

Article

Inference for Convolutionally Observed Diffusion Processes

Shogo H Nakakita ^{1,2,3,*}  and Masayuki Uchida ^{1,4}

¹ Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyamacho, Toyonaka, Osaka 560-0043, Japan; uchida@sigmath.es.osaka-u.ac.jp

² Japanese Society for the Promotion of Science, 5 Chome-3-1 Kōjimachi, Chiyoda City, Tokyo 102-0083, Japan

³ The Ronin Institute for Independent Scholarship, 127 Haddon Place, Montclair, Montclair, NJ 07043, USA

⁴ Center for Mathematical Modeling and Data Science, Osaka University, 1-3 Machikaneyamacho, Toyonaka, Osaka 560-0043, Japan

* Correspondence: nakakita@sigmath.es.osaka-u.ac.jp

Received: 21 August 2020; Accepted: 11 September 2020; Published: 15 September 2020



Abstract: We propose a new statistical observation scheme of diffusion processes named convolutional observation, where it is possible to deal with smoother observation than ordinary diffusion processes by considering convolution of diffusion processes and some kernel functions with respect to time parameter. We discuss the estimation and test theories for the parameter determining the smoothness of the observation, as well as the least-square-type estimation for the parameters in the diffusion coefficient and the drift one of the latent diffusion process. In addition to the theoretical discussion, we also examine the performance of the estimation and the test with computational simulation, and show an example of real data analysis for one EEG data whose observation can be regarded as smoother one than ordinary diffusion processes with statistical significance.

Keywords: convolutional observation; diffusion processes; parametric inference; partial observation; statistical modelling; stochastic differential equations

1. Introduction

We consider a d -dimensional diffusion process defined by the following stochastic differential equation,

$$dX_t = b(X_t, \beta) dt + a(X_t, \alpha) dw_t, \quad X_{-\lambda} = x_{-\lambda},$$

where $\lambda > 0$, $\{w_t\}_{t \geq -\lambda}$ is a standard r -dimensional Wiener process, $x_{-\lambda}$ is an \mathbf{R}^d -valued random variable independent of $\{w_t\}_{t \geq -\lambda}$, $\alpha \in \Theta_1$ and $\beta \in \Theta_2$ are unknown parameters, $\Theta_1 \subset \mathbf{R}^{m_1}$ and $\Theta_2 \subset \mathbf{R}^{m_2}$ are compact and convex parameter spaces, $a : \mathbf{R}^d \times \Theta_1 \rightarrow \mathbf{R}^d \otimes \mathbf{R}^r$ and $b : \mathbf{R}^d \times \Theta_2 \rightarrow \mathbf{R}^d$ are known functions. Our concern is statistical estimation for α and β from observation. $\theta_* = (\alpha_*, \beta_*)$ denotes the true value of $\theta := (\alpha, \beta)$.

We denote the observation as the discretised process $\{\bar{X}_{ih_n, n} : i = 0, \dots, n\}$ with discretisation step $h_n > 0$ such that $h_n \rightarrow 0$ and $T_n := nh_n \rightarrow \infty$, where the convoluted process $\{\bar{X}_{t, n}\}_{t \geq 0}$ is defined as

$$\bar{X}_{t, n} := \int_{t-\bar{\rho}h_n}^t V_{h_n}(t-s) X_s ds = \int_{\mathbf{R}} V_{h_n}(t-s) X_s ds = (V_{h_n} * X)(t),$$

where V_{h_n} is an $\mathbf{R}^d \otimes \mathbf{R}^d$ -valued kernel function whose support is a subset of $[0, \bar{\rho}h_n]$, and $\bar{\rho} > 0$ such that $\sup_n \bar{\rho}h_n \leq \lambda$. In this paper, we specify $V_{h_n} = V_{\rho, h_n}$ which is a parametric kernel function whose support is a subset of $[0, \bar{\rho}h_n]$ defined as

$$V_{\rho, h_n}^{(i,j)}(t) := \begin{cases} \left(\rho^{(i)}h_n\right)^{-1} \mathbf{1}_{[0, \rho^{(i)}h_n]}(t) & \text{if } i = j \text{ and } \rho^{(i)} > 0, \\ \delta(t) & \text{if } i = j \text{ and } \rho^{(i)} = 0, \\ 0 & \text{if } i \neq j, \end{cases}$$

$\delta(t)$ is the Dirac-delta function, $\rho = [\rho^{(1)}, \dots, \rho^{(d)}]^T \in \Theta_\rho := [0, \bar{\rho}]^d$ is the smoothing parameter determining the smoothness of observation. That is to say, the observed process is defined as follows:

$$\bar{X}_{ih_n, n}^{(\ell)} = \begin{cases} \left(\rho^{(\ell)}h_n\right)^{-1} \int_{(i-\rho^{(\ell)})h_n}^{ih_n} X_s^{(\ell)} ds & \text{if } \rho^{(\ell)} > 0, \\ X_{ih_n}^{(\ell)} & \text{if } \rho^{(\ell)} = 0, \end{cases}$$

for all $\ell = 1, \dots, d$. Let us consider both the problems that (i) ρ is a known parameter, and (ii) ρ is an unknown one whose true value is denoted by ρ_* and this is estimated by observation $\{\bar{X}_{ih_n, n}\}$, and the parameter space is denoted as $\Xi := \Theta_\rho \times \Theta$, where $\Theta := \Theta_1 \times \Theta_2$.

When assuming ρ as a known parameter, we can find literature for parametric estimation for α and/or β based on observation schemes which can be represented as special cases for some specific ρ . If $\rho = \mathbf{0}$, our scheme is simply equivalent to parametric inference based on discretely observed diffusion processes $\{X_{ih_n} : i = 0, \dots, n\}$ studied in [1–8] and references therein. If $\rho = [1, \dots, 1]^T$, we can regard the problem as parametric estimation for integrated diffusion processes discussed in Gloter [9], Ditlevsen and Sørensen [10], Gloter [11], Gloter and Gobet [12], Sørensen [13]. Even for the case $\rho = [0, \dots, 0, 1, \dots, 1]^T$ where some axes correspond to direct observation and the others do to integrated observation, we give consistent estimators for α and β by considering the scheme of convolutionally observed diffusion processes and this is one of the contributions of our study.

What is more, our contribution is to consider the scheme where ρ is unknown and succeed in representation of the microstructure noise which makes the observation smoother than the latent diffusion process itself, which can be seen in some biomedical time series data. Statistical modelling of biomedical time series data with stochastic differential equations has been one of the topics eagerly studied e.g., [14–17]. As [18] states, the existence of microstructure noise in financial data affects realised volatilities to increase as the subsampling frequency gets higher, for instance, see Figure 7.1 on p. 217 in [19]. However, realised volatilities of some biomedical data such as EEG decrease as subsampling frequency increases. For instance, some time series data for the 2nd participant in the dataset named Two class motor imagery (002-2014) of BNCI Horizon 2020 [20] show clear tendency of decreasing realised volatilities as subsampling frequency increases. Figure 1 shows the path of the 2nd axis of the data S02E.mat BNCI Horizon 2020 [20] for all 222 seconds (the observation frequency is 512 Hz, and hence the entire data size is 113664) and that for the first one second; it seems to perturb like a diffusion process.

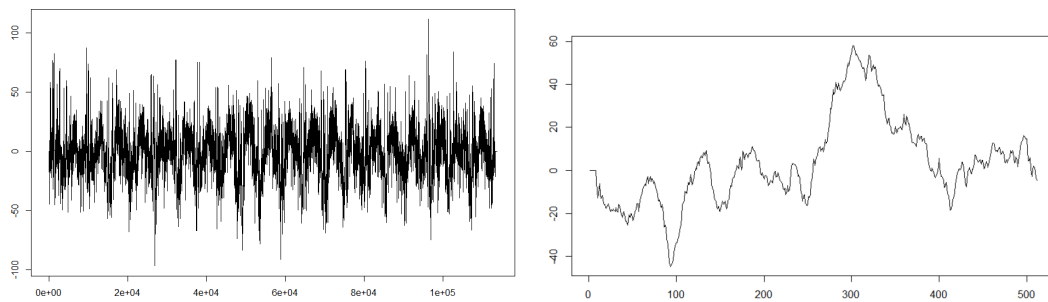


Figure 1. The path of the second column of S02E.mat of BNCI Horizon 2020 [20] for all 222 seconds (left) and the first one second (right).

We define realised volatilities with subsampling as for a sequence of real-valued observation $\{Y_i\}_{i=0,\dots,n}$,

$$RV(k) = \sum_{1 \leq i \leq [n/k]} (Y_{ik} - Y_{(i-1)k})^2,$$

where $k = 1, \dots, 100$ is the subsampling frequency parameter, and provide a plot of the realised volatilities the 2nd axis of the data S02E.mat in Figure 2:

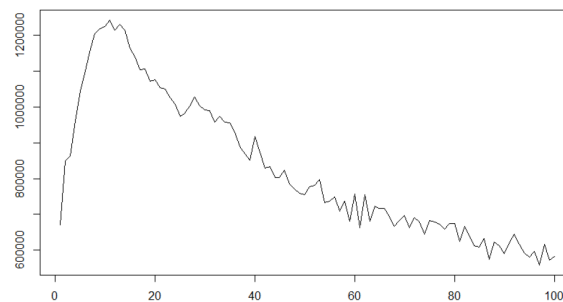


Figure 2. Realised volatilities with subsampling of the 2nd axis of data S02E.mat in two class motor imagery (002-2014) [20].

The altitudes of the graph represented in the y -axis correspond to the values of the realised volatilities $RV(k)$ with subsampling at every k observation represented in the x -axis. It is observable that the increasing subsampling frequency results in decreasing realised volatilities, which cannot be explained by the existent major microstructure noises, e.g., see [21–25]. To explain this phenomenon, we consider the smoother process than the latent one, though ordinarily microstructure noises make the observation rougher than the latent process, because quadratic variation of a sufficiently smooth function is zero. One way to deal with smoother observation than the latent state is convolutional observation. As a concrete example, we show a convolutionally observed diffusion process and its characteristics in realised volatilities: let us consider the following 1-dimensional stochastic differential equation defining an Ornstein–Uhlenbeck (OU) process:

$$dX_t = -20X_t dt + 10dw_t, X_{-\lambda} = 0,$$

where $\lambda = 10^{-2/5}$. We simulate the stochastic differential equation by Euler–Maruyama method see [26] with parameters $n = 10^7$, $h_n = 10^{-5}$, and $T_n = 10^2$ and its convolution approximated by summation with the smoothing parameter $\rho = 10$ (for details, see Section 5). Figure 3 shows the latent diffusion process and the convoluted observation on $[0, 1]$, and we can see that the observation is indeed smoothed compared to the latent state.

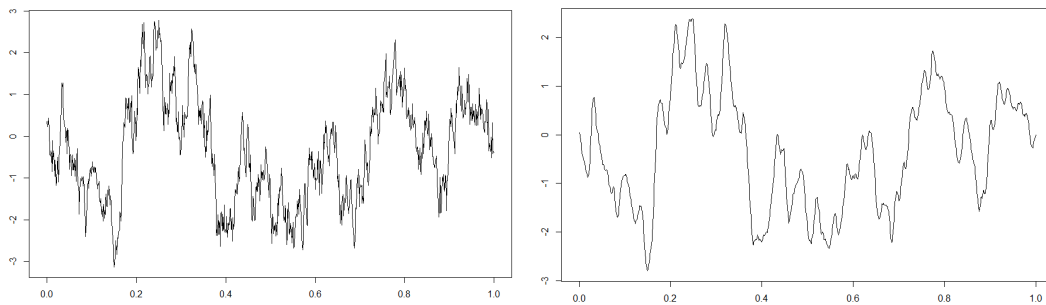


Figure 3. The left figure is the plot of the latent diffusion process, and the right one is that of the convolutionally observed process on $[0, 1]$ respectively.

In Figure 4, we also give the plot of realised volatilities of the convolutionally observed process with subsampling as Figure 2.

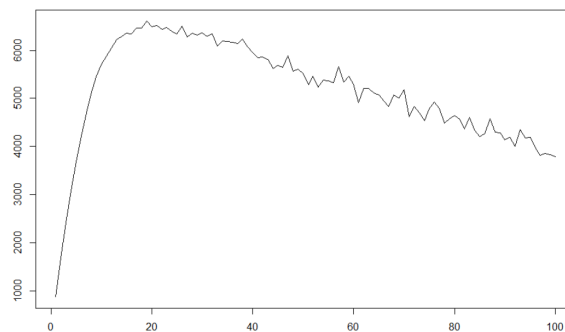


Figure 4. The realised volatilities of the convolutionally observed diffusion process with subsampling.

It is seen that the convolutional observation of a diffusion process also has the characteristics of decreasing realised volatilities as subsampling frequency increases, which can be seen in some biological data such as BNCI Horizon 2020 [20]. Of course, graphically comparing characteristics of simulation and real data is insufficient to verify the convolutional observation with smoothing parameter $\rho > 0$ in 1-dimensional case; therefore, we propose statistical estimation method for unknown ρ and hypothesis test with the null hypothesis $H_0 : \rho = 0$ and the alternative one $H_1 : \rho \neq 0$ from convolutional observation in Section 3. Moreover, in Section 6, we examine the real EEG data plotted in Figure 2 by the statistical hypothesis testing we propose, and see it is more appropriate to consider the data as a convolutional observation of a latent diffusion process with $\rho \neq 0$ rather than direct observation of the latent process, which indicates the validity to deal with the problem of the convolutional observation scheme with unknown ρ .

The paper is composed of the following sections: Section 2 gives the notations and assumptions used in this paper; Section 3 discusses the estimation and test for smoothing parameter ρ , and the discussion provides us with the tools to examine whether we should consider the convolutional observation scheme; Section 4 proposes the quasi-likelihood functions for the parameter of diffusion processes α and β , and corresponding estimators with consistency; Section 5 is for the computational simulation to examine the theoretical results in the previous sections; and Section 6 shows an application of the methods we propose in real data analysis.

2. Notations and Assumptions

2.1. Notations

First of all, we set $A(x, \alpha) := a(x, \alpha)^{\otimes 2}$, $a(x) := a(x, \alpha_*)$, $A(x) := A(x, \alpha_*)$ and $b(x) := b(x, \beta_*)$. We also give the notation for a matrix-valued function $\mathbb{G}(x, \alpha|\rho)$ such that $\mathbb{G}^{(i,j)}(x, \alpha|\rho) := A^{(i,j)}(x, \alpha) f_{\mathbb{G}}(\rho^{(i)}, \rho^{(j)})$, where

$$f_{\mathbb{G}}(\rho^{(i)}, \rho^{(j)}) := \begin{cases} 1 & \text{if } \rho^{(i)} = \rho^{(j)} = 0, \\ 1 - \frac{\rho^{(j)}}{2} & \text{if } \rho^{(i)} = 0, \rho^{(j)} \in (0, 1], \\ \frac{1}{2\rho^{(j)}} & \text{if } \rho^{(i)} = 0, \rho^{(j)} \in (1, \bar{\rho}], \\ 1 - \frac{\rho^{(i)}}{2} & \text{if } \rho^{(i)} \in (0, 1], \rho^{(j)} = 0, \\ \frac{1}{2\rho^{(i)}} & \text{if } \rho^{(i)} \in (1, \bar{\rho}], \rho^{(j)} = 0, \\ \frac{-3(\rho^{(i)})^2\rho^{(j)} + 3\rho^{(i)}(\rho^{(j)})^2 + 6\rho^{(i)}\rho^{(j)} - 2(\rho^{(j)})^3}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(i)}, \rho^{(j)} \in (0, 1], \rho^{(i)} > \rho^{(j)}, \\ \frac{3(\rho^{(i)})^2\rho^{(j)} - 3\rho^{(i)}(\rho^{(j)})^2 + 6\rho^{(i)}\rho^{(j)} - 2(\rho^{(i)})^3}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(i)}, \rho^{(j)} \in (0, 1], \rho^{(i)} \leq \rho^{(j)}, \\ \frac{3(\rho^{(j)})^2 + 3\rho^{(j)} - (\rho^{(i)})^3}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(i)} \in (1, \bar{\rho}], \rho^{(j)} \in (0, 1], \rho^{(i)} > \rho^{(j)} + 1, \\ \frac{6\rho^{(j)} - 1}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(i)}, \rho^{(j)} \in (1, \bar{\rho}], \rho^{(i)} > \rho^{(j)} + 1, \\ \frac{(\rho^{(i)} - \rho^{(j)})^3 - 3(\rho^{(i)})^2 + 6\rho^{(i)}\rho^{(j)} + 3\rho^{(i)} - 1 - (\rho^{(j)})^3}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(i)} \in (1, \bar{\rho}], \rho^{(j)} \in (0, 1], \rho^{(i)} \leq \rho^{(j)} + 1, \\ \frac{(\rho^{(i)} - \rho^{(j)})^3 - 3(\rho^{(i)})^2 + 6\rho^{(i)}\rho^{(j)} + 3\rho^{(i)} - 3(\rho^{(j)})^2 + 3\rho^{(j)} - 2}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(i)}, \rho^{(j)} \in (1, \bar{\rho}], \rho^{(j)} < \rho^{(i)} \leq \rho^{(j)} + 1, \\ \frac{-(\rho^{(i)} - \rho^{(j)})^3 + 6\rho^{(i)}\rho^{(j)} - (\rho^{(i)})^3 - 3(\rho^{(j)})^2 + 3\rho^{(j)} - 1}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(i)} \in (0, 1], \rho^{(j)} \in (1, \bar{\rho}], \rho^{(j)} \leq \rho^{(i)} + 1, \\ \frac{-(\rho^{(i)} - \rho^{(j)})^3 - 3(\rho^{(i)})^2 - 3(\rho^{(j)})^2 + 6\rho^{(i)}\rho^{(j)} + 3\rho^{(i)} + 3\rho^{(j)} - 2}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(i)}, \rho^{(j)} \in (1, \bar{\rho}], \rho^{(i)} \leq \rho^{(j)} \leq \rho^{(i)} + 1, \\ \frac{3(\rho^{(i)})^2 + 3\rho^{(i)} - (\rho^{(i)})^3}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(i)} \in (0, 1], \rho^{(j)} \in (1, \bar{\rho}], \rho^{(j)} > \rho^{(i)} + 1, \\ \frac{6\rho^{(i)} - 1}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(i)}, \rho^{(j)} \in (1, \bar{\rho}], \rho^{(j)} > \rho^{(i)} + 1. \end{cases}$$

The continuity of the function $f_{\mathbb{G}}$ is shown in the supplementary material. For the detailed discussion of the necessity of $f_{\mathbb{G}}$, see Remark 4.

In addition, we also give the notation used throughout this paper.

- For every matrix A , A^T is the transpose of A , and $A^{\otimes 2} := AA^T$.
- For every set of matrices A and B of the same size, $A[B] := \text{tr}(AB^T)$. Moreover, for any $m \in \mathbf{N}$, $A \in \mathbf{R}^m \otimes \mathbf{R}^m$ and $u, v \in \mathbf{R}^m$, $A[u, v] := v^T Au$.
- $v^{(\ell)}$ and $A^{(\ell_1, \ell_2)}$ denote the ℓ -th element of a vector v and the (ℓ_1, ℓ_2) -th one of a matrix A , respectively.
- For any vector v and any matrix A , $|v| := \sqrt{\text{tr}(v^T v)}$ and $\|A\| := \sqrt{\text{tr}(A^T A)}$.
- $(\Omega, P, \mathcal{F}, \mathcal{F}_t)$ denotes the stochastic basis, where $\mathcal{F}_t := \sigma(x_{-\lambda}, w_s : s \leq t)$.
- $\mu(f(\cdot))$ denotes the integral of a μ -integrable function f where μ is a measure.

2.2. Assumptions

With respect to X_t , we assume the following conditions.

[A1] (i) For a constant C , for all $x_1, x_2 \in \mathbf{R}^d$,

$$\sup_{\alpha \in \Theta_1} \|a(x_1, \alpha) - a(x_2, \alpha)\| + \sup_{\beta \in \Theta_2} |b(x_1, \beta) - b(x_2, \beta)| \leq C|x_1 - x_2|.$$

(ii) For all $p \geq 0$, $\sup_{t \geq -\lambda} \mathbf{E}_{\theta_*} [|X_t|^p] < \infty$.

- (iii) There exists a unique invariant measure ν_0 on $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ and for all $p \geq 1$ and $f \in L^p(\nu_0)$ with polynomial growth,

$$\frac{1}{T} \int_{-\lambda}^T f(X_t) dt \xrightarrow{P} \int_{\mathbf{R}^d} f(x) \nu_0(dx).$$

Remark 1. For sufficient conditions of these regularity ones, please see [A1] and Remark 1 in Uchida and Yoshida [7].

- [A2] There exists $C > 0$ such that $a : \mathbf{R}^d \times \Theta_1 \rightarrow \mathbf{R}^d \otimes \mathbf{R}^r$ and $b : \mathbf{R}^d \times \Theta_2 \rightarrow \mathbf{R}^d$ have continuous derivatives satisfying

$$\begin{aligned} \sup_{\alpha \in \Theta_1} \left| \partial_x^j \partial_\alpha^i a(x, \alpha) \right| &\leq C(1 + |x|)^C, \quad 0 \leq i \leq 2, \quad 0 \leq j \leq 2, \\ \sup_{\beta \in \Theta_2} \left| \partial_x^j \partial_\beta^i b(x, \beta) \right| &\leq C(1 + |x|)^C, \quad 0 \leq i \leq 2, \quad 0 \leq j \leq 2. \end{aligned}$$

With the invariant measure ν_0 , $\zeta := [\rho^T, \theta^T]^T$, and ζ_* denoting the true value of ζ , we define

$$\begin{aligned} \mathbb{V}_1(\alpha|\zeta_*) &:= - \int_{\mathbf{R}^d} \|\mathbb{G}(x, \alpha|\rho_*) - \mathbb{G}(x, \alpha_*|\rho_*)\|^2 \nu_0(dx), \\ \mathbb{V}_2(\beta|\zeta_*) &:= - \int_{\mathbf{R}^d} |b(x, \beta) - b(x, \beta_*)|^2 \nu_0(dx). \end{aligned}$$

For these functions, let us assume the following identifiability conditions hold.

- [A3] There exist $\chi_1(\alpha_*) > 0$ and $\chi'_1(\beta_*) > 0$ such that for all $\alpha \in \Theta_1$ and $\beta \in \Theta_2$, $\mathbb{V}_1(\alpha|\zeta_*) \leq -\chi_1(\alpha_*)|\alpha - \alpha_*|^2$ and $\mathbb{V}_2(\beta|\zeta_*) \leq -\chi'_1(\beta_*)|\beta - \beta_*|^2$.

3. Estimation and Test of the Smoothing Parameter

In this section, we discuss the case where the smoothing parameter ρ of the kernel function V_{ρ, h_n} is unknown. The estimation is significant for estimation of α and β since we utilise the estimate for ρ in quasi-likelihood functions of α and β . The test problem for hypotheses $H_0 : \rho = \mathbf{0}$ and $H_1 : \rho \neq \mathbf{0}$ is also important to examine whether our framework of convolutional observation is meaningful.

3.1. Estimation of the Smoothing Parameter

For simplicity of notation, let us consider the case $\bar{\rho} > 2$; otherwise the discussion is quite parallel. We should note that for all $i = 1, \dots, d$,

$$\mathbb{G}^{(i,i)}(x|\rho) = \begin{cases} A^{(i,i)}(x) & \text{if } \rho^{(i)} = 0, \\ A^{(i,i)}(x) \left(1 - \frac{\rho^{(i)}}{3}\right) & \text{if } \rho^{(i)} \in (0, 1], \\ A^{(i,i)}(x) \left(\frac{1}{\rho^{(i)}} - \frac{1}{3(\rho^{(i)})^2}\right) & \text{if } \rho^{(i)} \in (1, \bar{\rho}]. \end{cases}$$

Let us consider the estimation of $\rho^{(i)}$ with using the next statistics: the full quadratic variation

$$\frac{1}{nh_n} \sum_{k=1}^n \left(\bar{X}^{(i)}_{kh_n, n} - \bar{X}^{(i)}_{(k-1)h_n, n} \right)^2 \xrightarrow{P} \begin{cases} \nu_0(A^{(i,i)}(\cdot)) & \text{if } \rho_*^{(i)} = 0, \\ \nu_0(A^{(i,i)}(\cdot)) \left(1 - \frac{\rho_*^{(i)}}{3}\right) & \text{if } \rho_*^{(i)} \in (0, 1], \\ \nu_0(A^{(i,i)}(\cdot)) \left(\frac{1}{\rho_*^{(i)}} - \frac{1}{3(\rho_*^{(i)})^2}\right) & \text{if } \rho_*^{(i)} \in (1, \bar{\rho}], \end{cases}$$

because of Proposition 3 in supplementary material Appendix A, and the reduced quadratic variation defined as $\frac{1}{nh_n} \sum_{2 \leq 2k \leq n} (\bar{X}_{2kh_n,n}^{(i)} - \bar{X}_{(2k-2)h_n,n}^{(i)})^2$ converges in probability as follows.

Lemma 1. Under [A1], we have the convergence in probability such that

$$\frac{1}{nh_n} \sum_{2 \leq 2k \leq n} (\bar{X}_{2kh_n,n}^{(i)} - \bar{X}_{(2k-2)h_n,n}^{(i)})^2 \xrightarrow{P} \begin{cases} \nu_0(A^{(i,i)}(\cdot)) & \text{if } \rho_\star^{(i)} = 0, \\ \nu_0(A^{(i,i)}(\cdot)) \left(1 - \frac{\rho_\star^{(i)}}{6}\right) & \text{if } \rho_\star^{(i)} \in (0, 2], \\ \nu_0(A^{(i,i)}(\cdot)) \left(\frac{2}{\rho_\star^{(i)}} - \frac{4}{3(\rho_\star^{(i)})^2}\right) & \text{if } \rho_\star^{(i)} \in (2, \bar{\rho}]. \end{cases}$$

Then we define the ratio of those statistics such that

$$\begin{aligned} R_n^{(i)} &:= \left(\frac{1}{nh_n} \sum_{k=1}^n (\bar{X}_{kh_n,n}^{(i)} - \bar{X}_{(k-1)h_n,n}^{(i)})^2\right) \left(\frac{1}{nh_n} \sum_{2 \leq 2k \leq n} (\bar{X}_{2kh_n,n}^{(i)} - \bar{X}_{(2k-2)h_n,n}^{(i)})^2\right)^{-1} \\ &\xrightarrow{P} \begin{cases} 1 & \text{if } \rho_\star^{(i)} = 0, \\ \left(1 - \frac{\rho_\star^{(i)}}{3}\right) \left(1 - \frac{\rho_\star^{(i)}}{6}\right)^{-1} & \text{if } \rho_\star^{(i)} \in (0, 1], \\ \left(\frac{1}{\rho_\star^{(i)}} - \frac{1}{3(\rho_\star^{(i)})^2}\right) \left(1 - \frac{\rho_\star^{(i)}}{6}\right)^{-1} & \text{if } \rho_\star^{(i)} \in (1, 2], \\ \left(\frac{1}{\rho_\star^{(i)}} - \frac{1}{3(\rho_\star^{(i)})^2}\right) \left(\frac{2}{\rho_\star^{(i)}} - \frac{4}{3(\rho_\star^{(i)})^2}\right)^{-1} & \text{if } \rho_\star^{(i)} \in (2, \bar{\rho}] \end{cases} \\ &= \begin{cases} 1 & \text{if } \rho^{(i)} = 0, \\ (6 - 2\rho_\star^{(i)}) (6 - \rho_\star^{(i)})^{-1} & \text{if } \rho_\star^{(i)} \in (0, 1], \\ (6\rho_\star^{(i)} - 2) \left(6(\rho_\star^{(i)})^2 - (\rho_\star^{(i)})^3\right)^{-1} & \text{if } \rho_\star^{(i)} \in (1, 2], \\ (3\rho_\star^{(i)} - 1) (6\rho_\star^{(i)} - 4)^{-1} & \text{if } \rho_\star^{(i)} \in (2, \bar{\rho}], \end{cases} \\ &=: R(\rho_\star^{(i)}), \end{aligned}$$

where R has the next property.

Lemma 2. R is a $[(3\bar{\rho} - 1)(6\bar{\rho} - 4)^{-1}, 1]$ -valued monotonically decreasing continuous function, and has a continuous inverse $R^{-1} : [(3\bar{\rho} - 1)(6\bar{\rho} - 4)^{-1}, 1] \rightarrow [0, \bar{\rho}]$.

We define $\hat{\rho}_n$ such that

$$\hat{\rho}_n^{(i)} := \begin{cases} 0 & \text{if } R_n^{(i)} > 1, \\ R^{-1}(R_n^{(i)}) & \text{if } R_n^{(i)} \in [(3\bar{\rho} - 1)(6\bar{\rho} - 4)^{-1}, 1], \\ \bar{\rho} & \text{if } R_n^{(i)} < (3\bar{\rho} - 1)(6\bar{\rho} - 4)^{-1}, \end{cases}$$

and then continuous mapping theorem for convergence in probability verifies the next result.

Theorem 1. Under [A1], $\hat{\rho}_n$ has consistency, i.e., $\hat{\rho}_n \xrightarrow{P} \rho_\star$.

Remark 2. We can compute $y = R^{-1}(x)$, $x \in \left[(3\bar{\rho} - 1)(6\bar{\rho} - 4)^{-1}, 1 \right]$ by solving the following equations:

$$\begin{aligned} (i) \quad x &= (6 - 2y)(6 - y)^{-1} && \text{if } x \in (4/5, 1], \\ (ii) \quad x &= (6 - 2y)(6y^2 - y^3)^{-1} && \text{if } x \in (5/8, 4/5], \\ (iii) \quad x &= (3y - 1)(6y - 4)^{-1} && \text{if } x \in \left[(3\bar{\rho} - 1)(6\bar{\rho} - 4)^{-1}, 5/8 \right]. \end{aligned}$$

3.2. Test for Smoothed Observation

For all $i = 1, \dots, d$, we consider the next hypothesis testing:

$$H_0 : \rho^{(i)} = 0, \quad H_1 : \rho^{(i)} > 0.$$

Let us consider the following test statistic:

$$\begin{aligned} \mathcal{T}_{i,n} &:= \sqrt{\frac{n}{3nh_n^2 \sum_{k=1}^n \left(\bar{X}_{kh_n,n}^{(i)} - \bar{X}_{(k-1)h_n,n}^{(i)} \right)^4}} \\ &\times \left(\frac{1}{nh_n} \sum_{k=1}^n \left(\bar{X}_{kh_n,n}^{(i)} - \bar{X}_{(k-1)h_n,n}^{(i)} \right)^2 - \frac{1}{nh_n} \sum_{2 \leq 2k \leq n} \left(\bar{X}_{2kh_n,n}^{(i)} - \bar{X}_{(2k-2)h_n,n}^{(i)} \right)^2 \right) \\ &= \sqrt{\frac{3/2}{\sum_{k=1}^n \left(\bar{X}_{kh_n,n}^{(i)} - \bar{X}_{(k-1)h_n,n}^{(i)} \right)^4}} \\ &\times \left(\sum_{k=1}^n \left(\bar{X}_{k,n}^{(i)} - \bar{X}_{k-1,n}^{(i)} \right)^2 - \sum_{2 \leq 2k \leq n} \left(\bar{X}_{2k,n}^{(i)} - \bar{X}_{(2k-2),n}^{(i)} \right)^2 \right), \end{aligned}$$

and we abbreviate $\mathcal{T}_{i,n}$ to \mathcal{T}_n if $d = 1$. Under H_0 , we have the next result.

Theorem 2. Under H_0 and [A1], we have the convergence in law such that

$$\mathcal{T}_{i,n} \rightarrow^{\mathcal{L}} N(0, 1).$$

We also obtain the result to support the consistency of the test.

Theorem 3. Under H_1 and [A1], we have convergence such that for any $c \in \mathbf{R}$,

$$P(\mathcal{T}_{i,n} < c) \rightarrow 1.$$

Remark 3. These results are intuitive since the quadratic variation and the reduced one with some appropriate scaling should converge to the same value if H_0 holds and the quadratic variation with some appropriate scaling should converge to the value which is smaller than the value which the reduced one with scaling converge to.

Hence when we set the significance level $\alpha_{\text{sig}} \in (0, 1)$, then we have the rejection region

$$\mathcal{T}_{i,n} < \Phi^{-1}(\alpha_{\text{sig}})$$

where Φ is the distribution function of the standard Gaussian distribution. Theorem 3 supports the consistency of the test.

This test is essential in terms of examining the validity to consider the scheme of convolutional observation: if $\rho = \mathbf{0}$, then the ordinary observation scheme can be applied, but if $\rho \neq \mathbf{0}$, then we have the motivation to consider the convolutional observation scheme.

4. Least Square Estimation of the Diffusion and Drift Parameters

Let us set the least-square quasi-likelihood functions such that

$$\mathbb{H}_{1,n}(\alpha|\rho) := - \sum_{k=1}^n \left\| \frac{1}{h_n} \left(\bar{X}_{kh_n,n} - \bar{X}_{(k-1)h_n,n} \right)^{\otimes 2} - \mathbb{G} \left(\bar{X}_{(k-1)h_n,n}, \alpha|\rho \right) \right\|^2,$$

$$\mathbb{H}_{2,n}(\beta|\rho) := - \sum_{k=\lceil \max_i \rho^{(i)} \rceil + 2}^n \frac{1}{h_n} \left| \bar{X}_{kh_n,n} - \bar{X}_{(k-1)h_n,n} - h_n b \left(\bar{X}_{(k-2-\lceil \max_i \rho^{(i)} \rceil)h_n,n}, \beta \right) \right|^2,$$

and the least-square estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ satisfying

$$\mathbb{H}_{1,n}(\hat{\alpha}_n|\rho_\star) = \sup_{\alpha \in \Theta_1} \mathbb{H}_{1,n}(\alpha|\rho_\star), \quad \mathbb{H}_{2,n}(\hat{\beta}_n|\rho_\star) = \sup_{\beta \in \Theta_2} \mathbb{H}_{2,n}(\beta|\rho_\star).$$

when ρ_\star is known, and

$$\mathbb{H}_{1,n}(\hat{\alpha}_n|\hat{\rho}_n) = \sup_{\alpha \in \Theta_1} \mathbb{H}_{1,n}(\alpha|\hat{\rho}_n), \quad \mathbb{H}_{2,n}(\hat{\beta}_n|\hat{\rho}_n) = \sup_{\beta \in \Theta_2} \mathbb{H}_{2,n}(\beta|\hat{\rho}_n).$$

when ρ_\star is unknown.

Theorem 4. Under [A1]–[A3], $\hat{\alpha}_n$ and $\hat{\beta}_n$ are consistent, i.e., $\hat{\alpha}_n \rightarrow^P \alpha_\star$ and $\hat{\beta}_n \rightarrow^P \beta_\star$.

Remark 4. The function \mathbb{G} and $f_{\mathbb{G}}$ are indeed complex and confusing; hence, we can consider some alternative estimation methods with subsampling or pre-averaging. However, these methods also have the problem what size of subsampling or pre-averaging is proper and the result of the estimation can be dependent on tuning the subsampling size or pre-averaging one. Therefore, our work proposes the estimation method which uses the observation without subsampling or pre-averaging.

5. Simulations

In this simulation section, we only consider the case where ρ is unknown and should be estimated by data with the method proposed in Section 3.

5.1. 1-Dimensional Simulation

We examine the following 1-dimensional stochastic differential equation whose solution is a 1-dimensional Ornstein–Uhlenbeck (OU) process:

$$dX_t = \left(\beta^{(1)} X_t + \beta^{(2)} \right) dt + \alpha dw_t, \quad X_{-\lambda} = 0,$$

$\alpha \in \Theta_1 := [0.01, 10]$, $\beta \in \Theta_2 := [-10, -0.01] \times [-10, 10]$, and $\lambda = 10^{-7/3}$. The procedure of the simulation is as follows: in the first place we iterate an approximated OU process by Euler–Maruyama scheme, for example, see [26] with simulation parameters $n_{\text{sim}} = 10^{5+m}$, $h_{\text{sim}} = 10^{-10/3-m}$, $T_{\text{sim}} = 10^{5/3}$ where $m \in \mathbf{N}$ is a parameter to determine the precision of approximation; secondly, we give the approximation of convolution by summation such that

$$\bar{X}_{ih_n,n} \approx \begin{cases} \frac{1}{[10^m \rho]} \sum_{k=0}^{10^m-1} X_{ih_n-kh_{\text{sim}}} & \text{if } [10^m \rho] \geq 1, \\ X_{ih_n} & \text{if } [10^m \rho] < 0, \end{cases}$$

where $i = 0, \dots, n$, the sampling frequency $h_n = 10^{-10/3}$ and $n = 10^{5/3}$. In this Section 5.1, we fix the true value of α and β as $\alpha_\star = 3$ and $\beta_\star = [-2, 1]^T$, and change the true value of $\rho \in \Theta_\rho := [0, 100]$ to see the corresponding changes of performance of estimation for ζ , and test for ρ in comparison to estimation

by an existent method called local Gaussian approximation (LGA) for parametric estimation of discretely observed diffusion processes, e.g., see [4] which does not concern convolutional observation. All the numbers of iterations for different ρ 's are 1000.

In the first place, we see the estimation and test with small values of ρ_* such that $\rho_* = 0, 0.1, 0.2, \dots, 1$ to observe how the performance of statistics changes by difference in ρ . Table 1 summarises the results of simulation of $\hat{\rho}_n$ for ρ 's with respective empirical means and root mean square error (RMSE).

Table 1. Estimation performance of ρ with small ρ .

ρ	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
mean	0.00990	0.0971	0.198	0.298	0.398	0.498	0.598	0.699	0.799	0.899	0.999
RMSE	(0.0182)	(0.0256)	(0.0235)	(0.0215)	(0.0197)	(0.0180)	(0.0164)	(0.0150)	(0.0135)	(0.0123)	(0.0110)

We can see the proposed estimator $\hat{\rho}_n$ works well for small ρ . With respect to the performance of the test statistic \mathcal{T}_n proposed in Section 3.2, Table 2 shows the empirical ratio of the number of iterations whose \mathcal{T}_n is lower than some typical critical values where Φ indicates the distribution function of 1-dimensional standard Gaussian distribution as well as the maximum value of \mathcal{T}_n in 1000 iterations.

Table 2. Empirical ratio of test statistic \mathcal{T}_n less than some critical values, and the maximum value of \mathcal{T}_n in 1000 iterations.

	Empirical Ratio of \mathcal{T}_n Less Than ...					Max. of \mathcal{T}_n
	$\Phi^{-1}(0.10)$	$\Phi^{-1}(0.05)$	$\Phi^{-1}(0.025)$	$\Phi^{-1}(0.01)$	$\Phi^{-1}(0.001)$	
$\rho = 0.0$	0.101	0.053	0.025	0.005	0.000	3.060
$\rho = 0.1$	0.989	0.980	0.966	0.914	0.759	-0.710
$\rho = 0.2$	1.000	1.000	1.000	1.000	1.000	-4.593
$\rho = 0.3$	1.000	1.000	1.000	1.000	1.000	-9.341
$\rho = 0.4$	1.000	1.000	1.000	1.000	1.000	-13.985
$\rho = 0.5$	1.000	1.000	1.000	1.000	1.000	-19.152
$\rho = 0.6$	1.000	1.000	1.000	1.000	1.000	-24.816
$\rho = 0.7$	1.000	1.000	1.000	1.000	1.000	-30.848
$\rho = 0.8$	1.000	1.000	1.000	1.000	1.000	-37.557
$\rho = 0.9$	1.000	1.000	1.000	1.000	1.000	-44.829
$\rho = 1.0$	1.000	1.000	1.000	1.000	1.000	-52.759

Even for $\rho = 0.1$, the simulation result supports the theoretical discussion of the test with consistency. Because $\Phi(10^{-16}) = -8.222$, all the iterations with $\rho \geq 0.3$ result in rejection of H_0 with substantially significance level 10^{-16} . Let us see the estimation for α and β by our proposal method and that by the LGA in Table 3.

Table 3. Estimation of θ by the proposed method and LGA with small ρ .

		The Proposed Method			LGA		
		α	$\beta^{(1)}$	$\beta^{(2)}$	α	$\beta^{(1)}$	$\beta^{(2)}$
True Value		3.0	−2.0	1.0	3.0	−2.0	1.0
$\rho = 0.0$	mean	3.004	−2.091	1.036	2.999	−2.095	1.037
	RMSE	(0.0109)	(0.318)	(0.497)	(0.00679)	(0.320)	(0.497)
$\rho = 0.1$	mean	2.999	−2.091	1.035	2.949	−2.026	1.003
	RMSE	(0.0146)	(0.319)	(0.496)	(0.0509)	(0.297)	(0.480)
$\rho = 0.2$	mean	2.998	−2.091	1.035	2.898	−1.955	0.967
	RMSE	(0.0142)	(0.319)	(0.496)	(0.102)	(0.290)	(0.464)
$\rho = 0.3$	mean	2.998	−2.092	1.036	2.846	−1.885	0.932
	RMSE	(0.0139)	(0.319)	(0.497)	(0.155)	(0.299)	(0.452)
$\rho = 0.4$	mean	2.998	−2.091	1.036	2.792	−1.815	0.897
	RMSE	(0.0135)	(0.319)	(0.497)	(0.208)	(0.324)	(0.442)
$\rho = 0.5$	mean	2.998	−2.092	1.036	2.738	−1.744	0.862
	RMSE	(0.0132)	(0.319)	(0.497)	(0.262)	(0.361)	(0.436)
$\rho = 0.6$	mean	2.998	−2.091	1.036	2.683	−1.674	0.827
	RMSE	(0.0129)	(0.319)	(0.497)	(0.317)	(0.408)	(0.434)
$\rho = 0.7$	mean	2.998	−2.092	1.036	2.626	−1.604	0.792
	RMSE	(0.0126)	(0.319)	(0.497)	(0.374)	(0.460)	(0.434)
$\rho = 0.8$	mean	2.998	−2.092	1.036	2.568	−1.534	0.757
	RMSE	(0.0124)	(0.319)	(0.496)	(0.432)	(0.517)	(0.439)
$\rho = 0.9$	mean	2.998	−2.092	1.036	2.509	−1.464	0.722
	RMSE	(0.0121)	(0.319)	(0.497)	(0.491)	(0.578)	(0.445)
$\rho = 1.0$	mean	2.998	−2.091	1.036	2.449	−1.394	0.687
	RMSE	(0.0119)	(0.319)	(0.497)	(0.551)	(0.640)	(0.456)

Note that the biases of the estimation by LGA increase as the true value of ρ gets larger, while the estimation by our proposal method is not influenced by the true value of ρ . This result of the simulation supports the theoretical discussion in Section 4 stating the consistency of $\hat{\theta}_n$, and necessity to consider the convolutional observation scheme where the LGA method does not work properly.

Secondly, we consider the estimation and test with large ρ_* such that $\rho_* = 10, 15, 20$ to see if our proposal methods work even for large ρ . We note that the maximum values of \mathcal{T}_n for $\rho = 10, 15, 20$ in 1000 iterations are $-55.091, -68.462$ and -79.105 , and hence we can detect the smoothed observation easily. Table 4 shows the empirical means and RMSEs of $\hat{\rho}_n$ for $\rho = 10, 15, 20$ and we can see that the RMSEs increase as ρ 's increase; it indicates the difficulty to estimate ρ accurately when ρ_* is large.

Table 4. The performance of $\hat{\rho}_n$ for $\rho = 10, 15, 20$ in 1000 iterations.

	$\rho = 10$	$\rho = 15$	$\rho = 20$
mean of $\hat{\rho}_n$	9.919	14.980	19.751
RMSE of $\hat{\rho}_n$	(0.145)	(0.240)	(0.409)

Table 5 summarises the estimation for θ by means and RMSE, and tells us that although the large RMSE of $\hat{\rho}_n$ results in increase of RMSE of $\hat{\alpha}_n$, estimation by our method is substantially better than that by LGA.

Table 5. Estimation of θ by the proposed method with large ρ .

		The Proposed Method			LGA		
		α	$\beta^{(1)}$	$\beta^{(2)}$	α	$\beta^{(1)}$	$\beta^{(2)}$
True Value		3.0	−2.0	1.0	3.0	−2.0	1.0
$\rho = 10$	mean	2.989	−2.101	1.030	0.933	−0.204	0.0811
	RMSE	(0.0347)	(0.323)	(0.496)	(2.067)	(1.796)	(0.920)
$\rho = 15$	mean	2.996	−2.095	1.027	0.765	−0.138	0.0473
	RMSE	(0.0475)	(0.321)	(0.495)	(2.235)	(1.862)	(0.953)
$\rho = 20$	mean	2.977	−2.090	1.024	0.664	−0.104	0.0302
	RMSE	(0.0526)	(0.319)	(0.493)	(2.336)	(1.896)	(0.970)

5.2. 2-Dimensional Simulation

We consider the following 2-dimensional stochastic differential equation whose solution is a 2-dimensional OU process:

$$d \begin{bmatrix} X_t^{(1)} \\ X_t^{(2)} \end{bmatrix} = \left(\begin{bmatrix} \beta^{(1)} & \beta^{(2)} \\ \beta^{(4)} & \beta^{(5)} \end{bmatrix} \begin{bmatrix} X_t^{(1)} \\ X_t^{(2)} \end{bmatrix} + \begin{bmatrix} \beta^{(3)} \\ \beta^{(6)} \end{bmatrix} \right) dt + \begin{bmatrix} \alpha^{(1)} & \alpha^{(2)} \\ \alpha^{(2)} & \alpha^{(3)} \end{bmatrix} dw_t, X_{-\lambda} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

$\lambda = 10^{-7/3}$. The simulation is conducted with the settings as follows: firstly, we iterate the OU process by Euler–Maruyama scheme with the simulation sample size $n_{sim} = 10^{5+m}$, $T_{sim} = 10^{5/3}$ and discretisation step $h_{sim} = 10^{-10/3-m}$, where $m = 2$ is the precision parameter for approximation of convolution; in the second place, we approximate the convoluted process with summation such that

$$\bar{X}_{ih_n, n}^{(j)} \cong \begin{cases} \frac{1}{[10^m \rho^{(j)}]} \sum_{k=0}^{10^m-1} X_{ih_n - kh_{sim}}^{(j)} & \text{if } [10^m \rho^{(j)}] \geq 1, \\ X_{ih_n}^{(j)} & \text{if } [10^m \rho^{(j)}] < 0, \end{cases}$$

where $i = 0, \dots, n, j = 1, 2$, the sampling scheme for inference is defined as $n = 10^5$ and $h_n = 10^{-10/3}$; the true value of ρ, α and β are set as $\rho_* = [2, 4]^T, \alpha_* = [2, 0, 3]^T, \beta_* = [-2, -0.4, 0, 0.1, -3, 5]^T$; the parameter spaces are defined as $\Theta_\rho = [0, 10]^2, \Theta_1 = [1 + 10^{-8}, 10] \times [-1 + 10^{-8}, 1 - 10^{-8}] \times [1 + 10^{-8}, 10]$, and $\Theta_2 = [-10, 10]^6$; the total iteration number is set to 1000.

Table 6 summarises the estimation for ρ with the method proposed in Section 3 (the inverse of r is computationally obtained) with empirical means and empirical RMSEs of $\hat{\rho}_n$ in 1000 iterations. We can see that $\hat{\rho}_n$ is sufficiently precise to estimate the true value of ρ indeed in this result, which is significant to estimate the other parameters α and β .

Table 6. Summary for ρ estimate.

	$\rho^{(1)}$	$\rho^{(2)}$
true value	2.0	4.0
empirical mean	1.988	3.966
empirical RMSE	(0.0207)	(0.0514)

We also note that the maximum values of the test statistics for smoothed observation proposed in Section 3.2 in 1000 iterations are -17.947 and -33.159 for each axis. The p -value for them are smaller than 10^{-16} ; therefore, we can conclude that it is possible to detect the smoothed observation with the proposed test statistic in the case $\rho^{(i)} = 2.0$ if $d = 2$ from this result.

With respect to the estimation for α and β , we compare the estimates by our proposal method with that by LGA which does not concern convolutional observation. Table 7 is the summary for α estimate by both the methods:

Table 7. Summary for α estimate.

		$\alpha^{(1)}$	$\alpha^{(2)}$	$\alpha^{(3)}$
True Value		2.0	0.0	3.0
Our proposal	mean	1.993	0.000256	2.992
	RMSE	(0.0115)	(0.00739)	(0.0213)
LGA	mean	1.295	−0.00320	1.442
	RMSE	(0.705)	(0.0154)	(1.558)

We can see that the estimation precision for α by our proposal outperforms those by LGA. This results support validity of our estimation method if we have convolutional observation for diffusion processes. Regarding β , the simulation result is summarised in Table 8:

Table 8. Summary for β estimate.

		$\beta^{(1)}$	$\beta^{(2)}$	$\beta^{(3)}$	$\beta^{(4)}$	$\beta^{(5)}$	$\beta^{(6)}$
True Value		−2.0	−0.4	0.0	0.1	−3.0	5.0
Our proposal	mean	−2.137	−0.408	−0.0439	0.0788	−3.103	5.091
	RMSE	(0.362)	(0.252)	(0.540)	(0.473)	(0.399)	(0.777)
LGA	mean	−0.917	0.340	−0.326	−0.696	0.221	1.243
	RMSE	(1.093)	(0.802)	(0.386)	(0.804)	(3.242)	(3.765)

Though the estimation for $\beta^{(3)}$ by our method has the smaller bias in comparison to that by LGA, the RMSE of our method is larger than that of LGA; in the estimation for other parameters, our proposal method outperforms the method by LGA. We can conclude that our proposal for estimation of α and β concerning convolutional observation performs better than that not considering this observation scheme.

6. Real Data Analysis

In this section, we analyse the EEG dataset named S02E.mat provided in “2. Two class motor imagery (002-2014)” of BNCI Horizon 2020 [20]. The datasets including S02E.mat are also studied by Steryl et al. [27].

6.1. Estimation and Test for the Smoothing Parameters

In the first place, we pick up the first 15 axes of the dataset and compute $\hat{\rho}_n$ and \mathcal{T}_n proposed in Section 3.1 and Section 3.2 respectively. The results are shown in Table 9.

We can observe that all the 15 time series data have the smoothing parameter $\rho > 0$ with statistical significance when we assume ordinary significance levels. These results motivate us to use our methods for parametric estimation proposed in Section 4 when we fit stochastic differential equations for these data.

Table 9. The values of $\hat{\rho}_n$ and \mathcal{T}_n for the first 15 axes of S02.mat by BNCI Horizon 2020 [20].

	1st axis	2nd axis	3rd axis	4th axis	5th axis
$\hat{\rho}_n$	0.449	1.037	0.894	0.736	0.937
\mathcal{T}_n	−20.398	−58.631	−46.649	−35.201	−49.741
	6th axis	7th axis	8th axis	9th axis	10th axis
$\hat{\rho}_n$	0.951	0.971	1.017	0.958	0.967
\mathcal{T}_n	−51.392	−52.607	−55.455	−51.221	−51.797
	11th axis	12th axis	13th axis	14th axis	15th axis
$\hat{\rho}_n$	0.949	0.649	0.952	0.977	0.932
\mathcal{T}_n	−50.457	−30.094	−50.633	−50.978	−48.842

6.2. Parametric Estimation for a Diffusion Process

We fit a 1-dimensional OU process for the time series data in the 2nd column of the data file S02E.mat with 512 Hz observation for 222 s (the plot of the path can be seen in Figure 1), whose $\hat{\rho}_n = 1.037$ is the largest among those for the 15 axes and it is larger than 0 with statistical significance. According to the simulation result shown in Section 5.1, this size of the smoothing parameter gives critical biases when we estimate α and β with LGA method not concerning convolutional observation scheme.

The stochastic differential equation with parameters $\alpha \in \Theta_1 := [0.01, 200]$ and $\beta \in \Theta_2 := [-100, -0.01] \times [-100, 100]$ is defined as follows:

$$dX_t = \left(\beta^{(1)} X_t + \beta^{(2)} \right) dt + \alpha dw_t, \quad X_{-\lambda} = x_{-\lambda}.$$

We set 5 s as the time unit: hence $n = 113,664$ and $h_n = 1 / (5 \times 512)$. If we fit the OU process with the LGA method, i.e., we do not concern convolutional observation scheme, we obtain the fitting result such that

$$dX_t = ((-17.378) X_t + (-1.091)) dt + (122.892) dw_t, \quad X_{-\lambda} = x_{-\lambda}.$$

In the next place, we fit α and β with the least square method proposed in Section 4, and then we have the next fitting result:

$$dX_t = ((-2.146) X_t + (0.552)) dt + (151.919) dw_t, \quad X_{-\lambda} = x_{-\lambda}.$$

It is worth noting that this fitting result is substantially different to that by LGA as shown above: hence these results indicate the significance to examine if the observation is convoluted with the smoothing parameter $\rho > 0$ and otherwise the estimation is strongly biased.

7. Summary

We have discussed the convolutional observation scheme which deals with the smoothness of observation in comparison to ordinary diffusion processes. The first contribution is to propose this new observation scheme with the statistical test to confirm whether this scheme is valid in real data. The second one is to prove consistency of the estimator $\hat{\rho}_n$ for the smoothing parameter ρ , those for parameters in diffusion and drift coefficient, i.e., α and β , of the latent diffusion process $\{X_t\}$. Thirdly, we have examined the performance of those estimators and the test statistics in computational simulation, and verified these statistics work well in realistic settings. In the fourth place, we have shown a real example of observation where $\rho \neq 0$ holds with statistical significance.

If we combine the test for noise detection by Nakakita and Uchida [28] and that for smoothed observation proposed in this paper, we can test if the observed process is diffusion or not in terms that the observation is noisy or smoothed. Note that the realised volatilities of the noisy observation

of diffusion processes increase as observation frequency increases while those of the smoothed observation decreases as the frequency grows. On that point, the noisy observation in ordinary meaning and the smoothed one are ‘opposite’ to each other.

These contributions, especially the third one, will cultivate the motivation to study statistical approaches for convolutionally observed diffusion processes furthermore, such as estimation of kernel function V appearing in the convoluted diffusion $\bar{X}_t := (V * X)(t)$, test theory for parameters α and β as likelihood-ratio-type test statistics, for example, see [29,30], large deviation inequalities for quasi-likelihood functions and associated discussion of Bayes-type estimators, e.g., [6,31–33]. By these future works, it is expected that the applicability of stochastic differential equations in real data analysis and contributions to the areas with high frequency observation of phenomena such as EEG will be enhanced.

8. Proofs

8.1. the Results for Some Laws of Large Numbers

In this subsection, we give the notations and statements of propositions without proofs except for Proposition 3: the detailed proofs are given in supplementary material. We assume $\Delta \leq \lambda$, $k \in \mathbf{N}, M > 0$, and consider a class of $\mathbf{R}^k \otimes \mathbf{R}^d$ -valued kernel functions on \mathbf{R} denoted as $\mathcal{K}(\Delta, k, M)$ such that for all $\Phi_\Delta \in \mathcal{K}(\Delta, k, M)$, it holds:

- (i) $\text{supp}\Phi_\Delta \subset [0, \Delta]$,
- (ii) for all $f : [0, \Delta] \times \Omega \rightarrow \mathbf{R}^k, \omega \in \Omega, \left| \int_0^\Delta \Phi_\Delta(\Delta - s) f(s, \omega) ds \right| \leq M \sup_{s \in [0, \Delta]} |f(s, \omega)|$
- (iii) for all $t_0 \geq -\lambda, f : \mathbf{R}^d \rightarrow \mathbf{R}$ which is continuous and at most polynomial growth,
$$\mathbf{E} \left[\int_{t-\Delta}^t \Phi_\Delta(t-s) f(X_s) ds \middle| \mathcal{F}_{t_0} \right] = \int_{t-\Delta}^t \Phi_\Delta(t-s) \mathbf{E}[f(X_s) | \mathcal{F}_{t_0}] ds.$$

Remark 5. Note one sufficient condition for $\Phi_\Delta \in \mathcal{K}(\Delta, k, M)$ is (i) $\Phi_\Delta : \mathbf{R} \rightarrow \mathbf{R}^k \otimes \mathbf{R}^d$, (ii) $\text{supp}\Phi_\Delta \subset [0, \Delta]$, (iii) $\int_0^\Delta \|\Phi_\Delta(\Delta - s)\| ds \leq M$ and (iv) $\mathcal{B}([0, \Delta])$ -measurable since

$$\left| \int_0^\Delta \Phi_\Delta(\Delta - s) f(s, \omega) ds \right| \leq \int_0^\Delta \|\Phi_\Delta(\Delta - s)\| |f(s, \omega)| ds \leq M \sup_{s \in [0, \Delta]} |f(s, \omega)|$$

for the Cauchy–Schwarz inequality, and Fubini’s theorem.

It is easily checked that $V_{\rho, h_n} \in \mathcal{K}(\max_{i=1, \dots, d} \rho_i h_n, d, d)$.

Let p denote an integer such that $\sup_{n \in \mathbf{N}} ph_n \leq \lambda, \Delta_n := ph_n$. We set the sequence of the kernels $\{\Phi_{\Delta_n, n}\}_{n \in \mathbf{N}}$ such that $\Phi_{\Delta_n, n} \in \mathcal{K}(\Delta_n, d, M)$ for some $M > 0, \int_0^{\Delta_n} \Phi_{\Delta_n, n} ds = I_d$ and there exist a matrix $B \in \mathbf{R}^d \otimes \mathbf{R}^d$ such that

$$\left\| \int_0^{\Delta_n + h_n} (\Phi_{\Delta_n, n}((\Delta_n + h_n) - s) - \Phi_{\Delta_n, n}(\Delta_n - s)) s ds - h_n B \right\| \leq Ch_n^2 (1 + |x|)^C,$$

a set $\mathbb{L} \subset \{0, \dots, p\}$ such that there exist functions $D_\ell : \mathbf{R}^d \rightarrow \mathbf{R}^d \otimes \mathbf{R}^d$ for $\ell \in \mathbb{L}$ such that

$$\begin{aligned} & \left\| \mathbf{E} \left[\left(\int_0^{\Delta_n+(1+\ell)h_n} \Phi_{\Delta_n,n}(\Delta_n - s_1) \left(\int_0^{s_1} a(x) \, d\tau_{s_2} \right) \, ds_1 \right) \right. \right. \\ & \quad \left. \left(\int_0^{\Delta_n+(1+\ell)h_n} (\Phi_{\Delta_n,n}((\Delta_n + (1 + \ell) h_n) - s_1) - \Phi_{\Delta_n,n}(\Delta_n + \ell h_n - s_1)) \right. \right. \\ & \quad \quad \left. \left. \times \left(\int_0^{s_1} a(x) \, d\tau_{s_2} \right) \, ds_1 \right)^T \right] - h_n D_\ell(x) \right\| \\ & \leq Ch_n^2 (1 + |x|)^C, \end{aligned}$$

a function $G : \mathbf{R}^d \rightarrow \mathbf{R}^d \otimes \mathbf{R}^d$ such that

$$\begin{aligned} & \left\| \mathbf{E} \left[\left(\int_0^{\Delta_n+h_n} (\Phi_{\Delta_n,n}((\Delta_n + h_n) - s_1) - \Phi_{\Delta_n,n}(\Delta_n - s_1)) \left(\int_0^{s_1} a(x) \, d\tau_{s_2} \right) \, ds_1 \right) \right. \right. \\ & \quad \left. \left(\int_0^{\Delta_n+h_n} (\Phi_{\Delta_n,n}((\Delta_n + h_n) - s_1) - \Phi_{\Delta_n,n}(\Delta_n - s_1)) \left(\int_0^{s_1} a(x) \, d\tau_{s_2} \right) \, ds_1 \right)^T \right] \right. \\ & \quad \left. - h_n G(x) \right\| \\ & \leq Ch_n^2 (1 + |x|)^C. \end{aligned}$$

We define

$$\bar{X}_{t,n} = \int_{t-\Delta_n}^t \Phi_{\Delta_n,n}(t-s) X_s \, ds,$$

and the following random quantities such that

$$\begin{aligned} \bar{v}_n(f(\cdot, \xi)) &:= \frac{1}{n} \sum_{i=1}^n f(\bar{X}_{ih_n,n}, \xi), \\ \bar{I}_{\ell,n}(v(\cdot, \xi)) &:= \frac{1}{nh_n} \sum_{i=1+\ell}^n v(\bar{X}_{(i-1-\ell)h_n,n}, \xi) \left[\bar{X}_{ih_n,n} - \bar{X}_{(i-1)h_n,n} - (h_n B) b(\bar{X}_{(i-1-\ell)h_n,n}) \right], \\ \bar{Q}_n(M(\cdot, \xi)) &:= \frac{1}{nh_n} \sum_{i=1}^n M(\bar{X}_{(i-1)h_n,n}, \xi) \left[(\bar{X}_{ih_n,n} - \bar{X}_{(i-1)h_n,n})^{\otimes 2} \right], \end{aligned}$$

where $f : \mathbf{R}^d \times \Xi \rightarrow \mathbf{R}$, $v : \mathbf{R}^d \times \Xi \rightarrow \mathbf{R}^d$, $M : \mathbf{R}^d \times \Xi \rightarrow \mathbf{R}^d \otimes \mathbf{R}^d$ are in \mathcal{C}^2 -class, and their first and second derivatives and themselves are at most polynomial growth uniformly in $\xi \in \Xi$.

Proposition 1. Under [A1], $\bar{v}_n(f(\cdot, \xi)) \rightarrow^P v_0(f(\cdot, \xi))$ uniformly in $\xi \in \Xi$.

Proposition 2. If $\ell \in \mathbb{L}$ and [A1] hold, $\bar{I}_{\ell,n}(v(\cdot, \xi)) \rightarrow^P v_0(\partial_x v [D_\ell^T](\cdot, \xi))$ uniformly in $\xi \in \Xi$.

Proposition 3. Under [A1], $\bar{Q}_n(M(\cdot, \xi)) \rightarrow^P v_0(M[G](\cdot, \xi))$ uniformly in $\xi \in \Xi$.

We set $p = [\bar{\rho}] + 1$, $\Delta_n = ph_n$ and see the evaluation of B , D_ℓ and G when setting our kernel $\{\Phi_{\Delta_n,n}\} = \{V_{\rho,h_n}\}$ as follows (for the derivation of the evaluation, see the supplementary material): we have $\Delta_n = ph_n$, $B = I_d$, $D_0(x) = \mathbb{D}_0(x|\rho)|_{\rho=\rho^*}$, where $\mathbb{D}_0^{(i,j)}(x|\rho) = A^{(i,j)}(x) f_{\mathbb{D}_0}(\rho^{(i)}, \rho^{(j)})$,

$$f_{\mathbb{D}_0}(\rho^{(i)}, \rho^{(j)}) := \begin{cases} 0 & \text{if } \rho^{(j)} = 0, \\ \frac{\rho^{(j)}}{2} & \text{if } \rho^{(i)} = 0, \rho^{(j)} \in (0, 1], \\ \frac{2\rho^{(j)} - 1}{2\rho^{(j)}} & \text{if } \rho^{(i)} = 0, \rho^{(j)} \in (1, \bar{\rho}], \\ \frac{6\rho^{(i)}\rho^{(j)} - 3(\rho^{(i)})^2 - 3\rho^{(i)}}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(i)} > 0, \rho^{(i)} + 1 < \rho^{(j)}, \\ \frac{(\rho^{(i)} - \rho^{(j)})^3 + 3(\rho^{(j)})^2 - 3\rho^{(j)} + 1}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(j)} > 1, \rho^{(i)} < \rho^{(j)} \leq \rho^{(i)} + 1, \\ \frac{3(\rho^{(j)})^2 - 3\rho^{(j)} + 1}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(j)} > 1, \rho^{(i)} \geq \rho^{(j)}, \\ \frac{(\rho^{(i)} - \rho^{(j)})^3 + (\rho^{(j)})^3}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(j)} \in (0, 1], \rho^{(i)} \in (0, \rho^{(j)}), \\ \frac{(\rho^{(j)})^3}{6\rho^{(i)}\rho^{(j)}} & \text{if } \rho^{(j)} \in (0, 1], \rho^{(i)} \geq \rho^{(j)}, \end{cases}$$

$D_\ell = O$ for $\ell \geq \lceil \max_{i=1, \dots, d} \rho_*^{(d)} \rceil + 1$ because of independent increments of the Wiener process, and $G(x) = \mathbb{G}(x|\rho)|_{\rho=\rho_*}$ where $\mathbb{G}(x|\rho) = \mathbb{G}(x, \alpha|\rho)|_{\alpha=\alpha_*}$.

8.2. Proof of the Results in Section 3.1

Proof of Lemma 1. By following the proof of the Proposition 3, it is sufficient to evaluate

$$\int_0^{(p+2)h_n} \int_0^{(p+2)h_n} A^{(i,i)}(x) \min\{s, s'\} \left(V_{\rho, h_n}^{(i,i)}((p+2)h_n - s) - V_{\rho, h_n}^{(i,i)}(ph_n - s) \right) \left(V_{\rho, h_n}^{(i,i)}((p+2)h_n - s') - V_{\rho, h_n}^{(i,i)}(ph_n - s') \right) ds' ds$$

for the asymptotic behaviour of the reduced quadratic variation by choosing

$$M^{(\ell_1, \ell_2)} = \begin{cases} 1 & \text{if } \ell_1 = \ell_2 = i, \\ 0 & \text{otherwise.} \end{cases}$$

If $\rho^{(i)} = 0$,

$$\begin{aligned} & \int_0^{(p+2)h_n} \int_0^{(p+2)h_n} A^{(i,i)}(x) \min\{s, s'\} \left(V_{\rho, h_n}^{(i,i)}((p+2)h_n - s) - V_{\rho, h_n}^{(i,i)}(ph_n - s) \right) \left(V_{\rho, h_n}^{(i,i)}((p+2)h_n - s') - V_{\rho, h_n}^{(i,i)}(ph_n - s') \right) ds' ds \\ &= A^{(i,i)}(x) \int_0^{(p+2)h_n} \int_0^{(p+2)h_n} \min\{s, s'\} (\delta((p+2)h_n - s) - \delta(ph_n - s)) (\delta((p+2)h_n - s') - \delta(ph_n - s')) ds' ds \\ &= A^{(i,i)}(x) ((p+2)h_n - 2ph_n + ph_n) \\ &= 2h_n A^{(i,i)}(x), \end{aligned}$$

and if $\rho^{(i)} \in (0, \bar{\rho}]$,

$$\begin{aligned}
 & \int_0^{(p+2)h_n} \int_0^{(p+2)h_n} A^{(i,i)}(x) \min\{s, s'\} \left(V_{\rho, h_n}^{(i,i)}((p+2)h_n - s) - V_{\rho, h_n}^{(i,i)}(ph_n - s) \right) \\
 & \quad \left(V_{\rho, h_n}^{(i,i)}((p+2)h_n - s') - V_{\rho, h_n}^{(i,i)}(ph_n - s') \right) ds' ds \\
 &= \frac{A^{(i,i)}(x)}{(\rho^{(i)}h_n)^2} \int_0^{(p+2)h_n} \int_0^{(p+2)h_n} \min\{s, s'\} \\
 & \quad \times \left(\mathbf{1}_{[0, \rho^{(i)}h_n]}((p+2)h_n - s) - \mathbf{1}_{[0, \rho^{(i)}h_n]}(ph_n - s) \right) \\
 & \quad \times \left(\mathbf{1}_{[0, \rho^{(i)}h_n]}((p+2)h_n - s') - \mathbf{1}_{[0, \rho^{(i)}h_n]}(ph_n - s') \right) ds' ds \\
 &= \frac{A^{(i,i)}(x)}{(\rho^{(i)}h_n)^2} \int_{(p+2-\rho^{(i)})h_n}^{(p+2)h_n} \int_{(p+2-\rho^{(i)})h_n}^{(p+2)h_n} \min\{s, s'\} ds' ds \\
 & \quad - \frac{2A^{(i,i)}(x)}{(\rho^{(i)}h_n)^2} \int_{(p+2-\rho^{(i)})h_n}^{(p+2)h_n} \int_{(p-\rho^{(i)})h_n}^{ph_n} \min\{s, s'\} ds' ds \\
 & \quad + \frac{A^{(i,i)}(x)}{(\rho^{(i)}h_n)^2} \int_{(p-\rho^{(i)})h_n}^{ph_n} \int_{(p-\rho^{(i)})h_n}^{ph_n} \min\{s, s'\} ds' ds \\
 &= \frac{A^{(i,i)}(x)}{(\rho^{(i)}h_n)^2} \int_{(p+2-\rho^{(i)})h_n}^{(p+2)h_n} \left(\int_s^{(p+2)h_n} s ds' + \int_{(p+2-\rho^{(i)})h_n}^s s' ds' \right) ds \\
 & \quad - \frac{2A^{(i,i)}(x)}{(\rho^{(i)}h_n)^2} \mathbf{1}_{(2, \bar{\rho}]}(\rho^{(i)}) \int_{ph_n}^{(p+2)h_n} \int_{(p-\rho^{(i)})h_n}^{ph_n} s' ds' ds \\
 & \quad - \frac{2A^{(i,i)}(x)}{(\rho^{(i)}h_n)^2} \mathbf{1}_{(2, \bar{\rho}]}(\rho^{(i)}) \int_{(p+2-\rho^{(i)})h_n}^{ph_n} \left(\int_s^{ph_n} s ds' + \int_{(p-\rho^{(i)})h_n}^s s' ds' \right) ds \\
 & \quad - \frac{2A^{(i,i)}(x)}{(\rho^{(i)}h_n)^2} \mathbf{1}_{(0, 2]}(\rho^{(i)}) \int_{(p+2-\rho^{(i)})h_n}^{(p+2)h_n} \int_{(p-\rho^{(i)})h_n}^{ph_n} s' ds' ds \\
 & \quad + \frac{A^{(i,i)}(x)}{(\rho^{(i)}h_n)^2} \int_{(p-\rho^{(i)})h_n}^{ph_n} \left(\int_s^{ph_n} s ds' + \int_{(p-\rho^{(i)})h_n}^s s' ds' \right) ds \\
 &= \begin{cases} 2h_n A^{(i,i)}(x) \left(1 - \frac{\rho^{(i)}}{6} \right) & \text{if } \rho^{(i)} \in (0, 2], \\ 2h_n A^{(i,i)}(x) \left(\frac{2}{\rho^{(i)}} - \frac{4}{3(\rho^{(i)})^2} \right) & \text{if } \rho^{(i)} \in (2, \bar{\rho}]. \end{cases}
 \end{aligned}$$

Hence, we obtain the proof (for details, see the supplementary material). \square

Proof of Lemma 2. Continuity is obvious, and monotonicity is obtained as follows: if $\rho^{(i)} \in (0, 1]$,

$$\begin{aligned}
 \frac{d}{d\rho^{(i)}} \left(6 - 2\rho^{(i)} \right) \left(6 - \rho^{(i)} \right)^{-1} &= \left((-2) \left(6 - \rho^{(i)} \right) - \left(6 - 2\rho^{(i)} \right) (-1) \right) \left(6 - \rho^{(i)} \right)^{-2} \\
 &= (-12) \left(6 - \rho^{(i)} \right)^{-2} < 0,
 \end{aligned}$$

and if $\rho^{(i)} \in (1, 2]$,

$$\begin{aligned} & \frac{d}{d\rho^{(i)}} (6\rho^{(i)} - 2) \left(6(\rho^{(i)})^2 - (\rho^{(i)})^3\right)^{-1} \\ &= \left(6 \left(6(\rho^{(i)})^2 - (\rho^{(i)})^3\right) - (6\rho^{(i)} - 2) \left(12\rho^{(i)} - 3(\rho^{(i)})^2\right)\right) \left(6(\rho^{(i)})^2 - (\rho^{(i)})^3\right)^{-2} \\ &= 6\rho^{(i)} \left(-7\rho^{(i)} + 4 + 2(\rho^{(i)})^2\right) \left(6(\rho^{(i)})^2 - (\rho^{(i)})^3\right)^{-2} < 0, \end{aligned}$$

and if $\rho^{(i)} \in (2, \bar{\rho}]$,

$$\frac{d}{d\rho^{(i)}} (3\rho^{(i)} - 1) (6\rho^{(i)} - 4)^{-1} = (-18) (6\rho^{(i)} - 4)^{-2} < 0.$$

The inverse can be obtained directly. \square

Proof of Theorem 1. It follows from Lemma 1, 2 and continuous mapping theorem. \square

8.3. Proofs of the Results in Section 3.2

Proof of Theorem 2. We can clearly prove the result by using Lemma 7 in Kessler [4], Proposition 7 in Nakakita and Uchida [28], and Slutsky’s theorem. \square

Proof of Theorem 3. By Lemma 1, there exists a number $\ell < 0$ such that

$$\frac{1}{nh_n} \sum_{k=1}^n \left(\bar{X}_{kh_n,n}^{(i)} - \bar{X}_{(k-1)h_n,n}^{(i)}\right)^2 - \frac{1}{nh_n} \sum_{2 \leq k \leq n} \left(\bar{X}_{2kh_n,n}^{(i)} - \bar{X}_{(2k-2)h_n,n}^{(i)}\right)^2 \xrightarrow{P} \ell < 0,$$

and hence it is sufficient to show that

$$\sup_{n \in \mathbb{N}} \mathbf{E} \left[\frac{2}{3nh_n^2} \sum_{k=1}^n \left(\bar{X}_{kh_n,n}^{(i)} - \bar{X}_{(k-1)h_n,n}^{(i)}\right)^4 \right] < \infty;$$

and it is obvious that

$$\left(\bar{X}_{kh_n,n}^{(i)} - \bar{X}_{(k-1)h_n,n}^{(i)}\right)^4 \leq C \left(\bar{X}_{kh_n,n}^{(i)} - X_{(k-\bar{\rho}-1)h_n}\right)^4 + C \left(\bar{X}_{(k-1)h_n,n}^{(i)} - X_{(k-\bar{\rho}-1)h_n}\right)^4$$

and

$$\begin{aligned}
 & \mathbf{E} \left[\left(\overline{X}_{kh_n, n}^{(i)} - X_{(k-\bar{\rho}-1)h_n} \right)^4 \middle| \mathcal{F}_{(k-\bar{\rho}-1)h_n} \right] \\
 &= \mathbf{E} \left[\left(\frac{1}{\rho_\star^{(i)} h_n} \int_{(k-\rho_\star^{(i)})h_n}^{kh_n} \left(X_s^{(i)} - X_{(k-\bar{\rho}-1)h_n} \right) ds \right)^4 \middle| \mathcal{F}_{(k-\bar{\rho}-1)h_n} \right] \\
 &\leq \mathbf{E} \left[\left(\frac{1}{\rho_\star^{(i)} h_n} \int_{(k-\rho_\star^{(i)})h_n}^{kh_n} \left| X_s^{(i)} - X_{(k-\bar{\rho}-1)h_n} \right| ds \right)^4 \middle| \mathcal{F}_{(k-\bar{\rho}-1)h_n} \right] \\
 &\leq \mathbf{E} \left[\left(\frac{1}{\rho_\star^{(i)} h_n} \int_{(k-\rho_\star^{(i)})h_n}^{kh_n} \sup_{s' \in [(k-\rho_\star^{(i)})h_n, kh_n]} \left| X_{s'}^{(i)} - X_{(k-\bar{\rho}-1)h_n} \right| ds \right)^4 \middle| \mathcal{F}_{(k-\bar{\rho}-1)h_n} \right] \\
 &= \mathbf{E} \left[\sup_{s \in [(k-\rho_\star^{(i)})h_n, kh_n]} \left| X_s^{(i)} - X_{(k-\bar{\rho}-1)h_n} \right|^4 \middle| \mathcal{F}_{(k-\bar{\rho}-1)h_n} \right] \\
 &\leq Ch_n^2 \left(1 + \left| X_{(k-\bar{\rho}-1)h_n} \right| \right)^C
 \end{aligned}$$

by Proposition A in Gloter [9], and a parallel result holds for $\left(\overline{X}_{(k-1)h_n, n}^{(i)} - X_{(k-\bar{\rho}-1)h_n} \right)^4$. Hence we obtain the result. \square

8.4. Proof of the Results in Section 4

Proof of Theorem 4. We only deal with the case where ρ_\star is unknown because the discussion for the case where ρ_\star is known is parallel. First of all, we prove the consistency of $\hat{\alpha}_n$. We obtain that

$$\begin{aligned}
 & \left| \frac{1}{n} \mathbb{H}_{1,n}(\alpha|\hat{\rho}_n) - \frac{1}{n} \mathbb{H}_{1,n}(\alpha|\rho_\star) \right| \\
 &= \left| -\frac{1}{n} \sum_{k=1}^n \left\| \frac{1}{h_n} \left(\bar{X}_{kh_n,n} - \bar{X}_{(k-1)h_n,n} \right)^{\otimes 2} - \mathbb{G} \left(\bar{X}_{(k-1)h_n,n}, \alpha|\hat{\rho}_n \right) \right\|^2 \right. \\
 &\quad \left. + \frac{1}{n} \sum_{k=1}^n \left\| \frac{1}{h_n} \left(\bar{X}_{kh_n,n} - \bar{X}_{(k-1)h_n,n} \right)^{\otimes 2} - \mathbb{G} \left(\bar{X}_{(k-1)h_n,n}, \alpha|\rho_\star \right) \right\|^2 \right| \\
 &\leq \left| \frac{2}{nh_n} \sum_{k=1}^n \left(\mathbb{G} \left(\bar{X}_{(k-1)h_n,n}, \alpha|\hat{\rho}_n \right) - \mathbb{G} \left(\bar{X}_{(k-1)h_n,n}, \alpha|\rho_\star \right) \right) \left[\left(\bar{X}_{kh_n,n} - \bar{X}_{(k-1)h_n,n} \right)^{\otimes 2} \right] \right| \\
 &\quad + \left| \frac{1}{n} \sum_{k=1}^n \left(\left\| \mathbb{G} \left(\bar{X}_{(k-1)h_n,n}, \alpha|\hat{\rho}_n \right) \right\|^2 - \left\| \mathbb{G} \left(\bar{X}_{(k-1)h_n,n}, \alpha|\rho_\star \right) \right\|^2 \right) \right| \\
 &\leq \frac{2}{nh_n} \sum_{k=1}^n \left| \bar{X}_{kh_n,n} - \bar{X}_{(k-1)h_n,n} \right|^2 \\
 &\quad \times \sum_{i=1}^d \sum_{j=1}^d \left| A^{(i,j)} \left(\bar{X}_{(k-1)h_n,n}, \alpha \right) \right| \left| f_{\mathbb{G}} \left(\hat{\rho}_n^{(i)}, \hat{\rho}_n^{(j)} \right) - f_{\mathbb{G}} \left(\rho_\star^{(i)}, \rho_\star^{(j)} \right) \right| \\
 &\quad + \frac{1}{n} \sum_{k=1}^n \sum_{i=1}^d \sum_{j=1}^d \left| A^{(i,j)} \left(\bar{X}_{(k-1)h_n,n}, \alpha \right) \right|^2 \left| f_{\mathbb{G}}^2 \left(\hat{\rho}_n^{(i)}, \hat{\rho}_n^{(j)} \right) - f_{\mathbb{G}}^2 \left(\rho_\star^{(i)}, \rho_\star^{(j)} \right) \right| \\
 &\leq \frac{C}{nh_n} \sum_{k=1}^n \left(1 + \left| \bar{X}_{(k-1)h_n,n} \right| \right)^C \left| \bar{X}_{kh_n,n} - \bar{X}_{(k-1)h_n,n} \right|^2 \\
 &\quad \times \sum_{i=1}^d \sum_{j=1}^d \left| f_{\mathbb{G}} \left(\hat{\rho}_n^{(i)}, \hat{\rho}_n^{(j)} \right) - f_{\mathbb{G}} \left(\rho_\star^{(i)}, \rho_\star^{(j)} \right) \right| \\
 &\quad + \frac{C}{n} \sum_{k=1}^n \left(1 + \left| \bar{X}_{(k-1)h_n,n} \right| \right)^C \sum_{i=1}^d \sum_{j=1}^d \left| f_{\mathbb{G}}^2 \left(\hat{\rho}_n^{(i)}, \hat{\rho}_n^{(j)} \right) - f_{\mathbb{G}}^2 \left(\rho_\star^{(i)}, \rho_\star^{(j)} \right) \right| \\
 &\rightarrow^P 0 \text{ uniformly in } \alpha,
 \end{aligned}$$

because continuous mapping theorem holds. Therefore, it follows from Proposition 1 and Proposition 3 that

$$\begin{aligned}
 & \frac{1}{n} \mathbb{H}_{1,n}(\alpha|\hat{\rho}_n) - \frac{1}{n} \mathbb{H}_{1,n}(\alpha_\star|\rho_\star) \\
 &= \frac{2}{nh_n} \sum_{k=1}^n \mathbb{G} \left(\bar{X}_{(k-1)h_n,n}, \alpha|\rho_\star \right) \left[\left(\bar{X}_{kh_n,n} - \bar{X}_{(k-1)h_n,n} \right)^{\otimes 2} \right] \\
 &\quad - \frac{1}{n} \sum_{k=1}^n \left\| \mathbb{G} \left(\bar{X}_{(k-1)h_n,n}, \alpha|\rho_\star \right) \right\|^2 \\
 &\quad - \frac{2}{nh_n} \sum_{k=1}^n \mathbb{G} \left(\bar{X}_{(k-1)h_n,n}, \alpha_\star|\rho_\star \right) \left[\left(\bar{X}_{kh_n,n} - \bar{X}_{(k-1)h_n,n} \right)^{\otimes 2} \right] \\
 &\quad + \frac{1}{n} \sum_{k=1}^n \left\| \mathbb{G} \left(\bar{X}_{(k-1)h_n,n}, \alpha_\star|\rho_\star \right) \right\|^2 \\
 &\quad + o_P^*(1) \\
 &\rightarrow^P \mathbb{V}_1(\alpha|\zeta_\star) \text{ uniformly in } \alpha
 \end{aligned}$$

where $o_P^*(1)$ indicates the term converging in probability to zero uniformly in θ . Then we obtain that $\hat{\alpha}_n \rightarrow \alpha_\star$ in the same way as Kessler [4] with Assumption [A3].

In the next place, we consider the consistency of $\hat{\beta}_n$. Firstly, we consider the case $\max_i \rho_*^{(i)} \in (\ell - 1, \ell)$ for an integer $\ell \in \{1, \dots, [\bar{\rho}] + 1\}$. Then it is sufficient to show

$$\frac{1}{nh_n} \mathbb{H}_{2,n}(\beta|\hat{\rho}_n) - \frac{1}{nh_n} \mathbb{H}_{2,n}(\beta_*|\rho_*) \rightarrow^P \mathbb{V}_2(\beta|\zeta_*) \text{ uniformly in } \beta$$

due to Assumption [A3]. Because the evaluation $D_j(x) = O$ where $j \geq \lceil \max_{i=1, \dots, n} \rho_*^{(i)} \rceil + 1$ using independent increments of the Wiener process, Proposition 1 and Proposition 2 verify

$$F_\ell(\beta) - \frac{1}{nh_n} \mathbb{H}_{2,n}(\beta_*|\rho_*) \rightarrow^P \mathbb{V}_2(\beta|\zeta_*) \text{ uniformly in } \beta,$$

where

$$F_j(\beta) := -\frac{1}{nh_n^2} \sum_{k=1+j}^n \left| \bar{X}_{kh_n, n} - \bar{X}_{(k-1)h_n, n} - h_n b\left(\bar{X}_{(k-1-j)h_n, n}, \beta\right) \right|^2.$$

In addition, the exact convergences such that

$$P\left(\mathbf{1}_{\{\ell\}}\left(\left[\max_i \hat{\rho}_n^{(i)}\right] + 1\right) = 1\right) \rightarrow 1, P\left(\mathbf{1}_{\{j\}}\left(\left[\max_i \hat{\rho}_n^{(i)}\right] + 1\right) = 1\right) \rightarrow 0$$

hold for all $j \neq \ell$, since for all $j = 1, \dots, [\bar{\rho}] + 1$,

$$P\left(\mathbf{1}_{\{j\}}\left(\left[\max_i \hat{\rho}_n^{(i)}\right] + 1\right) = 1\right) = P\left(\max_i \hat{\rho}_n^{(i)} \in [j - 1, j)\right).$$

Therefore, for any $\epsilon > 0$,

$$\begin{aligned} &P\left(\sup_{\beta \in \Theta_2} \left| \frac{1}{nh_n} \mathbb{H}_{2,n}(\beta|\hat{\rho}_n) - \frac{1}{nh_n} \mathbb{H}_{2,n}(\beta_*|\rho_*) - \mathbb{V}_2(\beta|\zeta_*) \right| > \epsilon\right) \\ &\leq \sum_{j \neq \ell} P\left(\mathbf{1}_{\{j\}}\left(\left[\max_i \hat{\rho}_n^{(i)}\right] + 1\right) = 1\right) \\ &\quad + P\left(\left\{\mathbf{1}_{\{\ell\}}\left(\left[\max_i \hat{\rho}_n^{(i)}\right] + 1\right) = 1\right\} \cap \left\{\sup_{\beta \in \Theta_2} \left| F_\ell(\beta) - \frac{1}{nh_n} \mathbb{H}_{2,n}(\beta_*|\rho_*) - \mathbb{V}_2(\beta|\zeta_*) \right| > \epsilon\right\}\right) \\ &\rightarrow 0. \end{aligned}$$

For the case $\max_i \rho_*^{(i)} = \ell$ for an integer $\ell = \{0, \dots, [\bar{\rho}] + 1\}$, we similarly obtain

$$\frac{1}{nh_n} \mathbb{H}_{2,n}(\beta|\hat{\rho}_n) - \frac{1}{nh_n} \mathbb{H}_{2,n}(\beta_*|\rho_*) \rightarrow^P \mathbb{V}_2(\beta|\zeta_*) \text{ uniformly in } \beta$$

because we have

$$\begin{aligned} F_\ell(\beta) - \frac{1}{nh_n} \mathbb{H}_{2,n}(\beta_*|\rho_*) &\rightarrow^P \mathbb{V}_2(\beta|\zeta_*) \text{ uniformly in } \beta, \\ F_{\ell+1}(\beta) - \frac{1}{nh_n} \mathbb{H}_{2,n}(\beta_*|\rho_*) &\rightarrow^P \mathbb{V}_2(\beta|\zeta_*) \text{ uniformly in } \beta, \end{aligned}$$

and

$$P \left(\mathbf{1}_{\{\ell\}} \left(\left[\max_i \hat{\rho}_n^{(i)} \right] + 1 \right) + \mathbf{1}_{\{\ell+1\}} \left(\left[\max_i \hat{\rho}_n^{(i)} \right] + 1 \right) = 1 \right) \rightarrow 1,$$

$$P \left(\mathbf{1}_{\{j\}} \left(\left[\max_i \hat{\rho}_n^{(i)} \right] + 1 \right) = 0 \right) \rightarrow 1, \text{ for all } j \neq \ell, \ell + 1,$$

and it holds that for any $\epsilon > 0$,

$$P \left(\sup_{\beta \in \Theta_2} \left| \frac{1}{nh_n} \mathbb{H}_{2,n}(\beta | \hat{\rho}_n) - \frac{1}{nh_n} \mathbb{H}_{2,n}(\beta_* | \rho_*) - \mathbb{V}_2(\beta | \zeta_*) \right| > \epsilon \right)$$

$$\leq \sum_{j \neq \ell, \ell+1} P \left(\mathbf{1}_{\{j\}} \left(\left[\max_i \hat{\rho}_n^{(i)} \right] + 1 \right) = 1 \right)$$

$$+ P \left(\left\{ \mathbf{1}_{\{\ell\}} \left(\left[\max_i \hat{\rho}_n^{(i)} \right] + 1 \right) = 1 \right\} \right.$$

$$\quad \left. \cap \left\{ \sup_{\beta \in \Theta_2} \left| F_\ell(\beta) - \frac{1}{nh_n} \mathbb{H}_{2,n}(\beta_* | \rho_*) - \mathbb{V}_2(\beta | \zeta_*) \right| > \epsilon \right\} \right)$$

$$+ P \left(\left\{ \mathbf{1}_{\{\ell+1\}} \left(\left[\max_i \hat{\rho}_n^{(i)} \right] + 1 \right) = 1 \right\} \right.$$

$$\quad \left. \cap \left\{ \sup_{\beta \in \Theta_2} \left| F_{\ell+1}(\beta) - \frac{1}{nh_n} \mathbb{H}_{2,n}(\beta_* | \rho_*) - \mathbb{V}_2(\beta | \zeta_*) \right| > \epsilon \right\} \right)$$

$$\rightarrow 0.$$

Hence it is shown that $\hat{\beta}_n \xrightarrow{P} \beta_*$ with Assumption [A3]. \square

Supplementary Materials: The following are available online at <http://www.mdpi.com/1099-4300/22/9/1031/s1>.

Author Contributions: Conceptualization, S.H.N. and M.U.; Methodology, S.H.N. and M.U.; Software, S.H.N.; Validation, S.N.; Formal Analysis, S.H.N.; Investigation, S.H.N.; Resources, S.H.N. and M.U.; Data Curation, S.H.N.; Writing—Original Draft Preparation, S.H.N.; Writing—Review & Editing, M.U.; Visualization, S.H.N.; Supervision, M.U.; Project Administration, M.U.; Funding Acquisition, S.H.N. and M.U. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partially funded by JST CREST, JSPS KAKENHI Grant Number JP17H01100 and JP20J10058.

Acknowledgments: This work was partially supported by Cooperative Research Program of the Institute of Statistical Mathematics.

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

References

1. Florens-Zmirou, D. Approximate discrete time schemes for statistics of diffusion processes. *Statistics* **1989**, *20*, 547–557. [\[CrossRef\]](#)
2. Yoshida, N. Estimation for diffusion processes from discrete observation. *J. Multivar. Anal.* **1992**, *41*, 220–242. [\[CrossRef\]](#)
3. Bibby, B.M.; Sørensen, M. Martingale estimating functions for discretely observed diffusion processes. *Bernoulli* **1995**, *1*, 17–39. [\[CrossRef\]](#)
4. Kessler, M. Estimation of an ergodic diffusion from discrete observations. *Scand. J. Stat.* **1997**, *24*, 211–229. [\[CrossRef\]](#)
5. Kessler, M.; Sørensen, M. Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli* **1999**, *5*, 299–314. [\[CrossRef\]](#)

6. Yoshida, N. Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann. Inst. Stat. Math.* **2011**, *63*, 431–479. [[CrossRef](#)]
7. Uchida, M.; Yoshida, N. Adaptive estimation of an ergodic diffusion process based on sampled data. *Stoch. Process. Their Appl.* **2012**, *122*, 2885–2924. [[CrossRef](#)]
8. Uchida, M.; Yoshida, N. Adaptive Bayes type estimators of ergodic diffusion processes from discrete observations. *Stat. Inference Stoch. Process.* **2014**, *17*, 181–219. [[CrossRef](#)]
9. Gloter, A. Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Stat.* **2000**, *4*, 205–227. [[CrossRef](#)]
10. Ditlevsen, S.; Sørensen, M. Inference for observations of integrated diffusion processes. *Scand. J. Stat.* **2004**, *31*, 417–429. [[CrossRef](#)]
11. Gloter, A. Parameter estimation for a discretely observed integrated diffusion process. *Scand. J. Stat.* **2006**, *33*, 83–104. [[CrossRef](#)]
12. Gloter, A.; Gobet, E. LAMN property for hidden processes: The case of integrated diffusions. *Ann. Institut Henri Poincaré Probab. Stat.* **2008**, *44*, 104–128. [[CrossRef](#)]
13. Sørensen, M. Prediction-based estimating functions: Review and new developments. *Braz. J. Probab. Stat.* **2011**, *25*, 362–391. [[CrossRef](#)]
14. Donnet, S.; Samson, A. A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models. *Adv. Drug Deliv. Rev.* **2013**, *65*, 929–939. [[CrossRef](#)] [[PubMed](#)]
15. Delattre, M.; Genon-Catalot, V.; Samson, A. Mixtures of stochastic differential equations with random effects: Application to data clustering. *J. Stat. Plan. Inference* **2016**, *173*, 109–124. [[CrossRef](#)]
16. Picchini, U.; Forman, J.L. Bayesian inference for stochastic differential equation mixed effects models of a tumour xenography study. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2019**, *68*, 887–913. [[CrossRef](#)]
17. Ruse, M.G.; Samson, A.; Ditlevsen, S. Inference for biomedical data by using diffusion models with covariates and mixed effects. *J. R. Stat. Soc. Ser. C Appl. Stat.* **2020**, *69*, 167–193. [[CrossRef](#)]
18. Zhang, L.; Mykland, P.A.; Ait-Sahalia, Y. A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. *J. Am. Stat. Assoc.* **2005**, *100*, 1394–1411. [[CrossRef](#)]
19. Ait-Sahalia, Y.; Jacod, J. *High-Frequency Financial Econometrics*; Princeton University Press: Princeton, NJ, USA, 2014.
20. BNCI Horizon 2020. Two Class Motor Imagery. 2014. Available online: <http://bnci-horizon-2020.eu/database/data-sets> (accessed on 20 April 2019).
21. Jacod, J.; Li, Y.; Mykland, P.A.; Podolskij, M.; Vetter, M. Microstructure noise in the continuous case: The pre-averaging approach. *Stoch. Process. Their Appl.* **2009**, *119*, 2249–2276. [[CrossRef](#)]
22. Jacod, J.; Podolskij, M.; Vetter, M. Limit theorems for moving averages of discretized processes plus noise. *Ann. Stat.* **2010**, *38*, 1478–1545. [[CrossRef](#)]
23. Bibinger, M.; Hautsch, N.; Malec, P.; Reiß, M. Estimating the quadratic covariation matrix from noisy observations: Local method of moments and efficiency. *Ann. Stat.* **2014**, *42*, 1312–1346. [[CrossRef](#)]
24. Koike, Y. Quadratic covariation estimation of an irregularly observed semimartingale with jumps and noise. *Bernoulli* **2016**, *22*, 1894–1936. [[CrossRef](#)]
25. Ogihara, T. Parametric inference for nonsynchronously observed diffusion processes in the presence of market microstructure noise. *Bernoulli* **2018**, *24*, 3318–3383. [[CrossRef](#)]
26. Iacus, S.M. *Simulation and Inference for Stochastic Differential Equations: With R Examples*; Springer: New York, NY, USA, 2008.
27. Steyerl, D.; Scherer, R.; Faller, J.; Müller-Putz, G.R. Random forests in non-invasive sensorimotor rhythm brain-computer interfaces: a practical and convenient non-linear classifier. *Biomed. Tech.* **2016**, *61*, 77–86. [[CrossRef](#)]
28. Nakakita, S.H.; Uchida, M. Inference for ergodic diffusions plus noise. *Scand. J. Stat.* **2019**, *46*, 470–516. [[CrossRef](#)]
29. Kitagawa, H.; Uchida, M. Adaptive test statistics for ergodic diffusion processes sampled at discrete times. *J. Stat. Plan. Inference* **2014**, *150*, 84–110. [[CrossRef](#)]
30. Nakakita, S.H.; Uchida, M. Adaptive test for ergodic diffusions plus noise. *J. Stat. Plan. Inference* **2019**, *203*, 131–150. [[CrossRef](#)]
31. Ogihara, T.; Yoshida, N. Quasi-likelihood analysis for the stochastic differential equation with jumps. *Stat. Inference Stoch. Process.* **2011**, *14*, 189–229. [[CrossRef](#)]

32. Clinet, S.; Yoshida, N. Statistical inference for ergodic point processes and application to Limit Order Book. *Stoch. Process. Their Appl.* **2017**, *127*, 1800–1839. [[CrossRef](#)]
33. Nakakita, S.H.; Uchida, M. Quasi-likelihood analysis of an ergodic diffusion plus noise. *arXiv* **2018**, arXiv:1806.09401.



© 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/>).