi,\phi\rangle. \tag{6}$$

This isometry enables us to extend the definition of the Wiener integral w.r.t. a fBm to all elements of the completion of the space of deterministic step functions w.r.t. the scalar product $\langle .,. \rangle$ defined above. In result, Formula (6) holds true for all functions from this completion, see also [20]. This formula is extremely useful for the construction of the combined estimator below.

3.2. Solution to sEE in Physical and Frequency Domain

For the reader's convenience, we recall the concept of mild solution to (5) in physical domain. It reads:

$$X_t = e^{-A(\lambda)t}X_0 + \int_0^t e^{-A(\lambda)(t-s)}\Phi dB_s^H, \quad \text{where } e^{-A(\lambda)t} = \sum_{j=0}^{\infty}\frac{[-A(\lambda)]^j t^j}{j!}. \tag{7}$$

The existence of the mild solution was established (in more general setting of sEE on Hilbert spaces) in papers [21] for $H > 1/2$ and [22] for $H < 1/2$, or in paper [23].

When studying the solution, we can switch from the physical to the frequency domain. It provides us with an easy-to-handle formula of the solution, which will be useful both for the simulation and parameter estimation in the sequel. Start with the spectral decomposition of the self-adjoint, positive semidefinite matrix $A(\lambda)$:

$$A(\lambda) = G(\lambda)D(\lambda)G(\lambda)^T, \tag{8}$$

where $G(\lambda)$ is a $N \times N$ matrix whose columns are orthonormal eigenvectors $\{g_k(\lambda)\}_{k=1\ldots N}$ of $A(\lambda)$ and $D(\lambda) = \mathrm{diag}\{\alpha_1(\lambda), \ldots, \alpha_N(\lambda)\}$ is a diagonal matrix formed by the corresponding non-negative eigenvalues. It easily follows:

$$e^{-A(\lambda)t} = G(\lambda)\mathcal{E}_t(\lambda)G(\lambda)^T, \quad \text{where } \mathcal{E}_t(\lambda) = \mathrm{diag}\{e^{-\alpha_1(\lambda)t}, \ldots, e^{-\alpha_N(\lambda)t}\}. \tag{9}$$

With this decomposition at hand, we can consider the solution to (5) in frequency domain:

$$F_t = G(\lambda)^T X_t, \tag{10}$$

which is given by the mild formula

$$F_t = \mathcal{E}_t(\lambda)F_0 + \int_0^t \mathcal{E}_{t-s}(\lambda)d\tilde{B}_s^H, \tag{11}$$

where

$$\tilde{B}_t^H = G(\lambda)^T \Phi B_t^H, \quad t \geq 0. \tag{12}$$

Since this system is diagonal, we can write explicit formulas for individual entries of F_t. Denote

$$F_t = (f_1(t), \ldots, f_N(t))^T \quad \text{and } \tilde{B}_t^H = (\tilde{\beta}_1^H(t), \ldots, \tilde{\beta}_N^H(t))^T.$$

Then $\{f_k(t)\}_{t\geq 0}$ are real-valued fractional Ornstein-Uhlenbeck processes, as seen from the formulas

$$f_k(t) = e^{-\alpha_k(\lambda)t} f_k(0) + \int_0^t e^{-\alpha_k(\lambda)(t-s)} d\tilde{\beta}_k^H(s), \quad k = 1, \ldots, N. \tag{13}$$

To eliminate the stochastic Wiener integral from the expression above, we can apply the integration by parts formula to get

$$f_k(t) = e^{-\alpha_k(\lambda)t} f_k(0) + \tilde{\beta}_k^H(t) - \int_0^t \tilde{\beta}_k^H(s) e^{-\alpha_k(\lambda)(t-s)} \alpha_k(\lambda) ds, \quad k = 1, \ldots, N. \tag{14}$$

The solution formula above is particularly useful for simulations, because it contains only Riemann integral and no stochastic integration is needed.

4. Combined Estimator

To estimate the unknown parameter λ in (5), we observe a single realization of the process $\{X_{ih}\}_{i=0,\ldots n}$. It is an $(N \times (n+1))$-dimensional Gaussian vector, so a natural choice would be to take the maximum-likelihood estimator. However, the evaluation of such an estimator would be extremely consuming, both in terms of CPU time and memory, especially for large n. This is due to the complicated structure of the covariance matrix with respect to time domain $i = 0, \ldots n$ resulting from the long-range dependence of the fractional Brownian motion with $H > 1/2$. Hence, some approximation to MLE has to be considered (cf. e.g., [24]). In this section, we study an approximation based on the conditional least-squares technique. This approach has recently been introduced and studied (with promising results) in [15] for real-valued fractional Ornstein-Uhlenbeck processes.

4.1. Combined Estimator in Physical Domain

The main idea of the conditional least-squares technique is to consider the following linear model

$$X_{t+h} = \mathrm{E}_\lambda[X_{t+h}|X_t] + \eta_t = \mu_t(X_t, \lambda) + \eta_t, \tag{15}$$

where $\mathrm{E}_\lambda[X_{t+h}|X_t]$ means the conditional expectation of X_{t+h}, conditioned by X_t, where the process X is given by (5) with parameter value λ. Conditioning eliminates the unwanted correlations between explanatory variables and error terms in model (15), which enables the application of the least-squares technique.

To design the ultimate form of the least-squares criterion to be minimized, we have to decide about the space-norm to be applied to the vectors $X_{t+h} - \mathrm{E}_\lambda[X_{t+h}|X_t]$. In this aspect, we stick to the maximum-likelihood technique, which is well suited e.g., for the white-in-space noise (Φ is diagonal). Notice the N-dimensional Gaussian conditional distribution

$$\mathrm{Law}_\lambda(X_{t+h}|X_t) = \mathcal{N}\left(\mu_t(X_t, \lambda), \Sigma_t(\lambda)\right), \tag{16}$$

and the corresponding log-likelihood function (for a fixed t):

$$l_t(X_{t+h}, \lambda | X_t) = Const - \frac{1}{2}\log\left(\det[\Sigma_t(\lambda)]\right) - \frac{1}{2}\left(X_{t+h} - \mu_t(X_t, \lambda)\right)^T [\Sigma_t(\lambda)]^{-1}\left(X_{t+h} - \mu_t(X_t, \lambda)\right).$$

To express μ_t and Σ_t explicitly, recall

$$X_{t+h} = e^{-A(\lambda)h} X_t + \xi_t,$$

where

$$X_t = e^{-A(\lambda)t} X_0 + \int_0^t e^{-A(\lambda)(t-s)} \Phi dB^H(s), \quad \xi_t = \int_t^{t+h} e^{-A(\lambda)(t+h-s)} \Phi dB^H(s).$$

Now consider the more complicated case $H > 1/2$ (the simpler case $H = 1/2$ is discussed in Remark 1). Utilizing isometry (6), we get the (co)variance matrices

$$\mathrm{var}(X_t) = H(2H-1)\int_{-t}^0 \int_{-t}^0 e^{A(\lambda)r} \Phi\Phi^T e^{A(\lambda)s} |r-s|^{2H-2} ds dr, \tag{17}$$

$$\mathrm{var}(\xi_t) = H(2H-1)\int_{-h}^0 \int_{-h}^0 e^{A(\lambda)r} \Phi\Phi^T e^{A(\lambda)s} |r-s|^{2H-2} ds dr, \tag{18}$$

$$\mathrm{cov}(\xi_t, X_t) = H(2H-1) e^{-A(\lambda)h} \int_0^h \int_{-t}^0 e^{A(\lambda)r} \Phi\Phi^T e^{A(\lambda)s} |r-s|^{2H-2} ds dr. \tag{19}$$

The joint multivariate normality of X_t and ξ_t leads to

$$\mu_t(X_t, \lambda) = \mathrm{E}_\lambda[X_{t+h}|X_t] = \begin{cases} e^{-A(\lambda)h} X_t + \mathrm{cov}(\xi_t, X_t)[\mathrm{var}(X_t)]^{-1}(X_t - e^{-A(\lambda)t} X_0) & \text{for } t > 0 \\ e^{-A(\lambda)h} X_t & \text{for } t = 0, \end{cases} \tag{20}$$

$$\Sigma_t(\lambda) = \mathrm{var}_\lambda[X_{t+h}|X_t] = \begin{cases} \mathrm{var}(\xi_t) - \mathrm{cov}(\xi_t, X_t)[\mathrm{var}(X_t)]^{-1}[\mathrm{cov}(\xi_t, X_t)]^T & \text{for } t > 0 \\ \mathrm{var}(\xi_t) & \text{for } t = 0. \end{cases} \tag{21}$$

The combination of the conditional least-squares approach in the time domain with the maximum-likelihood approach in the space domain results in the following combined estimator:

$$\hat{\lambda} = \underset{\Lambda \in C}{\mathrm{argmin}}\, Q(\Lambda), \tag{22}$$

where C is a compact subset of $\mathbb{R}^d$, Λ is an unknown representing the parameter to be estimated (λ is its true value) and

$$Q(\Lambda) = \sum_{t \in \{0, h \ldots, (n-1)h\}} \left[\log\left(\det[\Sigma_t(\Lambda)]\right) + \left(X_{t+h} - \mu_t(X_t, \Lambda)\right)^T [\Sigma_t(\Lambda)]^{-1}\left(X_{t+h} - \mu_t(X_t, \Lambda)\right)\right], \tag{23}$$

with μ_t and Σ_t explicitly expressed in (20) and (21), respectively.

Remark 1. *If $H = 1/2$, which corresponds to the white-in-time noise, the combined estimator coincides with classical MLE. Indeed, X_t and ξ_t are independent in this case, the Gaussian process $\{X_{ih}\}_{i=0,\ldots,n}$ has Markov property and its joint distribution corresponds to the product of the conditional distributions (16). As a result, minimizing (22) is equivalent to minimizing log-likelihood function of the process. Moreover, we have simplified calculations*

$$\mu_t(X_t, \lambda) = e^{-A(\lambda)h} X_t,$$
$$\Sigma_t(\lambda) = \mathrm{var}(\xi_t) = \int_{-h}^0 e^{A(\lambda)s} \Phi\Phi^T e^{A(\lambda)s} ds.$$

4.2. Combined Estimator in Diagonal Case

If the number of vertices N is large, evaluation of the criterion function in Formula (22) might be computationally demanding, mainly due to inversions of $N \times N$ matrices $[\mathrm{var}(X_t)]^{-1}$ and $[\Sigma_t(\Lambda)]^{-1}$ for all $t \in \{h, 2h \ldots, nh\}$. This issue can be resolved by considering the whole problem in the spectral domain in the diagonal case. For this reason, we assume throughout this subsection that the noise term is white in space, i.e., Φ is an identity matrix (possibly multiplied by a constant characterizing noise intensity):

Assumption 1 (WN).

$$\Phi = \sigma \,\mathrm{Id}, \quad \sigma > 0. \tag{24}$$

Recall the spectral decomposition of $A(\lambda)$ and $e^{-A(\lambda)t}$ given in (8) and (9), respectively. In addition, assume that the eigenvectors of operator $A(\lambda)$ do not depend on the parameter λ, i.e.,

Assumption 2 (EV).

$$G(\lambda) = G. \tag{25}$$

Such an assumption is satisfied if the unknown parameter λ affects only the eigenvalues of $A(\lambda)$, for example, if λ is a multiplicative constant (diffusivity or heat conductivity): $A(\lambda) = \lambda A$.

To simplify the exposition, sort the eigenvalues in descendent order, so that the first M eigenvalues are positive and the others are zeros:

$$\alpha_1(\lambda) \geq \alpha_2(\lambda) \geq \cdots \geq \alpha_M(\lambda) > \alpha_{M+1}(\lambda) = \cdots = \alpha_N(\lambda) = 0. \tag{26}$$

It follows

$$\mathrm{var}(X_t) = H(2H-1)\sigma^2 G \left(\int_{-t}^{0} \int_{-t}^{0} \mathcal{E}_{-r-s}(\lambda) |r-s|^{2H-2} ds dr \right) G^T,$$

and after substitutions in integrals on the diagonal,

$$\mathrm{var}(X_t) = H(2H-1)\sigma^2 G \,\mathrm{diag}\left\{ \frac{I(\alpha_1(\lambda)t)}{\alpha_1(\lambda)^{2H}}, \ldots, \frac{I(\alpha_M(\lambda)t)}{\alpha_M(\lambda)^{2H}}, \frac{t^{2H}}{H(2H-1)}, \ldots, \frac{t^{2H}}{H(2H-1)} \right\} G^T,$$

where we denoted

$$I(x) = \int_{-x}^{0} \int_{-x}^{0} e^{r+s} |r-s|^{2H-2} ds dr. \tag{27}$$

Note that using change-of-variable formula ($u = r+s$, $v = r-s$) and analytical integrating over v yields the following formula with simple one-dimensional integrals, which is numerically more stable and easier to implement:

$$I(x) = \frac{1}{2H-1} \left[e^{-2x} \int_0^x e^u u^{2H-1} du + \int_0^x e^{-u} u^{2H-1} du \right].$$

The other two (co)variance matrices can be treated similarly:

$$\mathrm{var}(\xi_t) = H(2H-1)\sigma^2 G \,\mathrm{diag}\left\{ \frac{I(\alpha_1(\lambda)h)}{\alpha_1(\lambda)^{2H}}, \ldots, \frac{I(\alpha_M(\lambda)h)}{\alpha_M(\lambda)^{2H}}, \frac{h^{2H}}{H(2H-1)}, \ldots, \frac{h^{2H}}{H(2H-1)} \right\} G^T,$$

$$\mathrm{cov}(\xi_t, X_t) = H(2H-1)\sigma^2 G \mathcal{E}_h(\lambda) \,\mathrm{diag}\left\{ cov_1, \ldots, cov_N \right\} G^T,$$

where

$$cov_k = \begin{cases} \frac{J(\alpha_k(\lambda)h,\alpha_k(\lambda)t)}{\alpha_k(\lambda)^{2H}}, & k = 1,\dots,M, \\ \\ \frac{(t+h)^{2H}-t^{2H}-h^{2H}}{2H(2H-1)}, & k = M+1,\dots,N. \end{cases}$$

and where we define

$$J(x,y) = \int_0^x \int_{-y}^0 e^{r+s}|r-s|^{2H-2}dsdr, \tag{28}$$

After changing variables and analytical integration, we can again simplify this double integral into simple one-dimensional integrals, which are easier to approximate numerically:

$$J(x,y) = \frac{1}{2(2H-1)}\left[e^{2x}\int_x^{x+y} e^{-u}u^{2H-1}du - \int_0^y e^{-u}u^{2H-1}du + e^{-2y}\int_y^{x+y} e^{u}u^{2H-1}du - \int_0^x e^{u}u^{2H-1}du\right].$$

Ease of inversion of diagonal matrices provides us with the simplified formula

$$\begin{aligned} \mu_t(X_t,\lambda) &= e^{-A(\lambda)h}X_t \\ &+ G\,\mathrm{diag}\left\{m_1,\dots,m_N\right\}G^T(X_t - e^{-A(\lambda)t}X_0), \end{aligned}$$

where

$$m_k = \begin{cases} m(\alpha_k(\lambda),t) = e^{-\alpha_k(\lambda)h}\frac{J(\alpha_k(\lambda)h,\alpha_k(\lambda)t)}{I(\alpha_k(\lambda)t)}, & k = 1,\dots,M, \\ \\ \frac{(t+h)^{2H}-t^{2H}-h^{2H}}{2t^{2H}}, & k = M+1,\dots,N. \end{cases} \tag{29}$$

If we switch from physical to frequency domain, see (10), we get

$$\begin{aligned} \mu_t(X_t,\lambda) &= G\mathcal{E}_h(\lambda)F_t \\ &+ G\,\mathrm{diag}\left\{m_1,\dots,m_N\right\}(F_t - \mathcal{E}_t(\lambda)F_0). \end{aligned}$$

Similarly

$$\Sigma_t(\lambda) = H(2H-1)\sigma^2 G\,\mathrm{diag}\{s_1,\dots,s_N\}G^T,$$

where

$$s_k = \begin{cases} s(\alpha_k(\lambda),t) = \frac{I(\alpha_k(\lambda)h)}{\alpha_k(\lambda)^{2H}} - e^{-2\alpha_k(\lambda)h}\frac{J^2(\alpha_k(\lambda)h,\alpha_k(\lambda)t)}{I(\alpha_k(\lambda)t)}\frac{1}{\alpha_k(\lambda)^{2H}}, & k = 1,\dots,M, \\ \\ \frac{h^{2H}}{H(2H-1)} - \frac{[(t+h)^{2H}-t^{2H}-h^{2H}]^2}{t^{2H}4H(2H-1)}, & k = M+1,\dots,N. \end{cases} \tag{30}$$

Now we can construct the spectral version of the criterion (23). Let Λ be again an unknown representing the parameter to be estimated (λ is its true value) and denote by $f_k(t)$ the coefficients in frequency domain, i.e.,

$$(f_1(t),\dots,f_N(t))^T = (g_1^T X_t,\dots,g_N^T X_t)^T = G^T X_t.$$

Using this notation, we can write

$$\begin{aligned} Q(\Lambda) = \sum_{t\in\{0,h\dots,(n-1)h\}} &\Bigg[N\log\left(H(2H-1)\sigma^2\right) + \sum_{k=1}^N \log(s_k) \\ &+ \frac{1}{H(2H-1)\sigma^2}\sum_{k=1}^N \frac{1}{s_k}\left(f_k(t+h) - e^{-\alpha_k(\Lambda)h}f_k(t) - m_k[f_k(t) - e^{-\alpha_k(\Lambda)t}f_k(0)]\right)^2\Bigg], \end{aligned} \tag{31}$$

with $s_k = s_k(t, \Lambda)$ defined in (30) and $m_k = m_k(t, \Lambda)$ in (29). Note that by using spectral decomposition, we eliminated the matrix inversions from the criterion formula and integrals of matrices were substituted by real-valued integrals of the form (27) and (28). Moreover, if $k = M+1, \ldots, N$, then neither of $\alpha_k(\Lambda)$, $s_k(\Lambda, t)$ and $m_k(\Lambda, t)$ depend on Λ. As a result, minimizing $Q(\Lambda)$ with respect to Λ is the same as minimizing its reduced version:

$$\begin{aligned} Q_M(\Lambda) = \sum_{t\in\{0,h\ldots,(n-1)h\}} \Bigg[& M \log\left(H(2H-1)\sigma^2\right) + \sum_{k=1}^{M} \log(s(\alpha_k(\Lambda), t)) \\ & + \frac{1}{H(2H-1)\sigma^2} \sum_{k=1}^{M} \frac{1}{s(\alpha_k(\Lambda), t)} \left(f_k(t+h) - e^{-\alpha_k(\Lambda)h} f_k(t) - m(\alpha_k(\Lambda), t)[f_k(t) - e^{-\alpha_k(\Lambda)t} f_k(0)] \right)^2 \Bigg], \end{aligned} \tag{32}$$

which utilizes only projections to eigenvectors corresponding to strictly positive eigenvalues. This is not surprising, because in the white noise case, the projections $\{f_k(t)\}_{t\geq 0}$, $k = 1, \ldots, N$ are independent processes and

$$f_k(t) = f_k(0) + \tilde{\beta}_k^H(t), \quad k = M+1, \ldots, N.$$

Hence, $\{f_k(t)\}_{t\geq 0}$ for $k = M+1, \ldots, N$ represent (re-scaled and shifted) fractional Brownian motions that do not carry any information about the unknown parameter λ. They are, however, suitable for potential inference about σ. Such inference is outside the scope of this paper.

Remark 2. *If $H = 1/2$, the frequency version of the conditional expectation and variance is*

$$\begin{aligned} & \mu_t(X_t, \lambda) = G\mathcal{E}_h(\lambda) F_t, \\ & \Sigma_t(\lambda) = \sigma^2 G \operatorname{diag}\left\{ \frac{1 - e^{-2\alpha_1(\lambda)h}}{2\alpha_1(\lambda)}, \ldots, \frac{1 - e^{-2\alpha_M(\lambda)h}}{2\alpha_M(\lambda)}, h, \ldots, h \right\} G^T. \end{aligned}$$

The corresponding reduced criterion function for the combined estimator is given by the formula

$$\begin{aligned} Q_M(\Lambda) = \sum_{t\in\{0,h\ldots,(n-1)h\}} \Bigg[& M \log(\sigma^2) + \sum_{k=1}^{M} \log\left(\frac{1 - e^{-2\alpha_k(\Lambda)h}}{2\alpha_k(\Lambda)} \right) \\ & + \frac{1}{\sigma^2} \sum_{k=1}^{M} \frac{2\alpha_k(\Lambda)}{1 - e^{-2\alpha_k(\Lambda)h}} \left(f_k(t+h) - e^{-\alpha_k(\Lambda)h} f_k(t) \right)^2 \Bigg]. \end{aligned} \tag{33}$$

4.3. Consistency of the Combined Estimator in Diagonal Case

In this subsection, we prove the strong consistency of the spectral version of the combined estimator. The diagonality Assumptions 1 (**WN**) and 2 (**EV**), cf. Equations (24) and (25), are considered to hold throughout this subsection. In addition, let us assume the boundedness of the non-zero eigenvalues

$$0 < \inf_{\Lambda\in C} \alpha_k(\Lambda) \leq \sup_{\Lambda\in C} \alpha_k(\Lambda) < \infty \quad \forall k = 1, \ldots, M. \tag{34}$$

Motivated by the proof of the strong consistency of the conditional least-squares estimator in [15], we start with limiting behavior of the criterion function. Denote by

$$z_k(t) = \int_{-\infty}^{t} e^{-\alpha_k(\lambda)(t-s)} d\tilde{\beta}_k^H(s), \quad t \geq 0, k = 1, \ldots, M \tag{35}$$

the stationary fractional Ornstein–Uhlenbeck processes corresponding to (13). Note that (after some calculations)

$$\mathbb{E}[z_k(h)|z_k(0)] = e^{-\alpha_k(\lambda)h} z_k(0) + m(\alpha_k(\lambda), \infty) z_k(0) = \tilde{m}(\alpha_k(\lambda), \infty) z_k(0), \tag{36}$$

and

$$\mathrm{var}[z_k(h)|z_k(0)] = \mathbb{E}[(z_k(h) - \mathbb{E}[z_k(h)|z_k(0)])^2|z_k(0)] = H(2H-1)\sigma^2 s(\alpha_k(\lambda),\infty) = \tilde{s}(\alpha_k(\lambda),\infty). \quad (37)$$

Let us recall the function f defined in Equation (13) in [15] and considered in Lemma 3 therein. We have

$$f(\alpha_k(\Lambda)h) = \tilde{m}(\alpha_k(\Lambda),\infty) = e^{-\alpha_k(\Lambda)h} + m(\alpha_k(\Lambda),\infty), \quad \forall \Lambda \in C.$$

Lemma 1. *Let $H > 1/2$ and define*

$$\begin{aligned} Q_M^{(\infty)}(\Lambda) =& M\log\left(H(2H-1)\sigma^2\right) + \sum_{k=1}^{M}\log(s(\alpha_k(\Lambda),\infty)) \\ &+ \frac{1}{H(2H-1)\sigma^2}\sum_{k=1}^{M}\frac{1}{s(\alpha_k(\Lambda),\infty)}\mathbb{E}\left(z_k(h) - \tilde{m}(\alpha_k(\Lambda),\infty)z_k(0)\right)^2. \end{aligned} \quad (38)$$

Then

$$\lim_{n\to\infty}\sup_{\Lambda\in C}\left|\frac{1}{n}Q_M(\Lambda) - Q_M^{(\infty)}(\Lambda)\right| = 0 \quad \textit{almost surely (a.s.)} \quad (39)$$

Proof. To avoid repeating arguments and calculations from [15], we start by noting that

$$\sum_{t\in\{0,h\ldots,(n-1)h\}}\left(f_k(t+h) - e^{-\alpha_k(\Lambda)h}f_k(t) - m(\alpha_k(\Lambda),t)[f_k(t) - e^{-\alpha_k(\Lambda)t}f_k(0)]\right)^2 = S_n(\alpha_k(\Lambda)),$$

where S_n is the criterion for conditional LSE defined in Equation (26) in [15]. Indeed, $\{f_k(t)\}_{t\geq 0}$ is a fractional Ornstein–Uhlenbeck process with drift value $\alpha_k(\lambda)$, and

$$e^{-\alpha_k(\Lambda)h}f_k(t) + m(\alpha_k(\Lambda),t)[f_k(t) - e^{-\alpha_k(\Lambda)t}f_k(0)] = \mathbb{E}_{\alpha_k(\Lambda)}[f_k(t+h)|f_k(t)],$$

where $\mathbb{E}_{\alpha_k(\Lambda)}[f_k(t+h)|f_k(t)]$ is the conditional expectation for the fractional Ornstein–Uhlenbeck processes with drift value $\alpha_k(\Lambda)$. Similarly,

$$\mathbb{E}\left(z_k(h) - \tilde{m}(\alpha_k(\Lambda),\infty)z_k(0)\right)^2 = \mathbb{E}\left(z_k(h) - f(\alpha_k(\Lambda)h)z_k(0)\right)^2 = S_\infty(\alpha_k(\Lambda)),$$

with the function S_∞ defined in Lemma 3 in [15]. The above assumption on boundedness of eigenvalues (34) enables us to apply this lemma to get

$$\begin{aligned} \lim_{n\to\infty}\sup_{\Lambda\in C}\Bigg|\frac{1}{n}&\sum_{t\in\{0,h\ldots,(n-1)h\}}\left(f_k(t+h) - e^{-\alpha_k(\Lambda)h}f_k(t) - m(\alpha_k(\Lambda),t)[f_k(t) - e^{-\alpha_k(\Lambda)t}f_k(0)]\right)^2 \\ &- \mathbb{E}\left(z_k(h) - \tilde{m}(\alpha_k(\Lambda),\infty)z_k(0)\right)^2\Bigg| = 0 \quad \text{a.s.} \quad k = 1,\ldots,M. \end{aligned} \quad (40)$$

Another important ingredient is the uniform convergence of $s(\alpha_k(\Lambda),t)$. Observe

$$\left|s(\alpha,t) - s(\alpha,\infty)\right| = \frac{e^{-2\alpha h}}{\alpha^{2H}}\left|\frac{J^2(\alpha h,\alpha t)}{I(\alpha t)} - \frac{J^2(\alpha h,\infty)}{I(\infty)}\right|.$$

Consider a bounded interval $[\alpha_L,\alpha_U] \subset (0,\infty)$ and the uniform convergences

$$\lim_{t\to\infty}\sup_{\alpha\in[\alpha_L,\alpha_U]}|I(\alpha t) - I(\infty)| = \lim_{x\to\infty}|I(x) - I(\infty)| = 0,$$

$$\lim_{t\to\infty} \sup_{\alpha\in[\alpha_L,\alpha_U]} |J(\alpha h, \alpha t) - J(\alpha h, \infty)| = \lim_{y\to\infty} \sup_{x\in[\alpha_L h,\alpha_U h]} |J(x,y) - J(x,\infty)| = 0,$$

and the bounded limits

$$I(\infty) = \Gamma(2H-1), \quad \text{and} \quad \sup_{\alpha\in[\alpha_L,\alpha_U]\subset(0,\infty)} J(\alpha h, \infty) = J(\alpha_U h, \infty) < \infty.$$

These considerations imply the uniform convergence of s:

$$\lim_{t\to\infty} \sup_{\alpha\in[\alpha_L,\alpha_U]} \Big| s(\alpha,t) - s(\alpha,\infty) \Big| = 0,$$

and, in combination with (34),

$$\lim_{t\to\infty} \sup_{\Lambda\in C} \Big| s(\alpha_k(\Lambda),t) - s(\alpha_k(\Lambda),\infty) \Big| = 0, \quad k = 1,\ldots,M. \tag{41}$$

The last step is to show the positivity of the lower bound of $s(\alpha,\infty)$. Instead of analytical calculations; we use a simple probabilistic argument

$$s(\alpha,\infty) = \frac{1}{H(2H-1)\sigma^2} \mathrm{var}_\alpha[z_k(h)|z_k(0)] > 0, \quad \forall \alpha \in [\alpha_L,\alpha_U],$$

where $\mathrm{var}_\alpha[z_k(h)|z_k(0)]$ is the conditional variance of the stationary fractional Ornstein-Uhlenbeck process with drift value α. This, together with continuity of $s(\alpha,\infty)$ in α and the assumption (34), yields

$$\inf_{\Lambda\in C} s(\alpha_k(\Lambda),\infty) > 0, \quad k = 1,\ldots,M. \tag{42}$$

The uniform convergence (39) is now an easy consequence of (40)–(42). □

Next, we show that the limiting criterion $Q_M^{(\infty)}(\Lambda)$ is minimized for $\Lambda = \lambda$, where λ is the true value of the drift parameter of the process underlying the observed data.

Lemma 2. *Let the estimation problem be well-identified:*

$$(\alpha_1(\Lambda),\ldots,\alpha_M(\Lambda)) \neq (\alpha_1(\lambda),\ldots,\alpha_M(\lambda)) \quad \forall \Lambda \in C, \Lambda \neq \lambda, \tag{43}$$

and let the eigenvalues depend continuously on the estimated parameter:

$$\alpha_k(\Lambda) \text{ is continuous in } \Lambda \quad \forall \Lambda \in C, k = 1,\ldots,M. \tag{44}$$

Then the function $Q_M^{(\infty)}$, defined in (38) for $H > 1/2$, is continuous on C and

$$Q_M^{(\infty)}(\lambda) < Q_M^{(\infty)}(\Lambda) \quad \forall \Lambda \in C, \Lambda \neq \lambda. \tag{45}$$

Proof. Continuity of $Q_M^{(\infty)}$ follows immediately from the continuity of $s(.,\infty), \tilde{m}(.,\infty)$ and $\alpha_k(.)$. To prove (45), begin with the following decomposition, which is possible by considering the conditional expectation and (36):

$$\begin{aligned}
&\mathbb{E}\Big(z_k(h) - \tilde{m}(\alpha_k(\Lambda),\infty)z_k(0)\Big)^2 = \mathbb{E}\Big\{\mathbb{E}\Big[\Big(z_k(h) - \tilde{m}(\alpha_k(\Lambda),\infty)z_k(0)\Big)^2\Big|z_k(0)\Big]\Big\} \\
&= \mathbb{E}\Big\{\mathbb{E}\Big[\Big((z_k(h) - \tilde{m}(\alpha_k(\lambda),\infty)z_k(0)) + (\tilde{m}(\alpha_k(\lambda),\infty)z_k(0) - \tilde{m}(\alpha_k(\Lambda),\infty)z_k(0))\Big)^2\Big|z_k(0)\Big]\Big\} \\
&= \mathbb{E}\Big\{\mathbb{E}\Big[\Big(z_k(h) - \tilde{m}(\alpha_k(\lambda),\infty)z_k(0)\Big)^2\Big|z_k(0)\Big]\Big\} + \mathbb{E}\Big\{\Big(\tilde{m}(\alpha_k(\lambda),\infty)z_k(0) - \tilde{m}(\alpha_k(\Lambda),\infty)z_k(0)\Big)^2\Big\} \\
&= \tilde{s}(\alpha_k(\lambda),\infty) + \Big(\tilde{m}(\alpha_k(\lambda),\infty) - \tilde{m}(\alpha_k(\Lambda),\infty)\Big)^2 \mathbb{E}z_k(0)^2,
\end{aligned}$$

where the last equality is due to relations (36) and (37), which imply

$$\mathbb{E}\Big\{\mathbb{E}\Big[\Big(z_k(h) - \tilde{m}(\alpha_k(\lambda),\infty)z_k(0)\Big)^2\Big|z_k(0)\Big]\Big\} = \mathbb{E}\Big\{\mathrm{var}[z_k(h)|z_k(0)]\Big\} = \tilde{s}(\alpha_k(\lambda),\infty).$$

Using this decomposition and $\tilde{s}(\alpha_k(\Lambda),\infty) = H(2H-1)\sigma^2 s(\alpha_k(\Lambda),\infty)$ we can write

$$\begin{aligned}
Q_M^{(\infty)}(\Lambda) = &\sum_{k=1}^{M} \log(\tilde{s}(\alpha_k(\Lambda),\infty)) \\
&+ \sum_{k=1}^{M}\Bigg[\frac{\tilde{s}(\alpha_k(\lambda),\infty)}{\tilde{s}(\alpha_k(\Lambda),\infty)} + \frac{\mathbb{E}z_k(0)^2}{\tilde{s}(\alpha_k(\Lambda),\infty)}\Big(\tilde{m}(\alpha_k(\lambda),\infty) - \tilde{m}(\alpha_k(\Lambda),\infty)\Big)^2\Bigg].
\end{aligned}$$

Consequently

$$Q_M^{(\infty)}(\lambda) = \sum_{k=1}^{M} \log(\tilde{s}(\alpha_k(\lambda),\infty)) + \sum_{k=1}^{M} 1,$$

and

$$\begin{aligned}
Q_M^{(\infty)}(\Lambda) - Q_M^{(\infty)}(\lambda) = \sum_{k=1}^{M}\Bigg[&-\log\Big(\frac{\tilde{s}(\alpha_k(\lambda),\infty)}{\tilde{s}(\alpha_k(\Lambda),\infty)}\Big) + \frac{\tilde{s}(\alpha_k(\lambda),\infty)}{\tilde{s}(\alpha_k(\Lambda),\infty)} - 1 \\
&+ \frac{\mathbb{E}z_k(0)^2}{\tilde{s}(\alpha_k(\Lambda),\infty)}\Big(\tilde{m}(\alpha_k(\lambda),\infty) - \tilde{m}(\alpha_k(\Lambda),\infty)\Big)^2\Bigg].
\end{aligned}$$

The inequality

$$Q_M^{(\infty)}(\Lambda) \geq Q_M^{(\infty)}(\lambda), \quad \forall \Lambda \in C,$$

now follows from the fact that

$$\gamma(x) = -\log(x) + x - 1 \geq 0, \quad \forall x > 0.$$

To demonstrate the uniqueness of the minimum:

$$Q_M^{(\infty)}(\Lambda) = Q_M^{(\infty)}(\lambda) \quad \Rightarrow \quad \Lambda = \lambda,$$

note that $\tilde{m}(.,\infty)$ is strictly decreasing, which follows from Lemma 2 in [15] and the fact that $\tilde{m}(x,\infty) = f(x)$. □

Remark 3. *In this remark, we consider the simpler case $H = 1/2$ and outline validity of the properties in Lemmas 1 and 2. Recall $Q_M(\Lambda)$ from Remark 2 and define*

$$\begin{aligned} Q_M^{(\infty)}(\Lambda) =& M\log(\sigma^2) + \sum_{k=1}^{M} \log\left(\frac{1-e^{-2\alpha_k(\Lambda)h}}{2\alpha_k(\Lambda)}\right) \\ &+ \frac{1}{\sigma^2}\sum_{k=1}^{M} \frac{2\alpha_k(\Lambda)}{1-e^{-2\alpha_k(\Lambda)h}} \mathbb{E}\left(z_k(h) - e^{-\alpha_k(\Lambda)h} z_k(0)\right)^2, \end{aligned}$$

where $\{z_k(t)\}_{t\geq 0}$ is the stationary Ornstein–Uhlenbeck process defined in (35) with $H = 1/2$. By using arguments similar to those used in the proof of Lemma 1, including the boundedness assumption (42), we can show the uniform convergence

$$\lim_{n\to\infty} \sup_{\Lambda\in C} \left| \frac{1}{n} Q_M(\Lambda) - Q_M^{(\infty)}(\Lambda) \right| = 0 \quad a.s.$$

To show that λ is the unique minimum of $Q_M^{(\infty)}$, use the following notation

$$\tilde{s}(\alpha_k(\Lambda)) = \sigma^2 \frac{1-e^{-2\alpha_k(\Lambda)h}}{2\alpha_k(\Lambda)},$$

and write

$$\begin{aligned} Q_M^{(\infty)}(\Lambda) &= \sum_{k=1}^{M} \log(\tilde{s}(\alpha_k(\Lambda))) + \sum_{k=1}^{M} \frac{1}{\tilde{s}(\alpha_k(\Lambda))} \mathbb{E}\left(z_k(h) - e^{-\alpha_k(\Lambda)h} z_k(0)\right)^2 \\ &= \sum_{k=1}^{M} \left[\log(\tilde{s}(\alpha_k(\Lambda))) + \frac{1}{\tilde{s}(\alpha_k(\Lambda))} \left(\tilde{s}(\alpha_k(\lambda)) + \frac{\sigma^2}{2\alpha_k(\lambda)} \left(e^{-\alpha_k(\lambda)h} - e^{-\alpha_k(\Lambda)h}\right)^2 \right) \right]. \end{aligned}$$

Consequently

$$\begin{aligned} &Q_M^{(\infty)}(\lambda) - Q_M^{(\infty)}(\Lambda) \\ &= \sum_{k=1}^{M} \left[-\log\left(\frac{\tilde{s}(\alpha_k(\lambda))}{\tilde{s}(\alpha_k(\Lambda))}\right) + \frac{\tilde{s}(\alpha_k(\lambda))}{\tilde{s}(\alpha_k(\Lambda))} - 1 + \frac{1}{\tilde{s}(\alpha_k(\Lambda))} \frac{\sigma^2}{2\alpha_k(\lambda)} \left(e^{-\alpha_k(\lambda)h} - e^{-\alpha_k(\Lambda)h}\right)^2 \right]. \end{aligned}$$

Employing the identification assumption (43) and the continuity assumption (44) results in the desired property (45). Note that the continuity of $Q_M^{(\infty)}$ on C is immediate.

We are in a position to formulate the main result of this section—the strong consistency of the combined estimator.

Theorem 1. *Let the diagonality Assumptions 1 (**WN**) and 2 (**EV**), cf. Equations (24) and (25), hold true. Furthermore, assume that the boundedness (34), identifiability (43) and continuity (44) of the eigenvalues are satisfied. Then, the combined estimator defined in (22) is strongly consistent, i.e.,*

$$\hat{\lambda} \overset{n\to\infty}{\longrightarrow} \lambda \quad a.s.$$

Proof. The strong consistency can be demonstrated by standard arguments utilizing the compactness of C together with Lemmas 1 and 2 for $H > 1/2$ and Remark 3 for $H = 1/2$, respectively. For details, see, for example, the proof of Theorem 3 in [15]. □

5. Simulation Study

5.1. Algorithm for Simulations

Simulating stochastic evolution equations with fBm as the noise source is not an easy task. To avoid complications with simulating stochastic integral w.r.t. fBm and safe computational resources, we simulated the solution to (5) in frequency domain with use of the Formula (14), which replaces the stochastic integral with the standard Riemann integral. The algorithm we used for simulations is outlined below:

1. Simulate N-dimensional cylindrical fBm B_t^H as N independent real-valued fBm.
2. Calculate $\tilde{B}_t^H = (\tilde{\beta}_1^H(t), \ldots, \tilde{\beta}_N^H(t))^T$ by formula (12).
3. Find eigenvalues $\{\alpha_k(\lambda)\}_{k=1\ldots N}$ and eigenvectors $\{g_k(\lambda)\}_{k=1\ldots N}$ of $A(\lambda)$.
4. Transform initial condition into frequency domain $f_k(0) = g_k(\lambda)^T X_0$.
5. Calculate the whole solution in the frequency domain by formula (14).
6. Transform the solution back to physical domain by $X_t = G(\lambda)F_t$.

5.2. Examples

We choose to illustrate the performance of the combined estimator on two simple examples of stochastic heat equations on graphs. In both examples, we consider stationary and non-stationary initial conditions and two values of the Hurst index $H \in \{0.5, 0.75\}$. Moreover, to illustrate the effect of increasing number of vertices N on the combined estimator, we study an example with no interactions (edges) between vertices, so that the drift operator is a diagonal matrix.

Example 1. *Consider a stochastic heat Equation* (4) *on the graph shown in Figure 1. Such an equation can be understood as a space-discretization of a stochastic heat equation on a 1-dimensional bounded domain, with a zero Neumann boundary condition.*

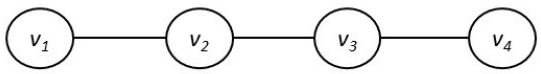

Figure 1. Stochastic heat equation on a linear graph.

Simulation setting:

$$\lambda = 1, \quad \Phi = Id = \mathrm{diag}\{1,1,1,1\}, \quad X_0 = \begin{cases} (0,100,0,10) \\ (10,10,10,10) \end{cases}, \quad H = \begin{cases} 0.5 \\ 0.75 \end{cases}.$$

The solutions have been simulated with time step $\delta t = 0.01$. For the purpose of estimating λ, we have sampled the simulated solution with time step $h = 0.1$, i.e., we have taken every 10th simulated value.

The simulated trajectories in the physical domain of a sample solution to Examples 1 with the non-stationary initial condition and $H = 0.75$ are shown in Figure 2.

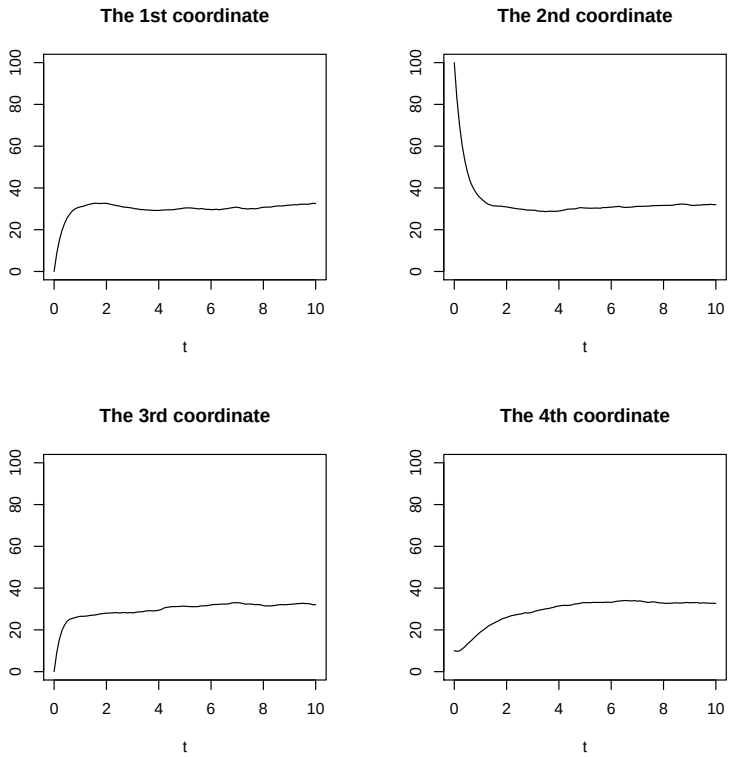

Figure 2. The trajectories of a sample solution to Example 1, where $H = 0.75$.

Example 2. *In the second example, we study a stochastic heat Equation* (4) *on the complete graph shown in Figure 3. Such an equation models situation when the diffusion of heat (or other quantity) occurs between all pairs of vertices.*

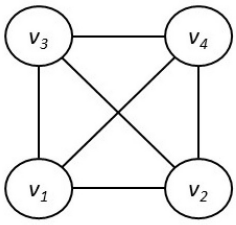

Figure 3. Stochastic heat equation on a complete graph.

Simulation setting:

$$\lambda = 1, \quad \Phi = Id = \mathrm{diag}\{1,1,1,1\}, \quad X_0 = \begin{cases} (0,100,0,10) \\ (10,10,10,10) \end{cases}, \quad H = \begin{cases} 0.5 \\ 0.75 \end{cases}.$$

Time step for simulations $\delta t = 0.01$, *time step for estimation* $h = 0.1$.

Trajectories of a single realization of the solution to Example (2) *with the non-stationary initial condition and* $H = 0.75$ *are depicted in Figure 4.*

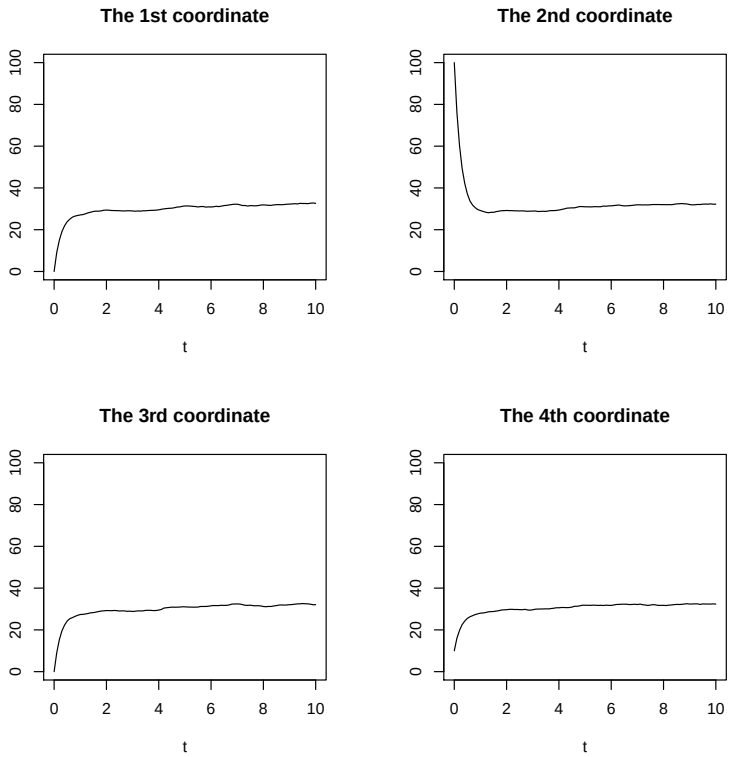

Figure 4. The trajectories of a sample solution to Example 2, where $H = 0.75$.

Example 3. *Consider an evolution equation on N isolated vertices without any interactions (there are no edges) with the drift operator in Equation (5) being a diagonal $N \times N$ matrix:*

$$A(\lambda) = \lambda \cdot Id = \mathrm{diag}\{\lambda, \lambda, \ldots, \lambda\}.$$

Such a simple equation is appropriate for studying how the performance of the estimator depends on the number of vertices N.
Simulation setting:
$\lambda = 1, \quad \Phi = Id = \mathrm{diag}\{1, 1, \ldots, 1\}, \quad X_0 = (0, 0, \ldots, 0), \quad H = 0.75.$

The solutions have been simulated with short time horizon $T = 1$ and with time step $\delta t = 0.01$. For the purpose of estimating λ, we have sampled the simulated solution with time step $h = 0.1$, i.e., we have taken every 10th simulated value.

5.3. Results and Discussion

5.3.1. Examples 1 and 2

The behavior of the combined estimator in different scenarios (various combinations of X_0, H, and time horizon T) for the two heat-equation examples is studied below. Besides application of the Formula (32) when $H = 0.75$ and (33) when $H = 0.5$, we also apply Formula (33) (the MLE criterion when $H = 0.5$) for trajectories generated by a model with $H = 0.75$. This illustrates the effect of model misspecification when the observed process has fractional noise, but we use the standard estimation technique from white-noise theory.

Boxplots of the resulting estimates obtained from the Monte Carlo simulations with 100 independent samples for each scenario are shown in Figure 5 (Example 1) and in Figure 6 (Example 2). The related root mean square errors (RMSE) can be found in Table 1.

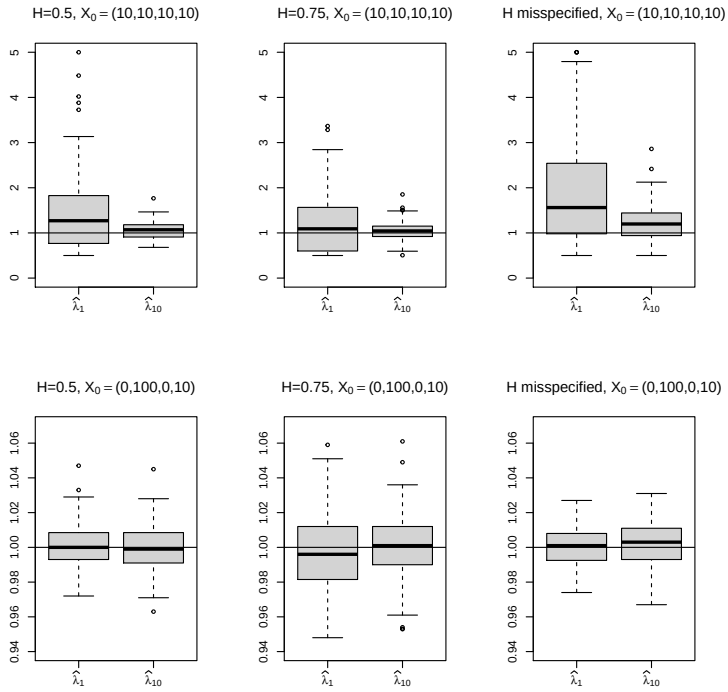

Figure 5. Boxplots of the estimated λ in Example 1 for time horizons $T = 1$ (denoted by $\widehat{\lambda}_1$) and $T = 10$ (denoted by $\widehat{\lambda}_{10}$) calculated for 100 samples in every case considered; the horizontal line shows the true value of the estimated parameter $\lambda = 1$.

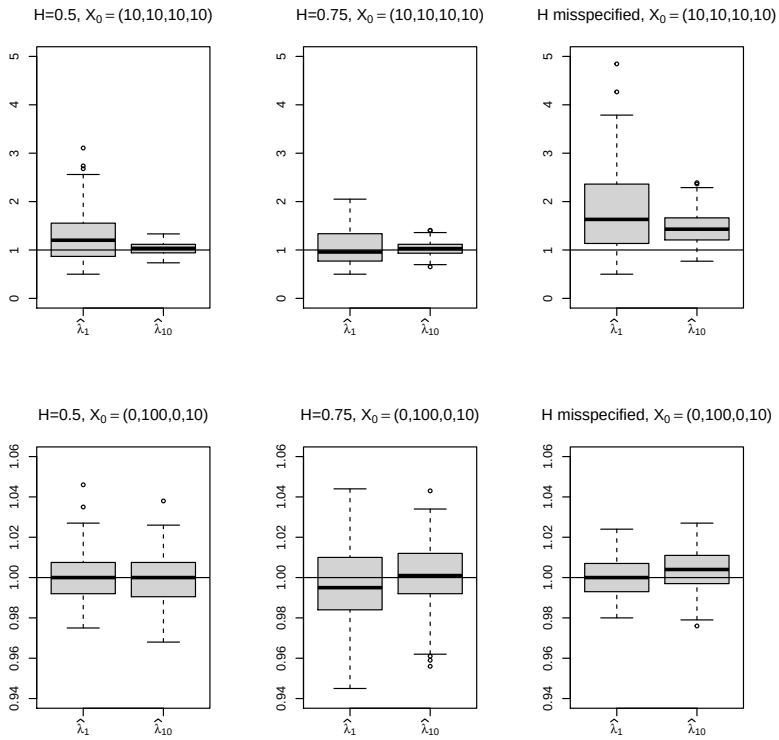

Figure 6. Boxplots of the estimated λ in Example 2 for time horizons $T = 1$ (denoted by $\widehat{\lambda}_1$) and $T = 10$ (denoted by $\widehat{\lambda}_{10}$) calculated for 100 samples in every case considered; the horizontal line shows the true value of the estimated parameter $\lambda = 1$.

Table 1. RMSE of the calculated estimates using 100 samples with $\lambda = 1$.

Initial Condition	T	H	Example 1	Example 2
(0, 100, 0, 10)	1	0.5	0.0132944	0.0126748
(0, 100, 0, 10)	1	0.75	0.0227583	0.0201745
(0, 100, 0, 10)	10	0.5	0.0134145	0.0125359
(0, 100, 0, 10)	10	0.75	0.019374	0.016389
(0, 100, 0, 10)	100	0.75	0.0180549	0.016476
(0, 100, 0, 10)	1	misspecified	0.0118246	0.0101789
(0, 100, 0, 10)	10	misspecified	0.0128522	0.0114258
(10, 10, 10, 10)	1	0.5	1.04229	0.637036
(10, 10, 10, 10)	1	0.75	0.654375	0.396853
(10, 10, 10, 10)	10	0.5	0.215363	0.139379
(10, 10, 10, 10)	10	0.75	0.217963	0.157472
(10, 10, 10, 10)	100	0.75	0.0714899	0.0515372
(10, 10, 10, 10)	1	misspecified	1.46259	1.18714
(10, 10, 10, 10)	10	misspecified	0.461442	0.568978

Estimating λ from solutions with the non-stationary initial condition provides significantly more accurate results compared to estimating from the stationary solution. This indicates that the combined estimator can efficiently utilize the valuable information about λ encoded in the speed of the convergence of the solution process to the stationary state in the initial phase. The solution in this phase is affected mainly by the drift operator; the effect of the random noise is marginal. This is in contrast to the stationary phase, where random noise prevails.

In line with the strong consistency of the combined estimator, we observe that increasing time horizon improves the accuracy of the estimates. This property is evident in the stationary case. In non-stationary case, only minor improvement has been observed, because most of the information about λ is stored in initial non-stationary phase. Time horizon T would have to be much longer to surpass this effect.

Analysis of the misspecified model reveals that it provides significantly biased estimates for stationary processes. Hence, there is a strong need to use specific methods for processes with fractional (autocorrelated) noise. The combined estimator overcomes this issue by the correction term for autocorrelation contained in Formula (32), but not in (33). In the non-stationary case, the effect of the initial condition suppresses the effect of the autocorrelated noise and the misspecified estimator shows competitive performance.

Comparison of Example 1 with Example 2 shows slightly better estimates for the latter (in terms of RMSE). More edges in the complete graph in Example 2 represent more transfer paths and more heat diffusion. In result, it contains more information about λ within the fixed time-window.

Figure 7 illustrates the path-wise (almost sure) convergence of the estimator to the true value of λ. It shows the evolution of the estimates with increasing time horizon T for three samples of the solution to Example 1 with $H = 0.75$ and $X_0 = (10, 10, 10, 10)$.

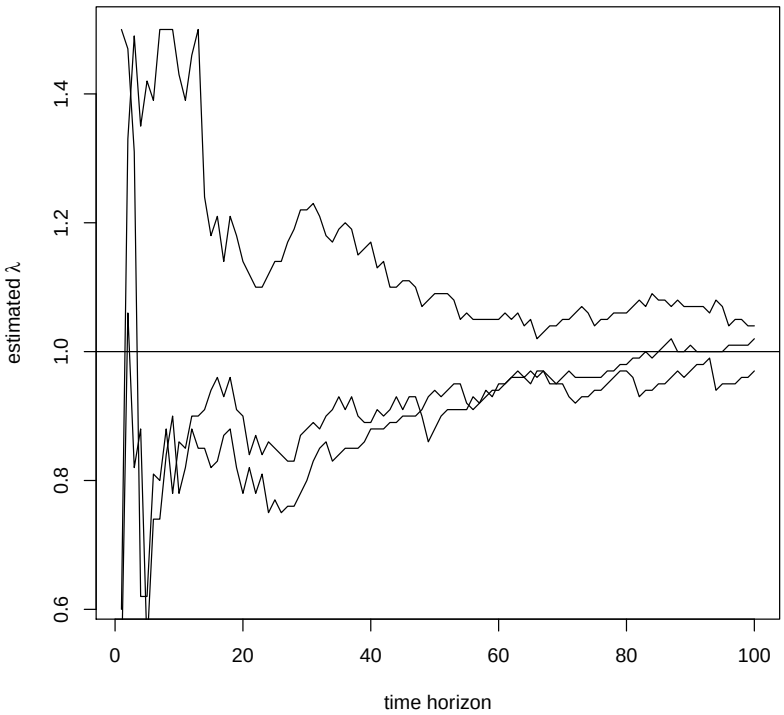

Figure 7. The convergence of $\hat{\lambda}$ for three sample solutions from Example 1 with $H = 0.75$ and $X_0 = (10, 10, 10, 10)$; the horizontal line shows the true value of the estimated parameter $\lambda = 1$.

5.3.2. Example 3

The dependence of the behavior of the combined estimator on the number of vertices in the setting of the diagonal example with no interactions between vertices (Example 3) is studied below.

The results of Monte Carlo simulations (with 100 runs for each setting) are shown in the table with corresponding RMSE (Table 2). Obviously, with an increasing number of vertices, the estimator improves, which suggests the space-consistency of the combined estimator. This consistency can be conjectured due to the fact that the combined estimator follows the maximum likelihood approach in the space domain. However, a detailed study of space asymptotics of the combined estimator is not in the scope of this paper, and may be the direction of the further research. We expect that the asymptotic behavior of eigenvalues of the drift operator $A(\lambda)$ will play important role here.

Table 2. RMSE of the calculated estimates using 100 samples with $\lambda = 1$ in Example 3.

N	RMSE
3	0.999612
5	0.756201
10	0.508276
20	0.472542

6. Conclusions

In this paper, we initiated a study of parameter estimation in the setting of stochastic evolution equations on graphs driven by fractional Brownian motions. We illustrated on the numerical simulation that such models can not be treated directly by estimation techniques known for equations with standard Brownian motion (see the misspecified model). Hence, we introduced a novel approach, which combines the conditional least-squares technique in the time domain with the maximum-likelihood technique in the space domain. This leads to a new estimator that we call the combined estimator. It has favourable asymptotic properties; we proved its strong consistency in the

long-span asymptotic regime. Unlike the minimum-contrast estimator studied in [14], which could also be applied in our setting, the combined estimator is very efficient also for non-stationary solutions. We also derived an easy-to-implement formula in diagonal setting and illustrate its behaviour in various scenarios by Monte Carlo simulations.

Although we documented promising properties of the combined estimator above, some important questions remain open and may determine the direction of further research. These include (but are not limited to) the speed of convergence and asymptotic distribution, the space asymptotics (increasing number of nodes), modification for the singular case with Hurst index $H < 1/2$ (negatively correlated increments), simultaneous estimation of H, Φ (or its magnitude) and λ, or parameter estimation in more complex and interesting examples with the application of more sophisticated graph theory.

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 the anonymous reviewers for their valuable and encouraging comments. They helped us significantly improve the paper.

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

References

1. Mesbahi, M.; Egerstedt, M. *Graph Theoretic Methods in Multiagent Networks (Princeton Series in Applied Mathematics)*; Princeton University Press: Princeton, NJ, USA, 2010.
2. Bonaccorsi, S.; Mugnolo, D. Existence of strong solutions for neuronal network dynamics driven by fractional Brownian motions. *Stoch. Dyn.* **2010**, *10*. [CrossRef]
3. Mishura, Y. *Stochastic Calculus for Fractional Brownian Motion and Related Processes*; Springer: Berlin/Heidelberg, Germany, 2008.
4. Biagini, F.; Hu, Y.; Øksendal, B.; Zhang, T. *Stochastic Calculus for Fractional Brownian Motion and Applications*; Springer: London, UK, 2008.
5. Zhang, Z.; Karniadakis, G.E. *Numerical Methods for Stochastic Partial Differential Equations with White Noise*; Springer International Publishing AG: Cham, Switzerland, 2017. [CrossRef]
6. Kutoyants, Y. *Statistical Inference for Ergodic Diffusion Processes*; Springer: London, UK, 2004.
7. Cialenco, I. Statistical inference for SPDEs: an overview. *Stat. Inference Stoch. Process.* **2018**, *21*, 309–329. [CrossRef]
8. Kubilius, K.; Mishura, Y.; Ralchenko, K. *Parameter Estimation in Fractional Diffusion Models*; Springer International Publishing AG: Cham, Switzerland, 2017.
9. Prakasa Rao, B.L.S. *Statistical Inference for Fractional Diffusion Processes*; John Wiley & Sons: Chichester, UK, 2010.
10. Estrada, E. *d*-Path Laplacians and Quantum Transport on Graphs. *Mathematics* **2020**, *8*, 527. [CrossRef]
11. Markussen, B. Likelihood inference for a discretely observed stochastic partial differential equation. *Bernoulli* **2003**, *9*, 745–762. [CrossRef]
12. Cialenco, I.; Delgado-Vences, F.; Kim, H. Drift estimation for discretely sampled SPDEs. *Stoch. Partial. Differ. Equ. Anal. Comput.* **2020**. [CrossRef]
13. Hildebrandt, F.; Trabs, M. Parameter estimation for SPDEs based on discrete observations in time and space. *arXiv* **2019**, arXiv:1910.01004.
14. Kříž, P.; Maslowski, B. Central limit theorems and minimum-contrast estimators for linear stochastic evolution equations. *Stochastics* **2019**, *91*, 1109–1140. [CrossRef]
15. Kříž, P.; Szała, L. Least-Squares Estimators of Drift Parameter for Discretely Observed Fractional Ornstein-Uhlenbeck Processes. *Mathematics* **2020**, *8*, 716. [CrossRef]
16. Istas, J.; Lang, G. Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Ann. l'Inst. Henri Poincare (B) Probab. Stat.* **1997**, *33*, 407–436. [CrossRef]
17. Coeurjolly, J. Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles. *Ann. Stat.* **2008**, *36*, 1404–1434. [CrossRef]

18. Rosenbaum, M. Estimation of the volatility persistence in a discretely observed diffusion model. *Stoch. Process. Appl.* **2008**, *118*, 1434–1462. [CrossRef]
19. 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]
20. Pipiras, V.; Taqqu, M. Integration questions related to fractional Brownian motion. *Probab. Theory Relat. Fields* **2000**, *118*, 251–291. [CrossRef]
21. Duncan, T. E.; Pasik-Duncan, B.; Maslowski, B. Fractional Brownian motion and stochastic equations in Hilbert spaces. *Stoch. Dyn.* **2002**, *2*, 225–250. [CrossRef]
22. Pasik-Duncan, B.; Duncan, T.E.; Maslowski, B. Linear stochastic equations in a Hilbert space with a fractional Brownian motion. *Int. Ser. Oper. Res. Manag. Sci.* **2006**, *94*, 201–222. [CrossRef]
23. Tindel, S.; Tudor, C.; Viens, F. Stochastic evolution equations with fractional Brownian motion. *Probab. Theory Relat. Fields* **2003**, *127*, 186–204. [CrossRef]
24. Beran, J.; Feng, Y.; Ghosh, S.; Kulik R. *Long-Memory Processes: Probabilistic Properties and Statistical Methods*; Springer: Berlin/Heidelberg, Germany, 2013.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).