



HAL
open science

Parameter estimation by contrast minimization for noisy observations of a diffusion process

Benjamin Favetto

► **To cite this version:**

Benjamin Favetto. Parameter estimation by contrast minimization for noisy observations of a diffusion process. 2010. hal-00493967v1

HAL Id: hal-00493967

<https://hal.science/hal-00493967v1>

Preprint submitted on 21 Jun 2010 (v1), last revised 31 Dec 2012 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Parameter estimation by contrast minimization for noisy observations of a diffusion process

BENJAMIN FAVETTO ¹

¹*Laboratoire MAP5, Université Paris Descartes, U.F.R. de Mathématiques et Informatique,
CNRS UMR 8145*

45, rue des Saints-Pères, 75270 Paris Cedex 06, France.

E-mail: benjamin.favetto@mi.parisdescartes.fr

url: http://www.math-info.univ-paris5.fr/~favetto/

Abstract We consider the estimation of unknown parameters in the drift and diffusion coefficients of a one-dimensional ergodic diffusion X when the observation Y is a discrete sampling of X with an additive noise, at times $i\delta, i = 1 \dots N$. Assuming that the sampling interval tends to 0 while the total length time interval tends to infinity, we prove limit theorems for functionals associated with the observations, based on local means of the sample. We apply these results to obtain a contrast function. The associated minimum contrast estimators are shown to be consistent. We provide an illustration on simulated and real data from neuronal activity.

Keywords: contrast function, diffusion process, hidden Markov models, parametric inference, discrete time noisy observations.

1. Introduction

Statistical inference for continuous time models based on high frequency data has been the subject of a huge number of recent papers. On one hand, continuous time stochastic processes are increasingly used for modelling purposes. On the other hand, such kind of data is now commonly available in various fields of applications whether in finance or in biology and medicine.

Among continuous time models, one-dimensional diffusion processes have received a lot of attention. In particular, diffusion models have been introduced in the studies of neuronal activity (see *e.g.* Ditlevsen and LANSKY (2005), Höpfner and Brodda (2006) and the references given in these papers). More precisely, let (X_t) be given by the stochastic differential equation:

$$dX_t = b(X_t, \kappa)dt + \sigma(X_t, \lambda)dB_t, \quad X_0 = \eta \quad (1.1)$$

with B a standard Wiener process and η a random variable independent of B , and $b(\cdot, \kappa), \sigma(\cdot, \lambda)$ real valued functions, defined on \mathbb{R} , depending on unknown parameters $(\kappa, \lambda) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Suppose that, for some positive sampling interval δ , a sample

$(X_{i\delta}, i \leq N)$ is observed and that is required to estimate $\theta = (\kappa, \lambda)$. As the exact likelihood of such observation is generally intractable, other methods have been developed to obtain explicit estimators : minimum contrast estimators, estimating equations, simulation based methods, ... See *e.g.* Florens-Zmirou (1989), Yoshida (1992), Kessler (1997), Bibby and Sørensen (1995), Sørensen (2009), Genon-Catalot (1990), Genon-Catalot and Jacod (1993), Pedersen (1995b), Pedersen (1995a), Aït-Sahalia (2002).

More recently, especially in the case of high frequency data, other kinds of observations have been investigated among which the case of noisy observations. Suppose that, instead of observing exactly $X_{i\delta}$, the observation at time $i\delta$ is given by

$$Y_{i\delta} = X_{i\delta} + \rho\varepsilon_{i\delta} \quad (1.2)$$

with $(\varepsilon_{i\delta}, i \geq 0)$ a sequence of i.i.d. random variables, satisfying $\mathbb{E}(\varepsilon_{i\delta}) = 0$, $\mathbb{E}((\varepsilon_{i\delta})^2) = 1$, independent of the process (X_t) . This kind of model takes into account measurement errors or, in the case of financial data, the so-called microstructure noise (see *e.g.* Zhang et al. (2005), Jacod et al. (2009)). It provides a model fitted to data which show non Markovian features.

The exact likelihood of $(Y_{i\delta}, i \leq N)$ given by (1.1)-(1.2) is generally intractable except for few models (essentially for Gaussian diffusions with additive Gaussian noise, see *e.g.* Cappé et al. (2005), Pedersen (1994), Favetto and Samson (2009)). For data within a fixed length-time interval ($\delta = \delta_N = \frac{1}{N}$, $N\delta_N = 1$), estimation for a general diffusion with additive Gaussian noise is investigated in Gloter and Jacod (2001). The authors use a contrast method. Only diffusion coefficient parameters can be consistently estimated in this case.

In this paper, we study observations given by (1.1)-(1.2) where $\delta = \delta_N \rightarrow 0$ while $N\delta_N \rightarrow \infty$, under ergodic properties for the hidden diffusion X and propose consistent estimators of both the drift and diffusion coefficient parameters (κ, λ) . The noise distribution is unknown, the variance ρ^2 of the noise term may be known or unknown and we assume either that ρ is fixed or that $\rho = \rho_N \rightarrow 0$.

Our starting idea is to reduce the influence of the noise by splitting the sample into sub-samples and taking empirical means of the sub-samples. More precisely, we split the sample into p blocks of size k , with $N = pk$, where $p = p_N$ and $k = k_N$ tend to infinity with N . Then, setting $\Delta_N = p_N\delta_N$ where p_N and δ_N are chosen such that $\Delta_N \rightarrow 0$, we build the empirical mean of the j^{th} block:

$$Y_{\bullet}^j = X_{\bullet}^j + \rho_N\varepsilon_{\bullet}^j, \quad j = 0, 1 \dots k_N - 1, \quad (1.3)$$

where, for $Z = Y, X, \varepsilon$,

$$Z_{\bullet}^j = \frac{1}{p_N} \sum_{i=0}^{p_N-1} Z_{j\Delta_N+i\delta_N}. \quad (1.4)$$

Thus, Δ_N defines a coarser sampling interval than δ_N , still tending to 0 while $N\delta_N = k_N\Delta_N \rightarrow \infty$. The empirical mean X_{\bullet}^j is close to the mean $\bar{X}_j = \frac{1}{\Delta_N} \int_{j\Delta_N}^{(j+1)\Delta_N} X_s ds$ of X over $[j\Delta_N, (j+1)\Delta_N]$, which, in turn, is close to $X_{j\Delta_N}$. The variance of ε_{\bullet}^j is reduced by a factor $\frac{1}{p_N}$.

Our statistical procedure is based on the k_N -sample $(Y_{\bullet}^j, j = 0 \dots k_N - 1)$ and follows a scheme analogous to the one in Gloter (2006). We study the differences $Y_{\bullet}^j - X_{j\Delta_N}$ (Proposition 3.2) and prove a regression type relation for the Y_{\bullet}^j 's (Proposition 3.3) which is the basement of the statistical applications. These results allow us to prove limit theorems for the variation and the quadratic variation of (Y_{\bullet}^j) (Theorems 4.1 and 4.2). We precise the adequate calibration of δ_N and p_N for the limit theorems to hold. Then we introduce two different contrasts: according to the rates of p_N, δ_N , the noise variance has or has not to be taken into account. We set $\delta_N = p_N^{-\alpha}$, with $1 < \alpha \leq 2$. For $\alpha = 2$ and $\rho_N = \rho$, the value ρ^2 appears in the contrast definition. In each case, we prove the consistency of the associated minimum contrast estimators. As could be expected, we have to deal with two-rate contrasts, which indicate that drift parameters estimators must have rate $\sqrt{k_N \Delta_N}$, while diffusion coefficient parameters estimators must have rate $\sqrt{k_N}$. The study of the asymptotic distributions of the minimum contrast estimators is postponed to a further paper. Estimators are implemented on simulated data and on real data of neuronal activities provided by Idoux et al. (2006).

The paper is organised as follows. In Section 2, we give our notations and assumptions on the model. Section 3 is devoted to the small sample properties of the empirical means sample (Y_{\bullet}^j) and Section 4 to uniform convergence in probability results. In Section 5, we introduce the contrasts and prove the consistency of the estimators. We also deal with the case $\rho_N = \rho$ unknown and prove that ρ^2 can be replaced by an estimator in the contrast formula.

Section 6 is devoted to examples and numerical results. For several models of hidden diffusions, we implement our estimators on simulated data for different choices of (N, δ_N, p_N) and of the noise level. We illustrate the estimation method on the set of neuronal data for one model of diffusion with estimated noise level. Section 7 contains some concluding remarks. Proofs are gathered in Section 8, and some auxiliary results are recalled in the Appendix.

2. Assumptions and Notations

Consider the one-dimensional stochastic differential equation

$$dX_t = b(X_t, \kappa_0)dt + \sigma(X_t, \lambda_0)dB_t, \quad X_0 = \eta \quad (2.1)$$

where B is a standard Brownian motion and η is a real valued random variable independent of B . The functions $b(x, \kappa)$ and $\sigma(x, \lambda)$ are respectively defined on $\mathbb{R} \times \Theta_1$ and $\mathbb{R} \times \Theta_2$ where Θ_1 (resp. Θ_2) is a compact convex subset of \mathbb{R}^{d_1} (resp. \mathbb{R}^{d_2}). For simplicity of notations, in proofs, we assume that $d_1 = d_2 = 1$. We denote by $\theta_0 = (\kappa_0, \lambda_0)$ the true value of the parameter and assume that $\theta_0 \in \overset{\circ}{\Theta}$ where $\Theta = \Theta_1 \times \Theta_2$.

From now on, we set $b(x) = b(x, \kappa_0)$ and $\sigma(x) = \sigma(x, \lambda_0)$ and make classical assumptions on functions b and σ ensuring that (2.1) admits an unique strong solution $(X_t)_{t \geq 0}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and that this solution is positive recurrent on \mathbb{R} .

(A1) Functions b and σ belong to $\mathcal{C}^2(\mathbb{R})$, $\sigma(x) > 0$ for all x , and there exists $c > 0$ such that for all $x \in \mathbb{R}$:

$$\begin{aligned} |b(x)| + |b'(x)| + |b''(x)| &\leq c(1 + |x|), \\ \sigma(x) + |\sigma'(x)| + |\sigma''(x)| &\leq c(1 + |x|). \end{aligned}$$

(A2) For $x_0 \in \mathbb{R}$, let $s(x) = \exp(-2 \int_{x_0}^x \frac{b(u)}{\sigma^2(u)} du)$ denote the scale density and $m(x) = \frac{1}{\sigma^2(x)s(x)}$ the speed density. Assume $\int_{-\infty}^{\infty} s(x)dx = \int_{-\infty}^{\infty} m(x)dx = \infty$ and $\int_{-\infty}^{\infty} m(x)dx = M < \infty$.

(A3) Let $\nu_0(dx) = \frac{1}{M}m(x)dx$. For all $k > 0$, ν_0 admits a finite moment of order k .

(A4) For all $k > 0$, $\sup_{t \geq 0} \mathbb{E}(|X_t|^k) < \infty$.

(A5) The common distribution of the random variables $\varepsilon_{i\delta_N}$ admits a 8th order moment, and is symmetric. We set $m_1 = \mathbb{E}(|\varepsilon_{i\delta_N}|)$, $m_4 = \mathbb{E}((\varepsilon_{i\delta_N})^4)$, $m_8 = \mathbb{E}((\varepsilon_{i\delta_N})^8)$.

(B1) $\rho_N = \rho > 0$.

(B2) $\rho_N \rightarrow 0$ when $N \rightarrow \infty$.

Assumption **(A1)** implies that (1.1) admits a unique strong solution on \mathbb{R} . Under **(A1)** and **(A2)**, ν_0 is the unique invariant probability of (2.1) and (X_t) satisfies the classical ergodic theorem (see *e.g.* Rogers and Williams (2000))

$$\forall f \in L^1(d\nu_0), \quad \frac{1}{T} \int_0^T f(X_s) ds \xrightarrow{T \rightarrow \infty} \nu_0(f) \quad a.s.$$

Moreover, under Assumption **(A1)**, for all $k \geq 1$, there exists a constant $c(k)$ such that, for all $t \geq 0$:

$$\mathbb{E} \left(\sup_{s \in [t, t+1]} |X_s|^k \middle| \mathcal{G}_t \right) \leq c(k)(1 + |X_t|^k). \quad (2.2)$$

where $\mathcal{G}_t = \sigma(B_s, s \leq t; \eta)$. (See *e.g.* Gloter (2000)).

Furthermore, Assumptions **(A1)**-**(A3)** imply **(A4)** if η has distribution ν_0 or η is deterministic (for the latter case, see Gloter (2006), Proposition 3).

Below, we first assume that the noise level ρ_N is known and discuss later the case where ρ_N is unknown. We distinguish the two cases **(B1)**-**(B2)** which yield different results. Assumption **(B2)** corresponds, for example, to the case $\rho_N \varepsilon_{i\delta_N} = V_{(i+1)\delta_N} - V_{i\delta_N}$, with (V_t) a Brownian motion independent of η and (B_t) . Here, $\rho_N = \sqrt{\delta_N}$.

We divide observations into k_N blocks of size p_N and set $\Delta_N = p_N \delta_N$. Since $X_{j\Delta_N}$ is unobserved, we build the local means (1.3). Notice that

$$\mathbb{E}((\varepsilon_{\bullet}^j)^2) = \frac{1}{p_N}, \quad \mathbb{E}((\varepsilon_{\bullet}^j)^4) = \frac{3}{p_N^2} + \frac{m_4 - 3}{p_N^3}, \quad \mathbb{E}((\varepsilon_{\bullet}^j)^8) \leq c \left(\frac{1}{p_N^4} + \frac{m_8}{p_N^7} \right).$$

for c a universal constant (the last inequality is obtained using the Rosenthal inequality (see Hall and Heyde (1980) p.23)). Define the σ -fields

$$\begin{aligned} \mathcal{G}_j^N &= \mathcal{G}_{j\Delta_N} = \sigma(B_s, s \leq j\Delta_N; \eta), \quad \mathcal{H}_j^N = \mathcal{G}_j^N \vee \mathcal{A}_j^N, \\ \mathcal{A}_j^N &= \sigma(\varepsilon_{k\Delta_N + i\delta_N}, i \leq p_N - 1, k \leq j - 1) = \sigma(\varepsilon_{l\delta_N}, l \leq j\Delta_N - \delta_N) \end{aligned} \quad (2.3)$$

For $0 \leq j \leq k_N - 1$, the random variable Y_{\bullet}^j is \mathcal{H}_{j+1}^N measurable. We introduce, for further use, a condition on functions $g : \mathbb{R} \times \Theta \rightarrow \mathbb{R}$:

(C1) The function g is the restriction of a function defined on $\mathbb{R} \times \mathcal{O}$ with \mathcal{O} an open neighbourhood of Θ and

$$\exists c > 0, \forall x \in \mathbb{R} \quad \sup_{\theta \in \Theta} |g(x, \theta)| \leq c(1 + |x|).$$

We need the following statistical assumptions (**(A6)** is the usual identifiability condition for this problem):

(A6)

$$\begin{aligned} \sigma(x, \lambda) = \sigma(x, \lambda_0) & \quad \nu_0 \text{ almost everywhere implies } \lambda = \lambda_0, \\ b(x, \kappa) = b(x, \kappa_0) & \quad \nu_0 \text{ almost everywhere implies } \kappa = \kappa_0. \end{aligned}$$

(A7) Functions $b, \sigma, \sigma^{-1}, \partial_x b, \partial_\kappa b, \partial_x \sigma, \partial_\lambda \sigma, \partial_{xx}^2 b, \partial_{\kappa\kappa}^2 b, \partial_{x\kappa}^2 b, \partial_{xx} \sigma, \partial_{\lambda\lambda}^2 \sigma$ and $\partial_{x\lambda}^2 \sigma$ exist, are continuous and satisfy Condition **(C1)**, where ∂ denotes the partial derivative.

3. Small sample properties of the local means sample

The following random variables appear in the expansions below:

$$\zeta_{j+1,N} = \frac{1}{p_N} \sum_{i=0}^{p_N-1} \int_{j\Delta_N+i\delta_N}^{(j+1)\Delta_N} dB_s, \quad \zeta'_{j+2,N} = \frac{1}{p_N} \sum_{i=0}^{p_N-1} \int_{(j+1)\Delta_N}^{(j+1)\Delta_N+i\delta_N} dB_s, \quad (3.1)$$

$$\xi'_{j+1,N} = \frac{1}{\Delta_N^{3/2}} \int_{(j+1)\Delta_N}^{(j+2)\Delta_N} ((j+2)\Delta_N - s) dB_s, \quad (3.2)$$

$$\xi'_{i+1,j,N} = \frac{1}{\delta_N^{3/2}} \int_{j\Delta_N+(i+1)\delta_N}^{j\Delta_N+(i+2)\delta_N} (j\Delta_N + (i+2)\delta_N - s) dB_s. \quad (3.3)$$

Some basic properties of these random variables are summarized in Lemma 8.1 and A.4 in the Appendix.

Proposition 3.1. *Let $\bar{X}_j = \Delta_N^{-1} \int_{j\Delta_N}^{(j+1)\Delta_N} X_s ds$. Under Assumption **(A1)**, we have*

$$\bar{X}_j - X_{\bullet}^j = \sqrt{\delta_N} \left(\frac{1}{p_N} \sum_{i=0}^{p_N-1} \sigma(X_{j\Delta_N+i\delta_N}) \xi'_{i,j,N} \right) + e_{j,N}$$

with (see (2.3))

$$|\mathbb{E}(e_{j,N} | \mathcal{H}_j^N)| \leq \delta_N C(1 + |X_{j\Delta_N}|), \quad \mathbb{E}(e_{j,N}^2 | \mathcal{H}_j^N) \leq \delta_N^2 C(1 + |X_{j\Delta_N}|^4).$$

The following proposition precises the local asymptotic behaviour of the observation blocks, by a first order comparison between Y_{\bullet}^j and $X_{j\Delta_N}$. It can be compared to Proposition 2.2 in Gloter (2000).

Proposition 3.2. *Under (A1), we have for $j \leq k_N - 1$,*

$$Y_{\bullet}^j - X_{j\Delta_N} = \sigma(X_{j\Delta_N})\sqrt{\Delta_N}\xi'_{j,N} + e'_{j,N} + \rho_N \varepsilon_{\bullet}^j \quad (3.4)$$

with $|\mathbb{E}(e'_{j,N} | \mathcal{H}_j^N)| \leq c\Delta_N(1 + |X_{j\Delta_N}|)$ and

$$\mathbb{E}(e'_{j,N}{}^2 | \mathcal{H}_j^N) \leq c\Delta_N^2(1 + |X_{j\Delta_N}|^4), \quad \mathbb{E}(e'_{j,N}{}^4 | \mathcal{H}_j^N) \leq c\Delta_N^3(1 + |X_{j\Delta_N}|^4).$$

If moreover (A5) holds, for $k \leq 8$, there exists $c > 0$ such that, for $j \leq k_N - 1$:

$$\mathbb{E}(|Y_{\bullet}^j - X_{j\Delta_N}|^k | \mathcal{H}_j^N) \leq C \left(\Delta_N^{k/2}(1 + |X_{j\Delta_N}|^k) + \rho_N^k \mathbb{E}(|\varepsilon_{\bullet}^j|^k) \right). \quad (3.5)$$

We deduce

Corollary 3.1. *Assume (A1) and (A5), and consider $f : \mathbb{R}^2 \times \Theta \rightarrow \mathbb{R}$ such that $f, \partial_x f, \partial_{xx}^2 f$ satisfy (C1). Then, there exists $c > 0$ such that, for all $j \geq 0$ and for all $\theta \in \Theta$:*

$$\left| \mathbb{E} \left(f(Y_{\bullet}^j, \theta) - f(X_{j\Delta_N}, \theta) \mid \mathcal{H}_j^N \right) \right| \leq c(\Delta_N(1 + |X_{j\Delta_N}|^2) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^4)}) \quad (3.6)$$

and for $l = 1, 2$

$$\begin{aligned} \mathbb{E} \left((f(Y_{\bullet}^j, \theta) - f(X_{j\Delta_N}, \theta))^{2l} \mid \mathcal{H}_j^N \right) &\leq c(1 + |X_{j\Delta_N}|^{2l} + \rho_N^{2l} \mathbb{E}((\varepsilon_{\bullet}^j)^{2l})) \\ &\quad \times (\Delta_N^l(1 + |X_{j\Delta_N}|^{2l}) + \rho_N^{2l} \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^{4l})}). \end{aligned} \quad (3.7)$$

The following proposition is essential for the limit theorems of Section 4 and for the statistical application.

Proposition 3.3. *Under Assumptions (A1) and (A5), we have*

$$Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N b(Y_{\bullet}^j) = \sigma(X_{j\Delta_N})(\zeta_{j+1,N} + \zeta'_{j+2,N}) + \tau_{j,N} + \rho_N(\varepsilon_{\bullet}^{j+1} - \varepsilon_{\bullet}^j)$$

where $\tau_{j,N}$ is \mathcal{H}_{j+2}^N measurable, and there exists a constant c such that

$$\begin{aligned} |\mathbb{E}(\tau_{j,N} | \mathcal{H}_j^N)| &\leq c\Delta_N(\Delta_N(1 + |X_{j\Delta_N}|^3) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^4)}), \\ \mathbb{E}(\tau_{j,N}^2 | \mathcal{H}_j^N) + |\mathbb{E}(\tau_{j,N}\zeta_{j+1,N} | \mathcal{H}_j^N)| + |\mathbb{E}(\tau_{j,N}\zeta'_{j+2,N} | \mathcal{H}_j^N)| &\leq \\ &c\Delta_N(1 + |X_{j\Delta_N}|^2 + \rho_N^2 \mathbb{E}((\varepsilon_{\bullet}^j)^2))(\Delta_N(1 + |X_{j\Delta_N}|^4) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^4)}), \\ \mathbb{E}(\tau_{j,N}^4 | \mathcal{H}_j^N) &\leq c(1 + |X_{j\Delta_N}|^4 + \rho_N^4 \mathbb{E}((\varepsilon_{\bullet}^j)^4))(\Delta_N^4(1 + |X_{j\Delta_N}|^4) + \rho_N^4 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^8)}). \end{aligned}$$

Note that, for $i = 1, 2$, $\rho_N^{2i} \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^{4i})} = O(\frac{\rho_N^{2i}}{\rho_N^i})$.

Remark: In Gloter (2000), Theorem 2.3., it is proved that (see Proposition 3.1)

$$\bar{X}_{j+1} - \bar{X}_j - \Delta_N b(\bar{X}_j) = \sqrt{\Delta_N} \sigma(X_{j\Delta_N}) (\xi_{j,N} + \xi'_{j+1,N}) + \bar{\tau}_{j,N}$$

where $\bar{\tau}_{j,N}$ satisfies $|\mathbb{E}(\bar{\tau}_{j,N} | \mathcal{G}_j^N)| \leq c\Delta_N^2(1 + |X_{j\Delta_N}|^3)$. In Proposition 3.3, additional terms due to the noise appear.

4. Uniform convergence in probability results

In this section, $f : \mathbb{R} \times \Theta \rightarrow \mathbb{R}$ denotes a \mathcal{C}^2 function, such that f , $\partial_x f$, $\partial_{xx}^2 f$ and $\partial_\theta f$ satisfy **(C1)**. The estimation results of Section 5 rely on the following statements.

Proposition 4.1. *Under Assumptions **(A1)**-**(A5)** and **(B1)** or **(B2)**, we have*

$$\bar{\nu}_N(f(\cdot, \theta)) = \frac{1}{k_N} \sum_{j=0}^{k_N-1} f(Y_\bullet^j, \theta) \longrightarrow \nu_0(f(\cdot, \theta)) \quad (4.1)$$

uniformly in θ , in probability, as $N \rightarrow \infty$, with $\delta_N \rightarrow 0$, $p_N \rightarrow \infty$, $k_N \rightarrow \infty$, $\Delta_N = p_N \delta_N \rightarrow 0$ and $N\delta_N = k_N \Delta_N \rightarrow \infty$.

The next theorem precises the variation of the process (Y_\bullet^j) .

Theorem 4.1. *Under Assumptions **(A1)**-**(A5)**, **(B1)** or **(B2)**, with $\delta_N = p_N^{-\alpha}$, $\alpha \in (1, 2]$, ($\Delta_N = p_N^{1-\alpha}$) we have*

$$\bar{I}_N(f(\cdot, \theta)) = \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1}, \theta) (Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1})) \xrightarrow{\mathbb{P}} 0 \quad (4.2)$$

uniformly in θ , as $N \rightarrow \infty$, with $\delta_N \rightarrow 0$, $p_N \rightarrow \infty$, $k_N \rightarrow \infty$, $\Delta_N \rightarrow 0$ and $N\delta_N \rightarrow \infty$.

The following result deals with the quadratic variation of Y_\bullet^j .

Theorem 4.2. *Assume **(A1)**-**(A5)**.*

1. *If $\delta_N = p_N^{-\alpha}$ with $\alpha \in (1, 2)$ ($\Delta_N = p_N^{1-\alpha}$) and **(B1)** ($\rho_N = \rho > 0$), then*

$$\bar{Q}_N(f(\cdot, \theta)) = \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1}, \theta) (Y_\bullet^{j+1} - Y_\bullet^j)^2 \xrightarrow{\mathbb{P}} \frac{2}{3} \nu_0(f(\cdot, \theta) \sigma^2), \quad (4.3)$$

2. *If $\delta_N = p_N^{-2}$ ($\Delta_N = \frac{1}{p_N}$) and **(B1)** ($\rho_N = \rho > 0$), then*

$$\bar{Q}_N(f(\cdot, \theta)) \xrightarrow{\mathbb{P}} \frac{2}{3} \nu_0(f(\cdot, \theta) \sigma^2) + 2\rho^2 \nu_0(f(\cdot, \theta)), \quad (4.4)$$

3. If $\delta_N = p_N^{-\alpha}$, $\alpha \in (1, 2]$ with **(B2)** ($\rho_N \rightarrow 0$), then

$$\bar{Q}_N(f(\cdot, \theta)) \xrightarrow{\mathbb{P}} \frac{2}{3} \nu_0(f(\cdot, \theta) \sigma^2), \quad (4.5)$$

where all the convergences in probability are uniform in $\theta \in \Theta$, as $N \rightarrow \infty$, with $\delta_N \rightarrow 0$, $p_N \rightarrow \infty$, $k_N \rightarrow \infty$, $\Delta_N \rightarrow 0$ and $N\delta_N \rightarrow \infty$.

The proofs of these two last theorems are based on the results of Proposition 3.3 and Lemma A.3 in the Appendix. Theorems 4.1 and 4.2 can be compared to the following classical results:

$$\frac{1}{k_N \Delta_N} \sum_{j=0}^{k_N-1} f(X_{j\Delta_N}, \theta) (X_{(j+1)\Delta_N} - X_{j\Delta_N} - \Delta_N b(X_{j\Delta_N})) = o_P(1), \quad (4.6)$$

$$\frac{1}{k_N \Delta_N} \sum_{j=0}^{k_N-1} f(X_{j\Delta_N}, \theta) (X_{(j+1)\Delta_N} - X_{j\Delta_N})^2 = \nu_0(f(\cdot, \theta) \sigma^2) + o_P(1). \quad (4.7)$$

Theorem 4.1 gives the analogous result as (4.6), under the condition $\delta_N = p_N^{-\alpha}$, $\alpha \in (1, 2]$ and provided that we introduce a lag to avoid correlation terms of order Δ_N (if no lag, the limit is not 0, see for instance Gloter (2006)). Theorem 4.2 underestimates $\nu_0(f(\cdot, \theta) \sigma^2)$ because the variance of $\zeta_{j+1, N} + \zeta'_{j+2, N}$ (see Proposition 3.3) is equivalent to $\frac{2}{3} \Delta_N$ and not to Δ_N . Moreover, for $\rho_N = \rho$ and $\delta_N = p_N^{-2}$, an additional bias appears due to the noise.

5. Statistical estimation by contrast minimization

5.1. Definition of the contrasts

Let $c(\cdot, \lambda) = \sigma^2(\cdot, \lambda)$ and define

$$\mathcal{E}_N(\theta) = \sum_{j=1}^{k_N-2} \left\{ \frac{3}{2\Delta_N} \frac{(Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N b(Y_{\bullet}^{j-1}, \kappa))^2}{c(Y_{\bullet}^{j-1}, \lambda)} + \log(c(Y_{\bullet}^{j-1}, \lambda)) \right\}. \quad (5.1)$$

When $\rho_N = \rho$ is fixed (**(B1)**) and $\delta_N = p_N^{-\alpha}$ with $\alpha \in (1, 2]$, let $c_{N, \rho}(x, \lambda) = c(x, \lambda) + 3\Delta_N^{\frac{2-\alpha}{\alpha-1}} \rho^2$ and define

$$\mathcal{E}_N^{\rho}(\theta) = \sum_{j=1}^{k_N-2} \left\{ \frac{3}{2\Delta_N} \frac{(Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N b(Y_{\bullet}^{j-1}, \kappa))^2}{c_{N, \rho}(Y_{\bullet}^{j-1}, \lambda)} + \log(c_{N, \rho}(Y_{\bullet}^{j-1}, \lambda)) \right\} \quad (5.2)$$

We have $\lim_{N \rightarrow \infty} c_{N, \rho}(x, \lambda) = c_{\rho}(x, \lambda)$ with $c_{\rho}(x, \lambda) = c(x, \lambda)$ if $1 < \alpha < 2$ and $c_{\rho}(x, \lambda) = c(x, \lambda) + 3\rho^2$ if $\alpha = 2$.

Let $\hat{\theta}_N$ and $\hat{\theta}_N^\rho$ be the associated minimum contrast estimators, defined as any solution of

$$\hat{\theta}_N = \underset{\theta \in \Theta}{\operatorname{arginf}} \mathcal{E}_N(\theta) \text{ and } \hat{\theta}_N^\rho = \underset{\theta \in \Theta}{\operatorname{arginf}} \mathcal{E}_N^\rho(\theta). \quad (5.3)$$

Theorem 5.1. *Assume (A1)-(A7), and consider $\hat{\theta}_N$ and $\hat{\theta}_N^\rho$ defined by (5.3).*

1. *If (B1) or (B2) holds, with $\delta_N = p_N^{-\alpha}$, $\alpha \in (1, 2)$, the estimator $\hat{\theta}_N$ is consistent: $\hat{\theta}_N \rightarrow \theta_0$ in \mathbb{P}_{θ_0} probability.*
2. *If (B1) holds, with $\delta_N = p_N^{-\alpha}$, $\alpha \in (1, 2]$, the estimator $\hat{\theta}_N^\rho$ is consistent.*

Note that point 1 does not require the knowledge of ρ_N .

5.2. Estimation with unknown observation noise level under (B1)

Assuming (B1) with unknown ρ , we consider the estimator $\hat{\rho}_N^2 = \frac{1}{2N} \sum_{i=0}^{N-1} (Y_{(i+1)\delta_N} - Y_{i\delta_N})^2$, which is the half of the quadratic variation of the observations.

Lemma 5.1. *Assume (A1)-(A5) and (B1). Then we have $\hat{\rho}_N^2 \xrightarrow{\mathbb{P}} \rho^2$, when $N \rightarrow \infty$, with $\delta_N \rightarrow 0$ and $N\delta_N \rightarrow \infty$. If, moreover, $N\delta_N^2 \rightarrow 0$, $\sqrt{N}(\hat{\rho}_N^2 - \rho^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 3\rho^4)$.*

The minimum contrast estimator $\hat{\theta}_N^{\hat{\rho}_N^2}$ based on the contrast $\mathcal{E}_N^{\hat{\rho}_N^2}(\theta)$ satisfies:

Corollary 5.1. *Assume (A1)-(A7), (B1) and $\delta_N = p_N^{-\alpha}$ with $\alpha \in (1, 2]$. The estimator $\hat{\theta}_N^{\hat{\rho}_N^2}$ is consistent.*

5.3. Link with the case of noisy observations of integrated diffusions

Consider (V_t) a standard Brownian motion independent of (X_t) and suppose that observations are

$$Y_{i\delta_N} = X_{i\delta_N} + \rho(V_{(i+1)\delta_N} - V_{i\delta_N}).$$

Setting $\rho_N = \rho\sqrt{\delta_N}$, $\varepsilon_{i\delta_N} = (V_{j\Delta_N + (i+1)\delta_N} - V_{j\Delta_N + i\delta_N})/\sqrt{\delta_N}$, we are in case (B2) and

$$Y_{\bullet}^j = X_{\bullet}^j + \frac{\rho}{p_N}(V_{(j+1)\Delta_N} - V_{j\Delta_N}).$$

This kind of observations can be compared with noisy observations of integrated diffusions (see *e.g.* Baltazar-Larios and Sørensen (2009)). Indeed, consider

$$\begin{cases} dX_t &= b(X_t, \kappa_0)dt + \sigma(X_t, \lambda_0)dB_t \\ dZ_t &= X_t dt + \varepsilon dV_t \end{cases}$$

and suppose that we want to estimate θ_0 from discrete observations $(Z_{j\Delta_N})$. We have

$$\Delta_N^{-1}(Z_{(j+1)\Delta_N} - Z_{j\Delta_N}) = \bar{X}_j + \frac{\epsilon}{\Delta_N} \int_{j\Delta_N}^{(j+1)\Delta_N} dV_t.$$

This relation may be compared to $Y_{\bullet}^j = X_{\bullet}^j + \rho_N \varepsilon_{\bullet}^j$ where ε_{\bullet}^j is a $\mathcal{N}(0, p_N^{-1})$ random variable if we set $\rho_N = \frac{\epsilon}{\sqrt{\delta_N}}$. As X_{\bullet}^j and \bar{X}_j have similar properties, we can use the previous contrasts with (Y_{\bullet}^j) replaced by $\Delta_N^{-1}(Z_{(j+1)\Delta_N} - Z_{j\Delta_N})$ to estimate θ_0 provided that $\epsilon = \epsilon_N$ is such that $\frac{\epsilon_N}{\sqrt{\delta_N}} = O(1)$.

6. Examples

In this section, simulation results are given for several examples of diffusion models on simulated data. For the Ornstein-Uhlenbeck model, an implementation on real neuronal data is proposed.

6.1. The Ornstein-Uhlenbeck process (simulation)

The hidden diffusion solves

$$dX_t = \kappa X_t dt + \lambda dB_t \quad (6.1)$$

with $\kappa < 0$ and $\lambda > 0$, and X_0 is deterministic or follows the stationary distribution of X . We assume $\rho_N = \rho > 0$ and consider several distributions for the noise.

In this model, we can compute explicitly the estimator $\hat{\theta}_N$ by minimizing the contrast. With the expressions of $\partial_{\kappa} \mathcal{E}_N(\theta)$ and $\partial_{\lambda} \mathcal{E}_N(\theta)$, we find

$$\begin{aligned} \hat{\lambda}_N^2 &= \frac{3}{2k_N \Delta_N} \sum_{j=1}^{k_N-2} (Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N \hat{\kappa}_N Y_{\bullet}^{j-1})^2 - 3\rho^2 \mathbf{1}_{\{\alpha=2\}}; \\ \hat{\kappa}_N &= \frac{1}{\Delta_N} \frac{\sum_{j=1}^{k_N-2} Y_{\bullet}^{j-1} (Y_{\bullet}^{j+1} - Y_{\bullet}^j)}{\sum_{j=1}^{k_N-2} (Y_{\bullet}^{j-1})^2}. \end{aligned}$$

We can replace $\hat{\lambda}_N^2$ by

$$\tilde{\lambda}_N^2 = \frac{3}{2k_N \Delta_N} \sum_{j=1}^{k_N-2} (Y_{\bullet}^{j+1} - Y_{\bullet}^j)^2, \text{ as } \hat{\lambda}_N^2 - \tilde{\lambda}_N^2 = o_P(1).$$

In Tables 1-5, the common distribution of $\varepsilon_{i\delta}$ is $\mathcal{N}(0, 1)$ and Table 6 presents some results with different distributions. Tables 1, 2 and 3 give mean and variance of $\hat{\theta}_N$ for different values of δ, α and N . The values of the parameters are $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$. We have used 500 replications, and we give the empirical mean and standard deviation in parenthesis.

$$N = 5000, \delta = 0.01 \ (N\delta = 50, N\delta^2 = 0.5) \ \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$$

	$\alpha = 1.17(p = 50, k = 100)$	$\alpha = 1.5(p = 22, k = 227)$	$\alpha = 2(p = 10, k = 500)$
$\hat{\kappa}_N \ (10^2 \text{ Var})$	-0.58 (1.53)	-0.76 (2.75)	-0.82 (3.26)
$\hat{\lambda}_N^2 \ (10^2 \text{ Var})$	0.76 (1.19)	1.07 (1.25)	0.86 (2.61)

Table 1. Influence of the size of blocks on the estimators, Ornstein-Uhlenbeck model. ($N = 5000$ observations, $\delta = 0.01$, 500 replications) $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$

$$N = 20000, \delta = 0.005 \ (N\delta = 100, N\delta^2 = 0.5) \ \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$$

	$\alpha = 1.35(p = 50, k = 400)$	$\alpha = 1.5(p = 34, k = 588)$	$\alpha = 2(p = 14, k = 1428)$
$\hat{\kappa}_N \ (10^2 \text{ Var})$	-0.74 (1.08)	-0.81 (1.47)	-0.87 (1.51)
$\hat{\lambda}_N^2 \ (10^3 \text{ Var})$	0.95 (3.87)	1.05 (3.88)	0.92 (11.07)

Table 2. Influence of the size of blocks on the estimators, Ornstein-Uhlenbeck model. ($N = 20000$ observations, $\delta = 0.005$, 500 replications) $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$

$$N = 100000, \delta = 0.001 \ (N\delta = 100, N\delta^2 = 0.1) \ \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$$

	$\alpha = 1.3(p = 200, k = 500)$	$\alpha = 1.5(p = 100, k = 1000)$	$\alpha = 2(p = 32, k = 3125)$
$\hat{\kappa}_N \ (10^2 \text{ Var})$	-0.81 (1.36)	-0.89 (1.49)	-0.96 (1.95)
$\hat{\lambda}_N^2 \ (10^3 \text{ Var})$	0.90 (2.74)	1.02 (1.99)	0.92 (3.85)

Table 3. Influence of the size of blocks on the estimators, Ornstein-Uhlenbeck model. ($N = 100000$ observations, $\delta = 0.001$, 500 replications) $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$

First, we remark that the empirical variance is bigger in the case $\alpha = 2$ than in the other cases. The parameter κ_0 is always underestimated, but the estimation of κ_0 is clearly improved as N grows, and δ is close to 0. The estimation of λ_0 is better in Table 2 than in Table 1, and rather close in Tables 2 and 3. The variance decreases strongly in the case $\alpha = 2$.

In Table 4, we study the influence of the noise on the estimators, in the case $\alpha = \frac{3}{2}$. We use 500 replications, with $\delta = 0.001$ and $N = 10^5$, and we give the empirical mean and standard deviation in parenthesis.

$$N = 10^5, \delta = 10^{-3}, \alpha = 1.5, \kappa_0 = -1, \lambda_0 = 1$$

	$\rho^2 = 0.1$	$\rho^2 = 1$	$\rho^2 = 2$	$\rho^2 = 5$
$\hat{\kappa}_N \ (10^2 \text{ Var})$	-0.91 (1.49)	-0.89 (1.50)	-0.86 (1.75)	-0.83 (1.52)
$\hat{\lambda}_N^2 \ (10^3 \text{ Var})$	0.96 (1.71)	1.17 (2.92)	1.47 (4.33)	2.37 (13.42)

Table 4. Influence of the observation noise variance on the estimators, Ornstein-Uhlenbeck model. (500 replications, $N = 10^5$, $\delta = 0.001$, $\alpha = \frac{3}{2}$)

A strong bias appears for $\hat{\lambda}_N$ when ρ^2 is bigger than 1, whereas there are no significant changes in the estimation of the drift parameter κ_0 . The empirical variances for the estimation of λ_0 also increases: the presence of noise in the observations contaminates the estimation of the diffusion coefficient in this case.

In Table 5, we study in Table 5 the influence of the value of the diffusion coefficient on

the estimators, in the case $\alpha = \frac{3}{2}$. We use 500 replications, with $\delta = 0.001$ and $N = 10^5$, and we give the empirical mean and standard deviation in parenthesis.

$N = 10^5, \delta = 10^{-3}, \alpha = 1.5, \kappa_0 = -1, \rho^2 = 1$				
	$\lambda_0^2 = 0.1$	$\lambda_0^2 = 0.5$	$\lambda_0^2 = 1$	$\lambda_0^2 = 2$
$\hat{\kappa}_N$ (10^2 Var)	-0.81 (1.48)	-0.87 (1.54)	-0.90 (1.64)	-0.89 (1.62)
$\hat{\lambda}_N^2$ (10^3 Var)	0.23 (0.12)	0.58 (0.78)	1.01 (1.95)	2.01 (6.93)

Table 5. Influence of the diffusion coefficient on the estimators, Ornstein-Uhlenbeck model. (500 replications, $N = 10^5$, $\delta = 0.001$, $\alpha = \frac{3}{2}$)

The smallest value of λ_0^2 is overestimated by $\hat{\lambda}_N^2$, and this result confirms the ones of Table 4 about high levels of noise. For the other values of λ_0^2 , no bias is observed.

We finally study in Table 6 the influence of the distribution of the errors on the estimators. We choose in this case $\alpha = \frac{3}{2}$, $\kappa_0 = -1$, $\lambda_0 = 1$, $\rho^2 = 0.5$. We use 500 replications, with $\delta = 0.001$ and $N = 10^5$, and we give the empirical mean and standard deviation in parenthesis. We make the appropriate corrections on the distributions of $\varepsilon_{i\delta}$ to have unitary variance.

$N = 10^5, \delta = 10^{-3}, \alpha = 1.5, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$				
	$\mathcal{N}(0, 1)$	Laplace($0, \frac{1}{\sqrt{2}}$)	Uniform($-\sqrt{3}, \sqrt{3}$)	Logistic($0, \frac{\sqrt{3}}{\pi}$)
$\hat{\kappa}_N$ (10^2 Var)	-0.89 (1.65)	-0.90 (1.52)	-0.87 (1.53)	-0.89 (1.65)
$\hat{\lambda}_N^2$ (10^3 Var)	1.02 (2.11)	1.02 (2.18)	1.31 (3.45)	1.02 (2.10)

Table 6. Influence of the distribution of the noise on the estimation, Ornstein-Uhlenbeck model. ($N = 10^5$ observations, $\delta = 0.0001$, $\alpha = \frac{3}{2}$)

We observe that, except in the case of a Uniform distribution, the estimators give results close to the Gaussian case. For the case $\varepsilon_{i\delta} \sim \text{Uniform}(-\sqrt{3}, \sqrt{3})$, a significant positive bias is observed, and the variance is bigger in this case than in the case of Gaussian distribution.

These simulations stress two facts : first, the value $\alpha = \frac{3}{2}$ for the local mean size parameter appears as a good compromise, with a bias in the estimation of κ lower than the bias observed for values of α close to 1, and an empirican variance on simulations lower than the variance observed for $\alpha = 2$. The second remark is about the number of observations: for $N = 5000$ observations, κ is underestimated, for all the values of α considered. Then, the context of high frequency data requires a large number of observations, with a very small discretization step, to be taken into consideration.

6.2. Comparison with a discretely observed Ornstein-Uhlenbeck process

In this section, we are interested in the comparison, on simulated datasets, of our method with the methods based on the direct observation of the diffusion at discrete time (see e.g. Genon-Catalot (1990) and Kessler (1997)). To compare the quality of the noise reduction

and its influence to the estimation of the parameters, we compare the results for discrete observations with noise, based on the minimization of the contrast built on the (Y_{\bullet}^j) with those obtained for the discrete observations without noise, based on the minimization of the contrast built on the $(X_{i\delta_N})$. In both cases the same datasets of N observations with a δ_N -step of discretization are considered. The hidden diffusion (X_t) is an Ornstein-Uhlenbeck process (6.1). We compare the estimators based on the discrete observations $(X_{i\delta_N})$ with the estimator based on (Y_{t_i}) with $Y_{t_i} = X_{t_i} + \rho\varepsilon_{t_i}$, $t_i = i\delta_N$. The results based on the direct observations are given in Table 7, and we refer to Tables 1, 2 and 3 for the results based on noisy observations.

$\alpha = 1.5, \kappa_0 = -1, \lambda_0 = 1, \text{ no noise}$			
	$N = 5.10^3, \delta = 10^{-2}$	$N = 2.10^4, \delta = 5.10^{-3}$	$N = 10^5, \delta = 10^{-3}$
$\hat{\kappa}_N$ (Var)	-1.04 (0.21)	-1.02 (0.13)	-1.01 (0.14)
$\hat{\lambda}_N^2$ (Var)	0.99 (1.98×10^{-2})	0.99 (9.80×10^{-3})	1.00 (4.30×10^{-3})

Table 7. Parameter estimation with direct observations of the Ornstein-Uhlenbeck model, for several numbers of observations. (500 replications, $\alpha = 1.5, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$)

The estimation of κ_0 is better for a direct observation of the diffusion, but in this case, the whole set of N observations is taken into account, whereas the size of the set of local means is $k_N = N\delta_N^{\frac{1}{\alpha}}$.

6.3. The Ornstein-Uhlenbeck process (neuronal data)

Diffusion-based model has been introduced in the 90's in the field of neuronal studies. The Ornstein-Uhlenbeck diffusion is classical (see *e.g.* Ditlevsen and Lansky (2005)), and the estimation of the parameters has been studied in Ditlevsen and Ditlevsen (2007), for example, when the diffusion is assumed to be observed directly.

Let us consider the stochastic differential equation

$$dZ_t = \left(-\frac{Z_t}{\tau} + \kappa\right)dt + \lambda dB_t, \quad X_0 = x. \quad (6.2)$$

Assume that the observations are at discrete time $t_0 < \dots < t_N$ and that they are given by

$$Y_{t_i} = Z_{t_i} + \rho\varepsilon_{t_i}$$

where (ε_{t_i}) is a sequence of independent $\mathcal{N}(0, 1)$ random variables and ρ is supposed to be known (or preliminary estimated).

Minimum contrast estimators of τ, κ, λ are given by:

$$\begin{pmatrix} -\frac{1}{k_N} \sum_{j=1}^{k_N-2} Y_{\bullet}^{j-1} & 1 \\ -\frac{1}{k_N} \sum_{j=1}^{k_N-2} (Y_{\bullet}^{j-1})^2 & \frac{1}{k_N} \sum_{j=1}^{k_N-2} Y_{\bullet}^{j-1} \end{pmatrix} \begin{pmatrix} \hat{\tau}_N^{-1} \\ \hat{\kappa}_N \end{pmatrix} = \begin{pmatrix} \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} (Y_{\bullet}^{j+1} - Y_{\bullet}^j) \\ -\frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} Y_{\bullet}^{j-1} (Y_{\bullet}^{j+1} - Y_{\bullet}^j) \end{pmatrix}$$

and

$$\hat{\lambda}_{2N}^2 = \frac{3}{2k_N \Delta_N} \sum_{j=1}^{k_N-2} (Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N \left(-\frac{Y_{\bullet}^{j-1}}{\hat{\tau}_N} + \hat{\kappa}_N\right))^2.$$

The dataset used for this study has been formerly presented in Idoux et al. (2006). An example is given in Figure 3.

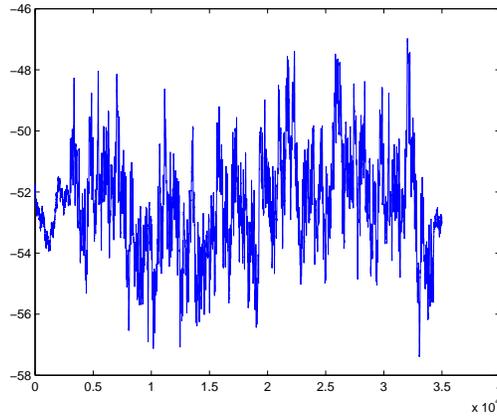


Figure 1. The neuronal dataset ($N = 35000$ observations, $\delta = 0.02 \times 10^{-2}$ seconds).

In this dataset, we have $\delta = 0.02$ (10^{-2} seconds), $N = 35000$ observations, $t_N = N\delta = 700$ (10^{-2} seconds). We estimate ρ^2 with:

$$\hat{\rho}_N^2 = \frac{1}{2N} \sum_{i=0}^{N-1} (Y_{t_{i+1}} - Y_{t_i})^2.$$

We have for this dataset $\hat{\rho}_N^2 = 0.0014$.

Then we compute the estimators $(\hat{\tau}_N, \hat{\kappa}_N, \hat{\lambda}_N^2)$ with different choices of p . Results are presented in Table 8.

	$p = 23(\alpha = 1.25)$	$p = 14(\alpha = 1.5)$	$p = 7(\alpha = 2)$
$\hat{\tau}_N$ (10^{-2} seconds)	13.52	5.85	4.67
$\hat{\kappa}_N$ (10^2 mV/ sec)	-3.87	-8.95	-11.23
$\hat{\lambda}_N^2$	1.94	1.78	1.10

Table 8. Parameter estimation for neuronal data (with measurement error).

Due to the low level of noise, we also provide the estimators corresponding to a direct observation of the discretized diffusion in Table 9.

	$\tilde{\tau}_N$ (10^{-2} seconds)	$\tilde{\kappa}_N$ (10^2 mV/ sec)	$\tilde{\lambda}_N^2$
$N = 35000, \delta = 0.0002$	40	-1.31	0.14

Table 9. Parameter estimation for neuronal data (without measurement error).

The results presented in Table 8 and 9 are rather different, but the mean of the stationary distribution $\mu = \tau\kappa$ is well estimated. Indeed, we find

- $\hat{\mu}_N = \hat{\tau}_N \hat{\kappa}_N = -52.28mV$ for the estimator based on the noisy observations model,
- $\tilde{\mu}_N = \tilde{\tau}_N \tilde{\kappa}_N = -52.45mV$ for the estimator based on the direct observations.

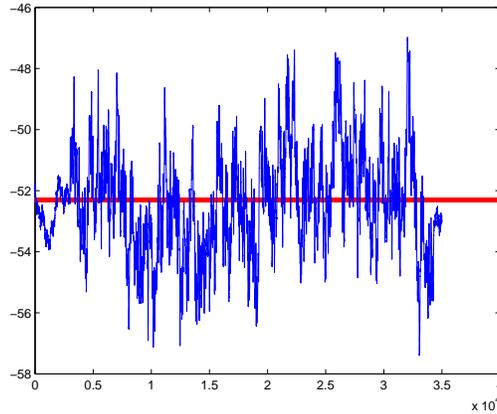


Figure 2. The observations with the estimated mean $\hat{\mu}_N$.

However, the estimated values of τ and λ^2 are significantly different, and $\hat{\tau}$ increases with the size p of the blocks.

6.4. The Cox-Ingersoll-Ross process

Consider the one-dimensional diffusion process (Cox-Ingersoll-Ross process), solution of

$$dX_t = (\kappa X_t + \kappa')dt + \lambda\sqrt{X_t}dB_t, \quad X_0 = \eta, \quad (6.3)$$

with $\kappa < 0$, $\kappa' \in \mathbb{R}$ and $\lambda > 0$, and η a positive random variable independent of $(B_t)/$

We assume that the observations at time $t_0 < \dots < t_N$ are given by

$$Y_{t_i} = X_{t_i} \exp(\varepsilon_{t_i})$$

where (ε_{t_i}) is a sequence of independent $\mathcal{N}(0, \rho^2)$ random variables. Hence the noise is multiplicative, and the observations remain positive.

We consider $U_{t_i} = \log(Y_{t_i})$ to have real valued observations.

The process $Z_t = \log(X_t)$ solves the stochastic differential equation

$$dZ_t = \left(\kappa + \kappa' - \frac{\lambda^2}{2}\right) \exp(-Z_t)dt + \lambda \exp\left(-\frac{Z_t}{2}\right)dB_t.$$

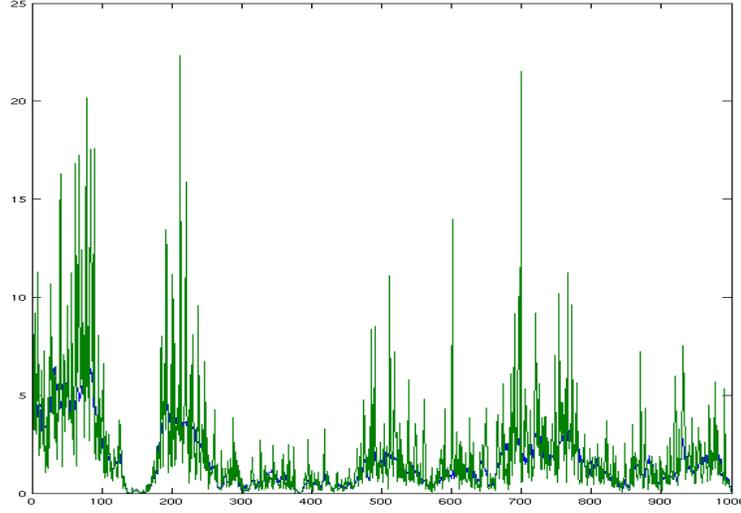


Figure 3. An example of Cox-Ingersoll-Ross process observed with a mutliplicative noise.

We set $\kappa'' = \kappa' - \frac{\lambda^2}{2}$.

In this case, the scale density is $s(x) = \exp\left(-\frac{2\kappa}{\lambda^2}e^x - \frac{2\kappa''}{\lambda^2}x\right)$ and the speed density is $m(x) = \frac{1}{\lambda^2} \exp\left(\left(\frac{2\kappa''}{\lambda^2} + 1\right)x + \frac{2\kappa}{\lambda^2}e^x\right)$. Provided $\kappa < 0$ and $\frac{2\kappa''}{\lambda^2} + 1 > 0$, Assumptions **(A2)**, **(A3)** are ensured, and **(A4)** holds with $\eta \sim \nu_0$. However, Assumption **(A1)** does not holds, but $\hat{\theta}_N$ is explicit, and the consistence can be proved directly.

Explicit expressions for the estimator $\hat{\theta}_N = (\hat{\kappa}_N, \hat{\kappa}''_N, \hat{\lambda}_N^2)$ are derived: $(\hat{\kappa}_N, \hat{\kappa}''_N)$ is solution of the system

$$\begin{pmatrix} \Delta_N \sum_{j=1}^{k_N-2} \exp(Y_{\bullet}^{j-1}) & \Delta_N k_N \\ \Delta_N k_N & \Delta_N \sum_{j=1}^{k_N-2} \exp(-Y_{\bullet}^{j-1}) \end{pmatrix} \begin{pmatrix} \hat{\kappa}_N \\ \hat{\kappa}''_N \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{k_N-2} \exp(Y_{\bullet}^{j-1})(Y_{\bullet}^{j+1} - Y_{\bullet}^j) \\ \sum_{j=1}^{k_N-2} (Y_{\bullet}^{j+1} - Y_{\bullet}^j) \end{pmatrix}$$

and

$$\hat{\lambda}_N^2 = \frac{3}{2k_N \Delta_N} \sum_{j=1}^{k_N-2} \exp(Y_{\bullet}^{j-1})(Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N(\hat{\kappa}_N + \hat{\kappa}''_N \exp(-Y_{\bullet}^{j-1})))^2.$$

Recall that the following explicit expressions for the estimator $\tilde{\theta}_N = (\tilde{\kappa}_N, \tilde{\kappa}''_N, \tilde{\lambda}_N^2)$ when the diffusion (X_t) is directly observed (Kessler (1997)):

$$\begin{pmatrix} \Delta_N \sum_{j=1}^{k_N-2} X_{j\Delta_N} & \Delta_N k_N \\ \Delta_N k_N & \Delta_N \sum_{j=1}^{k_N-2} \frac{1}{X_{j\Delta_N}} \end{pmatrix} \begin{pmatrix} \tilde{\kappa}_N \\ \tilde{\kappa}''_N \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{k_N-2} (X_{(j+1)\Delta_N} - X_{j\Delta_N}) \\ \sum_{j=1}^{k_N-2} \frac{1}{X_{j\Delta_N}} (X_{(j+1)\Delta_N} - X_{j\Delta_N}) \end{pmatrix}$$

and

$$\tilde{\lambda}_N^2 = \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} \frac{1}{X_{j\Delta_N}} (X_{(j+1)\Delta_N} - X_{j\Delta_N} - \Delta_N(\tilde{\kappa}_N X_{j\Delta_N} + \tilde{\kappa}'_N))^2.$$

Simulation results are presented in Table 10 (with noise) and Table 11 (directly observed). For this study, $N = 500$ trajectories are simulated with parameters $\kappa_0 = -2, \kappa'_0 = 3, \lambda_0 = 4, \rho^2 = 0.5$, and then $\kappa''_0 = 1$. Due to the simulation study for the Ornstein-Uhlenbeck process, the value $\alpha = \frac{3}{2}$ as local mean size parameter.

	$\kappa_0 = -2, \kappa'_0 = 1, \lambda_0 = 4, \rho^2 = 0.5, \alpha = 1.5$		
	$N = 5 \cdot 10^3, \delta = 10^{-2}$	$N = 2 \cdot 10^4, \delta = 5 \cdot 10^{-3}$	$N = 10^5, \delta = 10^{-3}$
$\hat{\kappa}_N$ (10^2 Var)	-1.43 (6.28)	-1.56 (3.14)	-1.78 (3.37)
$\hat{\kappa}''_N$ (10^2 Var)	0.99 (4.57)	1.03 (2.12)	1.13 (2.44)
$\hat{\lambda}_N^2$ (10^2 Var)	4.23 (37.61)	4.35 (15.15)	4.40 (8.15)

Table 10.

Parameter estimation for the Cox-Ingersoll-Ross process with a multiplicative noise for different values of α . (500 replications, $\kappa_0 = -2, \kappa'_0 = 1, \lambda_0 = 4, \rho^2 = 0.5, \alpha = \frac{3}{2}$)

In Table 10, we observe that $\kappa''_0 = 1$ is well estimated, whereas the estimation of κ_0 is negatively biased. The empirical variance, for $\hat{\kappa}_N$ and $\hat{\kappa}''_N$ decreases between $N = 5000$ and $N = 20000$ observations, but there is no significative improvement between $N = 20000$ and $N = 100000$ observations. For the diffusion parameter λ_0 , the estimator $\hat{\lambda}_N$ is positively biased, with a variance decreasing as the number of observations grows.

These results can be compared with the case of direct observations, given in Table 11.

	$\kappa_0 = -2, \kappa'_0 = 3, \lambda_0 = 4, \alpha = 1.5$		
	$N = 5 \cdot 10^3, \delta = 10^{-2}$	$N = 2 \cdot 10^4, \delta = 5 \cdot 10^{-3}$	$N = 10^5, \delta = 10^{-3}$
$\hat{\kappa}_N$ (10^2 Var)	-2.04 (11.03)	-2.03 (6.65)	-2.46 (53.45)
$\hat{\kappa}'_N$ (10^2 Var)	3.02 (13.47)	3.03 (8.17)	3.45 (65.44)
$\hat{\lambda}_N^2$ (10^2 Var)	4.11 (0.95)	4.05 (0.20)	4.01 (0.36)

Table 11. Parameter estimation for the Cox-Ingersoll-Ross process with direct observations for different values of α . (500 replications, $\kappa_0 = -2, \kappa'_0 = 3, \lambda_0 = 4, \rho^2 = 0.5, \alpha = \frac{3}{2}$)

Notice that there is no bias in the estimation of κ_0 and κ'_0 for $N = 5000$ and $N = 20000$, contrary to the noisy case. Moreover, the estimation of λ_0^2 is more accurate, with a lower empirical variance for $\hat{\lambda}_N^2$.

6.5. The hyperbolic diffusion

Consider the one dimensional diffusion process solution of

$$dX_t = \kappa X_t dt + \lambda \sqrt{1 + X_t^2} dB_t, \quad X_0 = \eta \in \mathbb{R}, \quad (6.4)$$

where η is a random variable independent of (B_t) , $\kappa < 0$ and $\lambda > 0$. In this case, the model is positive recurrent if $|\kappa| + \frac{\lambda^2}{2} > 0$, and in this case, its stationary distribution

has density

$$\nu(x) \propto \frac{1}{(1+x^2)^{1+\frac{|\kappa|}{\lambda^2}}}.$$

If $X_0 = \eta$ has distribution $\nu_0(x)dx$, then, $\sqrt{1 + \frac{2\kappa}{\lambda^2}}\eta$ has Student distribution which can be easily simulated. Even if η_0 has only a finite number of finite moments, and **(A4)** does not holds, for $2(1 + \frac{|\kappa|}{\lambda^2}) > k + 1$, ν_0 has a finite moment of order k .

Now we consider

$$G(x) = \int_0^x \frac{dx}{\lambda\sqrt{1+x^2}} = \frac{1}{\lambda} \arg \sinh(x).$$

By the Ito formula, $Z_t = G(X_t)$ satisfies

$$dZ_t = \beta(Z_t)dt + dB_t$$

with

$$\beta(z) = -\left(\frac{\kappa}{\lambda} + \frac{\lambda}{2}\right) \tanh(\lambda z).$$

Sample paths of this diffusion can be simulated exactly with the retrospective exact simulation algorithms proposed in Beskos et al. (2006).

We can compute explicitly the estimator $\hat{\theta}_N = (\hat{\kappa}_N, \hat{\lambda}_N^2)$ in this case:

$$\hat{\kappa}_N = \frac{1}{\Delta_N} \frac{\sum_{j=1}^{k_N-2} \frac{Y_{\bullet}^{j-1}}{1+(Y_{\bullet}^{j-1})^2} (Y_{\bullet}^{j+1} - Y_{\bullet}^j)}{\sum_{j=1}^{k_N-2} \frac{(Y_{\bullet}^{j-1})^2}{1+(Y_{\bullet}^{j-1})^2}}$$

and

$$\hat{\lambda}_N^2 = \frac{3}{2k_N\Delta_N} \sum_{j=1}^{k_N-2} \frac{(Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N \hat{\kappa}_N Y_{\bullet}^{j-1})^2}{1+(Y_{\bullet}^{j-1})^2}.$$

Some simulation results are given in Tables 12 and 13, with different distributions for the observation noise. In the different cases, $N = 500$ replications are made, and the empirical mean is given with the associated standard deviation in parenthesis. We consider for the values of the parameters: $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$.

	$\alpha = \frac{3}{2}, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5, \varepsilon \sim \mathcal{N}(0, 1)$		
	$N = 5 \cdot 10^3, \delta = 10^{-2} (p = 22)$	$N = 2 \cdot 10^4, \delta = 5 \cdot 10^{-3} (p = 34)$	$N = 10^5, \delta = 10^{-3} (p = 100)$
$\hat{\kappa}_N$ (Var)	-0.75 (0.21)	-0.82 (0.16)	-0.90 (0.17)
$\hat{\lambda}_N^2$ (Var)	1.14 (0.14)	1.11 (0.08)	1.07 (0.06)

Table 12. Parameter estimation for the hyperbolic diffusion, with a Gaussian noise, for different values of N . (500 replications, $\alpha = \frac{3}{2}, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$)

A negative bias is observed in the estimation of κ_0 , whereas a positive one appears in the estimation of λ_0^2 . For the two different noise distributions, empirical means and variances are very close in this model.

$$\alpha = \frac{3}{2}, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5, \varepsilon \sim \text{Laplace}(0, \frac{1}{\sqrt{2}})$$

	$N = 5 \cdot 10^3, \delta = 10^{-2} (p = 22)$	$N = 2 \cdot 10^4, \delta = 5 \cdot 10^{-3} (p = 34)$	$N = 10^5, \delta = 10^{-3} (p = 100)$
$\hat{\kappa}_N$ (Var)	-0.74 (0.21)	-0.82 (0.16)	-0.89 (0.18)
$\hat{\lambda}_N^2$ (Var)	1.15 (0.15)	1.12 (0.09)	1.08 (0.07)

Table 13. Parameter estimation for the hyperbolic diffusion, with a Laplace noise, for different values of N . (500 replications, $\alpha = \frac{3}{2}, \kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$)

7. Concluding remarks

The contrasts presented in this work give associated estimators for parameters involved in a non-Markovian setting: one-dimensional diffusions observed with a noise. The consistency of these minimum contrast estimators is illustrated on several simulations, and the estimated values are close to the values obtained for a direct observation of the diffusion. The importance of the sampling rate of local means, depending on the choice of α , appears in the case $\alpha = 2$ with $\rho_N = \rho$, where the variance of the observation noise ρ^2 appears in the limit theorem for the quadratic variation. The asymptotic normality is studied in a companion paper.

Acknowledgements

The author is greatly indebted to Valentine Genon-Catalot, his PhD advisor, for her help and her numerous comments on the redaction of this work, and Adeline Samson for the discussions about the numerical results and her insightful suggestions. He also thank Lee E. Moore for the dataset from his experiment on neuronal activity, and Yves Rozeholc for his help and his comments on the exact simulation algorithm.

8. Proofs

It is useful to introduce the intervals $I_{j,k,N} := I_{j,k} = [j\Delta_N + k\delta_N, j\Delta_N + (k+1)\delta_N)$, for $k = 0, \dots, p_N - 1, j = 0, \dots, k_N - 1$, which satisfy for all j , if $k \neq k'$ $I_{j,k} \cap I_{j,k'} = \emptyset$ and for $j \neq j'$ and all $k, k', I_{j,k} \cap I_{j',k'} = \emptyset$.

Lemma 8.1. *The random variables $\zeta_{j+1,N}$ and $\zeta'_{j+1,N}$ are $\mathcal{G}_{(j+1)\Delta_N}$ measurable, $\zeta'_{j+2,N}$ is independent of $\mathcal{G}_{(j+1)\Delta_N}$, and the following holds:*

$$\zeta_{j+1,N} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1) \int_{I_{j,k}} dB_s, \quad \zeta'_{j+2,N} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (p_N - 1 - k) \int_{I_{j+1,k}} dB_s. \quad (8.1)$$

Moreover, we have

$$\begin{aligned}\mathbb{E}(\zeta_{j,N}|\mathcal{G}_j^N) &= 0, & \mathbb{E}(\zeta'_{j+1,N}|\mathcal{G}_j^N) &= 0, & \mathbb{E}(\zeta_{j+1,N}\zeta'_{j+1,N}|\mathcal{G}_j^N) &= \frac{\Delta_N}{6} \left(1 - \frac{1}{p_N^2}\right), \\ \mathbb{E}((\zeta_{j+1,N})^2|\mathcal{G}_j^N) &= \Delta_N \left(\frac{1}{3} + \frac{1}{2p_N} + \frac{1}{6p_N^2}\right), & \mathbb{E}((\zeta'_{j+1,N})^2|\mathcal{G}_j^N) &= \Delta_N \left(\frac{1}{3} - \frac{1}{2p_N} + \frac{1}{6p_N^2}\right).\end{aligned}$$

Proof of Lemma 8.1 Using (3.1), we can rearrange terms to exhibit non-overlapping intervals, hence conditionally independent variables, and obtain (8.1). Afterwards, the proof is achieved by elementary computations.

□

Proof of Proposition 3.1 First, note that, as $(X_t, t \geq 0)$ and $(\varepsilon_{k\delta_N})$ are independent, for $l = 1, 2$,

$$\mathbb{E}(e_{j,N}^l|\mathcal{H}_{j,N}) = \mathbb{E}(e_{j,N}^l|\mathcal{G}_{j,N}).$$

Thus we study the expectations given $\mathcal{G}_{j,N}$. Using $\Delta_N = p_N\delta_N$ yields

$$R_{j,N} = \bar{X}_j - X_{\bullet}^j = \frac{1}{p_N} \sum_{k=0}^{p_N-1} \frac{1}{\delta_N} \int_{I_{j,k}} (X_s - X_{j\Delta_N+k\delta_N}) ds.$$

By the Fubini theorem, we get

$$R_{j,N} = \sqrt{\delta_N} \left(\frac{1}{p_N} \sum_{k=0}^{p_N-1} \sigma(X_{j\Delta_N+k\delta_N}) \xi'_{k,j,N} \right) + e_{j,N}$$

where $e_{j,N} = \alpha_{j,N} + \beta_{j,N}$, with

$$\alpha_{j,N} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} \frac{1}{\delta_N} \int_{I_{j,k}} (j\Delta_N + (k+1)\delta_N - s)(\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N})) dB_s$$

and

$$\beta_{j,N} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} \frac{1}{\delta_N} \int_{I_{j,k}} \int_{j\Delta_N+i\delta_N}^s b(X_u) du ds.$$

Under Assumption **(A1)**, we have $|\beta_{j,N}| \leq c\delta_N(1 + \sup_{s \in [j\Delta_N, (j+1)\Delta_N]} |X_s|)$. And for all $p \geq 0$, by (2.2),

$$\mathbb{E}(|\beta_{j,N}|^p|\mathcal{G}_j^N) \leq c\delta_N^p(1 + |X_{j\Delta_N}|^p).$$

Also $\mathbb{E}(\alpha_{j,N}|\mathcal{G}_j^N) = 0$, so we get $|\mathbb{E}(e_{j,N}|\mathcal{G}_j^N)| \leq \delta_N c(1 + |X_{j\Delta_N}|)$. Furthermore, we get with the Jensen inequality, the Ito isometry and the Fubini theorem

$$\mathbb{E}((\alpha_{j,N})^2|\mathcal{G}_j^N) \leq c \frac{1}{p_N} \sum_{k=0}^{p_N-1} \int_{I_{j,k}} \mathbb{E}((\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N}))^2|\mathcal{G}_j^N) ds.$$

With Proposition A.1 in the Appendix, it comes $\mathbb{E}(|\alpha_{j,N}|^2 | \mathcal{G}_j^N) \leq C\delta_N^2(1 + |X_{j\Delta_N}|^4)$. This implies the result. \square

Proof of Proposition 3.2 We have

$$Y_{\bullet}^j - X_{j\Delta_N} = X_{\bullet}^j - \bar{X}_j + \bar{X}_j - X_{j\Delta_N} + \rho_N \varepsilon_{\bullet}^j,$$

where ε_{\bullet}^j is independent of \mathcal{H}_j^N . Proposition 2.2 in Gloter (2000) states that, using the random variables (3.2),

$$\bar{X}_j - X_{j\Delta_N} = \sigma(X_{j\Delta_N})\sqrt{\Delta_N}\xi'_{j,N} + \bar{e}_{j,N}$$

with $|\mathbb{E}(\bar{e}_{j,N} | \mathcal{H}_j^N)| = |\mathbb{E}(\bar{e}_{j,N} | \mathcal{G}_j^N)| \leq c\Delta_N(1 + |X_{j\Delta_N}|)$ and $\mathbb{E}(\bar{e}_{j,N}^2 | \mathcal{H}_j^N) = \mathbb{E}(\bar{e}_{j,N}^2 | \mathcal{G}_j^N) \leq c\Delta_N^2(1 + |X_{j\Delta_N}|^4)$. With Proposition 3.1, setting $e'_{j,N} = e_{j,N} + \bar{e}_{j,N}$, we get the first part of Proposition 3.2. Now we need to prove that, for some $c > 0$

$$\mathbb{E}\left(|r_{j,N}|^k | \mathcal{H}_j^N\right) = \mathbb{E}\left(|r_{j,N}|^k | \mathcal{G}_j^N\right) \leq c(1 + |X_{j\Delta_N}|^k) \quad (8.2)$$

where

$$r_{j,N} = \frac{1}{p_N} \sum_{i=0}^{p_N-1} \sigma(X_{j\Delta_N+i\delta_N})\xi'_{i,j,N}$$

and $\xi'_{i,j,N}$ is defined in (3.3). With elementary computations on conditional expectation, we get (see notation (2.3))

$$\mathbb{E}\left(|r_{j,N}|^k | \mathcal{G}_j^N\right) \leq \frac{1}{p_N} \sum_{i=0}^{p_N-1} \mathbb{E}(|\sigma(X_{j\Delta_N+i\delta_N})|^k |\mathbb{E}(|\xi'_{i,j,N}|^k | \mathcal{G}_{j\Delta_N+i\delta_N}) | \mathcal{G}_j^N).$$

As $\xi'_{i,j,N}$ is independent of $\mathcal{G}_{j\Delta_N+i\delta_N}$,

$$\mathbb{E}\left(|r_{j,N}|^k | \mathcal{G}_j^N\right) \leq c \frac{1}{p_N} \sum_{i=0}^{p_N-1} \mathbb{E}(1 + |X_{j\Delta_N+i\delta_N}|^k | \mathcal{G}_j^N)$$

which implies (8.2). Finally, $\mathbb{E}(|\varepsilon_{\bullet}^j|^k | \mathcal{H}_j^N) = \mathbb{E}(|\varepsilon_{\bullet}^j|^k)$ because ε_{\bullet}^j is independent of \mathcal{H}_j^N . \square

Proof of Corollary 3.1 We have, with Taylor's formula (order two):

$$D_j := f(Y_{\bullet}^j, \theta) - f(X_{j\Delta_N}, \theta) = \partial_x f(X_{j\Delta_N}, \theta)(Y_{\bullet}^j - X_{j\Delta_N}) + \frac{1}{2} \partial_{xx}^2 f(Z, \theta)(Y_{\bullet}^j - X_{j\Delta_N})^2$$

with $Z \in (Y_{\bullet}^j, X_{j\Delta_N})$. Then, with the Cauchy Schwarz inequality, using that the derivatives satisfy **(C1)**, and Proposition 3.2, there exists a constant $c > 0$ such that, for all $\theta \in \Theta$,

$$\begin{aligned} |\mathbb{E}(D_j | \mathcal{H}_j^N)| &\leq c(1 + |X_{j\Delta_N}|) |\mathbb{E}(e'_{j,N} | \mathcal{H}_j^N)| \\ &\quad + c(1 + |X_{j\Delta_N}| + \rho_N \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^2)}) \sqrt{\mathbb{E}((Y_{\bullet}^j - X_{j\Delta_N})^4 | \mathcal{H}_j^N)} \\ &\leq c\Delta_N(1 + |X_{j\Delta_N}|^2) \\ &\quad + c(1 + |X_{j\Delta_N}| + \rho_N \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^2)}) \\ &\quad \times (\Delta_N(1 + |X_{j\Delta_N}|^2) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^4)}). \end{aligned}$$

With Taylor's formula (order one), there exists a random variable $\tilde{Z} \in (Y_{\bullet}^j, X_{j\Delta_N})$ and a constant $c > 0$ independent of θ such that $D_j^2 = (\partial_x f(\tilde{Z}, \theta))^2 (Y_{\bullet}^j - X_{j\Delta_N})^2$ and

$$D_j^2 \leq c(1 + \sup_{s \in [j\Delta_N, (j+1)\Delta_N]} [X_s]^2 + \rho_N^2 |\varepsilon_{\bullet}^j|^2) (Y_{\bullet}^j - X_{j\Delta_N})^2.$$

Using the Cauchy-Schwarz inequality and condition **(C1)**,

$$\mathbb{E}(D_j^2 | \mathcal{H}_j^N) \leq c(1 + |X_{j\Delta_N}|^2 + \rho_N^2 \mathbb{E}((\varepsilon_{\bullet}^j)^2)) (\Delta_N (1 + |X_{j\Delta_N}|^2) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^4)}).$$

Analogously, $D_j^4 = (\partial_x f(\tilde{Z}, \theta))^4 (Y_{\bullet}^j - X_{j\Delta_N})^4$ and

$$D_j^4 \leq c(1 + \sup_{s \in [j\Delta_N, (j+1)\Delta_N]} [X_s]^4 + \rho_N^4 |\varepsilon_{\bullet}^j|^4) (Y_{\bullet}^j - X_{j\Delta_N})^4,$$

with c independent of θ . Using the Cauchy-Schwarz inequality, it comes

$$\mathbb{E}(D_j^4 | \mathcal{H}_j^N) \leq c(1 + |X_{j\Delta_N}|^4 + \rho_N^4 \mathbb{E}((\varepsilon_{\bullet}^j)^4)) (\Delta_N^2 (1 + |X_{j\Delta_N}|^4) + \rho_N^4 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^8)}).$$

□

Proof of Proposition 3.3 In this proof, we study all conditional expectation given \mathcal{G}_j^N as they are identical to conditional expectations given \mathcal{H}_j^N in all the terms involved below. We have

$$Y_{\bullet}^{j+1} - Y_{\bullet}^j = X_{\bullet}^{j+1} - X_{\bullet}^j + \rho_N (\varepsilon_{\bullet}^{j+1} - \varepsilon_{\bullet}^j).$$

Setting $C_j = X_{\bullet}^{j+1} - X_{\bullet}^j$ and rearranging terms yields

$$\begin{aligned} C_j &= \frac{1}{p_N} \sum_{k=0}^{p_N-1} (X_{(j+1)\Delta_N+k\delta_N} - X_{j\Delta_N+k\delta_N}) \\ &= \frac{1}{p_N} \sum_{k=0}^{p_N-1} \sum_{l=0}^{p_N-1} \int_{I_{j,k+l}} dX_s \\ &= \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1) \int_{I_{j,k}} dX_s + \frac{1}{p_N} \sum_{k=0}^{p_N-1} (p_N - k - 1) \int_{I_{j+1,k}} dX_s. \end{aligned}$$

We use

$$\begin{aligned} \int_{I_{j,k}} dX_s &= b(X_{j\Delta_N+k\delta_N})\delta_N + \int_{I_{j,k}} (b(X_s) - b(X_{j\Delta_N+k\delta_N})) ds \\ &\quad + \sigma(X_{j\Delta_N+k\delta_N}) \int_{I_{j,k}} dB_s + \int_{I_{j,k}} (\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N})) dB_s. \end{aligned}$$

By splitting Δ_N into $\Delta_N = (k+1)\delta_N + (p_N - k - 1)\delta_N$ for all k , we get (see notation 3.1)

$$\begin{aligned} Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N b(Y_{\bullet}^j) &= C_j - \Delta_N b(Y_{\bullet}^j) + \rho_N (\varepsilon_{\bullet}^{j+1} - \varepsilon_{\bullet}^j) \\ &= \sigma(X_{j\Delta_N}) (\zeta_{j+1,N} + \zeta'_{j+2,N}) + \tau_{j,N} + \rho_N (\varepsilon_{\bullet}^{j+1} - \varepsilon_{\bullet}^j) \end{aligned}$$

where $\tau_{j,N} = \sum_{\ell=1}^4 \tau_{j,N}^{(\ell)}$ and for $\ell = 1, \dots, 4$, $\tau_{j,N}^{(\ell)} = r_{j,N}^{(\ell)} + s_{j,N}^{(\ell)}$ with

$$r_{j,N}^{(1)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1) \delta_N (b(X_{j\Delta_N+k\delta_N}) - b(Y_{\bullet}^j)), \quad (8.3)$$

$$s_{j,N}^{(1)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (p_N - k - 1) \delta_N (b(X_{(j+1)\Delta_N+k\delta_N}) - b(Y_{\bullet}^j)), \quad (8.4)$$

$$r_{j,N}^{(2)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1) \sigma(X_{j\Delta_N+k\delta_N}) \int_{I_{j,k}} dB_s - \sigma(X_{j\Delta_N}) \zeta_{j+1,N}, \quad (8.5)$$

$$s_{j,N}^{(2)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (p_N - k - 1) \sigma(X_{(j+1)\Delta_N+k\delta_N}) \int_{I_{j+1,k}} dB_s - \sigma(X_{j\Delta_N}) \zeta'_{j+2,N}, \quad (8.6)$$

$$r_{j,N}^{(3)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1) \int_{I_{j,k}} (b(X_s) - b(X_{j\Delta_N+k\delta_N})) ds, \quad (8.7)$$

$$s_{j,N}^{(3)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (p_N - k - 1) \int_{I_{j+1,k}} (b(X_s) - b(X_{(j+1)\Delta_N+k\delta_N})) ds, \quad (8.8)$$

$$r_{j,N}^{(4)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1) \int_{I_{j,k}} (\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N})) dB_s, \quad (8.9)$$

$$s_{j,N}^{(4)} = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (p_N - k - 1) \int_{I_{j+1,k}} (\sigma(X_s) - \sigma(X_{(j+1)\Delta_N+k\delta_N})) dB_s. \quad (8.10)$$

We mainly treat the terms $r_{j,N}^{(\ell)}$ because the others are analogous. We have $\mathbb{E}(r_{j,N}^{(\ell)} | \mathcal{G}_j^N) = 0$ and $\mathbb{E}(s_{j,N}^{(\ell)} | \mathcal{G}_j^N) = 0$ for $\ell = 2, 4$. Next,

$$|\mathbb{E}(r_{j,N}^{(1)} | \mathcal{G}_j^N)| \leq \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1) \delta_N |\mathbb{E}(b(X_{j\Delta_N+k\delta_N}) - b(Y_{\bullet}^j) | \mathcal{G}_j^N)|$$

We use, for $k = 0 \dots p_N - 1$ and $s \in I_{j,k}$, the inequality

$$|\mathbb{E}(b(X_s) - b(X_{j\Delta_N+k\delta_N}) | \mathcal{G}_j^N)| \leq c \Delta_N (1 + |X_{j\Delta_N}|^3).$$

With (3.6), it comes $|\mathbb{E}(r_{j,N}^{(1)} | \mathcal{G}_j^N)| \leq c \Delta_N (\Delta_N (1 + |X_{j\Delta_N}|^2) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^4)})$. Then, with the Fubini theorem, we derive $|\mathbb{E}(\tau_{j,N}^{(3)} | \mathcal{G}_j^N)| \leq c \Delta_N^2 (1 + |X_{j\Delta_N}|^3)$. Hence

$$|\mathbb{E}(\tau_{j,N} | \mathcal{G}_j^N)| \leq c \Delta_N (\Delta_N (1 + |X_{j\Delta_N}|^3) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^4)}).$$

Now we deal with $\mathbb{E}((r_{j,N}^{(1)})^2 | \mathcal{G}_j^N)$. With Proposition 3.1 and the Cauchy-Schwarz inequality, it comes

$$\mathbb{E}((b(Y_{\bullet}^j) - b(X_{j\Delta_N}))^2 | \mathcal{G}_j^N) \leq c(1 + |X_{j\Delta_N}|^2 + \rho_N^2 \mathbb{E}((\varepsilon_{\bullet}^j)^2)) (\Delta_N (1 + |X_{j\Delta_N}|^2) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^4)}).$$

Applying the Cauchy-Schwarz inequality, and after elementary computations, we obtain

$$\mathbb{E}((r_{j,N}^{(1)})^2 | \mathcal{G}_j^N) \leq c\Delta_N^2(1 + |X_{j\Delta_N}|^2 + \rho_N^2 \mathbb{E}((\varepsilon_{\bullet}^j)^2))(\Delta_N(1 + |X_{j\Delta_N}|^2) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^4)}).$$

With analogous techniques, we have

$$\begin{aligned} \mathbb{E}((\tau_{j,N}^{(3)})^2 | \mathcal{G}_j^N) &\leq c\Delta_N^2 \sup_{s \in [j\Delta_N, (j+2)\Delta_N]} \mathbb{E}((b(X_s) - b(X_{j\Delta_N}))^2 | \mathcal{G}_j^N) \\ &\leq c\Delta_N^3(1 + |X_{j\Delta_N}|^4). \end{aligned}$$

Using Lemma 8.1, we obtain

$$\begin{aligned} r_{j,N}^{(2)} &= \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1)(\sigma(X_{j\Delta_N+k\delta_N}) - \sigma(X_{j\Delta_N})) \int_{I_{j,k}} dB_s, \\ s_{j,N}^{(2)} &= \frac{1}{p_N} \sum_{k=0}^{p_N-1} (p_N - k - 1)(\sigma(X_{(j+1)\Delta_N+k\delta_N}) - \sigma(X_{j\Delta_N})) \int_{I_{j+1,k}} dB_s. \end{aligned}$$

Thus $r_{j,N}^{(2)} = \int_{j\Delta_N}^{(j+1)\Delta_N} f(s) dB_s$ with

$$f(s) = \frac{1}{p_N} \sum_{k=0}^{p_N-1} (k+1)(\sigma(X_{j\Delta_N+k\delta_N}) - \sigma(X_{j\Delta_N})) \mathbf{1}_{I_{j,k}}(s) \quad (8.11)$$

With the Ito isometry and the Fubini theorem, we have

$$\begin{aligned} \mathbb{E}((r_{j,N}^{(2)})^2 | \mathcal{G}_j^N) &= \frac{1}{p_N^2} \sum_{k=0}^{p_N-1} (k+1)^2 \delta_N \mathbb{E}((\sigma(X_{j\Delta_N+k\delta_N}) - \sigma(X_{j\Delta_N}))^2 | \mathcal{G}_j^N) \\ &\leq c\Delta_N^2(1 + |X_{j\Delta_N}|^4) \end{aligned}$$

We use similar techniques with $r_{j,N}^{(4)}$ and $s_{j,N}^{(4)}$ to obtain

$$\mathbb{E}((\tau_{j,N}^{(2)})^2 + (\tau_{j,N}^{(4)})^2 | \mathcal{G}_j^N) \leq c\Delta_N^2(1 + |X_{j\Delta_N}|^4).$$

Collecting terms, we get the bound for $\mathbb{E}(\tau_{j,N}^2 | \mathcal{G}_j^N)$.

Now, using (8.3), (3.1), Lemma 8.1 and the Cauchy Schwarz inequality we have

$$|\mathbb{E}(r_{j,N}^{(1)} \zeta_{j+1,N} | \mathcal{G}_j^N)| \leq c\Delta_N^{\frac{3}{2}} \sqrt{\mathbb{E}((b(X_{j\Delta_N+k\delta_N}) - b(Y_{\bullet}^j))^2 | \mathcal{G}_j^N)}.$$

Corollary 3.1 implies

$$|\mathbb{E}(r_{j,N}^{(1)} \zeta_{j+1,N} | \mathcal{G}_j^N)| \leq c\Delta_N^{\frac{3}{2}}(1 + |X_{j\Delta_N}| + \rho_N \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^2)})(\sqrt{\Delta_N}(1 + |X_{j\Delta_N}|) + \rho_N(\mathbb{E}((\varepsilon_{\bullet}^j)^4))^{\frac{1}{4}}).$$

The same inequality holds for $\mathbb{E}(r_{j,N}^{(1)} \zeta'_{j+2,N} | \mathcal{G}_j^N)$, $\mathbb{E}(s_{j,N}^{(1)} \zeta_{j+1,N} | \mathcal{G}_j^N)$ and $\mathbb{E}(s_{j,N}^{(1)} \zeta'_{j+2,N} | \mathcal{G}_j^N)$.

We can write $\zeta_{j+1,N} = \int_{j\Delta_N}^{(j+1)\Delta_N} g(s)dB_s$ with $g(s) = \frac{1}{p_N} \sum_{l=0}^{p_N-1} (l+1)\mathbf{1}_{I_{j,l}}(s)$. Using (8.11) and Corollary 3.1, we obtain

$$\begin{aligned} |\mathbb{E}(r_{j,N}^{(2)}\zeta_{j+1,N}|\mathcal{G}_j^N)| &\leq \frac{1}{p_N^2} \sum_{k=0}^{p_N-1} (k+1)^2 \delta_N |\mathbb{E}(\sigma(X_{j\Delta_N+k\delta_N}) - \sigma(X_{j\Delta_N})|\mathcal{G}_j^N)| \\ &\leq c\Delta_N(1 + |X_{j\Delta_N}| + \rho_N^2 \mathbb{E}((\varepsilon_{\bullet}^j)^2))(\Delta_N(1 + |X_{j\Delta_N}|^2) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^4)}). \end{aligned}$$

The same inequality holds for $|\mathbb{E}(r_{j,N}^{(2)}\zeta'_{j+2,N}|\mathcal{G}_j^N)|$.

For $r_{j,N}^{(3)}$ (see (8.7)), we use the Cauchy Schwarz inequality:

$$|\mathbb{E}(r_{j,N}^{(3)}\zeta_{j+1,N}|\mathcal{G}_j^N)| \leq \frac{1}{p_N^2} \sum_{k,l=0}^{p_N-1} (k+1)(l+1)\delta_N^{3/2} \mathbb{E} \left(\sup_{s \in I_{j,k}} (b(X_s) - b(X_{j\Delta_N+k\delta_N}))^2 \middle| \mathcal{G}_j^N \right)^{\frac{1}{2}}.$$

Hence

$$|\mathbb{E}(r_{j,N}^{(3)}\zeta_{j+1,N}|\mathcal{G}_j^N)| \leq c\Delta_N^2(1 + |X_{j\Delta_N}|^2).$$

Furthermore $\mathbb{E}(r_{j,N}^{(3)}\zeta'_{j+2,N}|\mathcal{G}_j^N) = 0$.

With the Fubini theorem and the Ito isometry, we have

$$\mathbb{E}(r_{j,N}^{(4)}\zeta_{j+1,N}|\mathcal{G}_j^N) = \frac{1}{p_N^2} \sum_{k=0}^{p_N-1} (k+1)^2 \int_{I_{j,k}} \mathbb{E}(\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N})|\mathcal{G}_j^N) ds$$

Introducing $Lf = \frac{\sigma^2}{2}f'' + bf'$ yields

$$\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N}) = \int_{j\Delta_N+k\delta_N}^s L\sigma(X_u)du + \frac{1}{2} \int_{j\Delta_N+k\delta_N}^s \sigma(X_u)\sigma'(X_u)dB_u.$$

Therefore, $|\mathbb{E}(\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N})|\mathcal{G}_j^N)| \leq c\Delta_N(1 + |X_{j\Delta_N}|^4)$ which implies

$$|\mathbb{E}(r_{j,N}^{(4)}\zeta_{j+1,N}|\mathcal{G}_j^N)| \leq c\Delta_N^2(1 + |X_{j\Delta_N}|^4).$$

Furthermore $\mathbb{E}(r_{j,N}^{(4)}\zeta'_{j+2,N}|\mathcal{G}_j^N) = 0$. The terms containing $s_{j,N}^{(3)}$ and $s_{j,N}^{(4)}$ are treated analogously. This gives the bound for $|\mathbb{E}(\tau_{j,N}\zeta_{j+1,N}|\mathcal{G}_j^N)|$ and $|\mathbb{E}(\tau_{j,N}\zeta'_{j+2,N}|\mathcal{G}_j^N)|$.

Finally, we have to bound the fourth order conditional moment of $\tau_{j,N}$. We only study the terms $r_{j,N}^{(2)}$ and $r_{j,N}^{(1)}$. Using (8.11), the Burkholder - Davies - Gundy inequality and Proposition A.1, we have

$$\begin{aligned} \mathbb{E}((r_{j,N}^{(2)})^4|\mathcal{G}_j^N) &\leq c\mathbb{E} \left(\left(\int_{j\Delta_N}^{(j+1)\Delta_N} f(s)^2 ds \right)^2 \middle| \mathcal{G}_j^N \right) \\ &\leq c\Delta_N^2 \mathbb{E} \left(\sup_{s \in [j\Delta_N, (j+1)\Delta_N]} (\sigma(X_s) - \sigma(X_{j\Delta_N}))^4 \middle| \mathcal{G}_j^N \right) \leq c\Delta_N^4(1 + |X_{j\Delta_N}|^4). \end{aligned}$$

With similar computations, we derive $\mathbb{E}((\tau_{j,N}^{(2)})^4 + (\tau_{j,N}^{(4)})^4 | \mathcal{G}_j^N) \leq c\Delta_N^4(1 + |X_{j\Delta_N}|^4)$.

Using Proposition 3.1, we get

$$\begin{aligned} \mathbb{E}((r_{j,N}^{(1)})^4 | \mathcal{G}_j^N) &\leq c \frac{\delta_N^4}{p_N} \sum_{k=0}^{p_N-1} (k+1)^4 \mathbb{E}((b(Y_{\bullet}^j) - b(X_{j\Delta_N+k\delta_N}))^4 | \mathcal{G}_j^N) \\ &\leq c(1 + |X_{j\Delta_N}|^4 + \rho_N^4 \mathbb{E}((\varepsilon_{\bullet}^j)^4)) (\Delta_N^6(1 + |X_{j\Delta_N}|^4) + \rho_N^4 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^8)}). \end{aligned}$$

Analogously, using Proposition A.1, $\mathbb{E}((r_{j,N}^{(3)})^4 | \mathcal{G}_j^N) \leq c\Delta_N^6(1 + |X_{j\Delta_N}|^4)$. Finally, we get the bound for $\mathbb{E}(\tau_{j,N}^4 | \mathcal{G}_j^N)$.

□

Proof of Proposition 4.1 By Lemma A.2, it is enough to prove the L^1 convergence to zero of

$$\sup_{\theta \in \Theta} \frac{1}{k_N} \sum_{j=0}^{k_N-1} |f(Y_{\bullet}^j, \theta) - f(X_{j\Delta_N}, \theta)|.$$

By Taylor expansion and condition **(C1)** we derive the bound

$$A_j := \sup_{\theta \in \Theta} |f(Y_{\bullet}^j, \theta) - f(X_{j\Delta_N}, \theta)| \leq c(1 + |X_{j\Delta_N}| + |Y_{\bullet}^j|) |Y_{\bullet}^j - X_{j\Delta_N}|.$$

Hence, the Cauchy Schwarz inequality and Assumption **(A2)** imply

$$\mathbb{E}(A_j | \mathcal{H}_j^N) \leq c(1 + |X_{j\Delta_N}| + \rho_N \sqrt{\mathbb{E}((\varepsilon_{\bullet}^j)^2)}) \sqrt{\mathbb{E}(|Y_{\bullet}^j - X_{j\Delta_N}|^2 | \mathcal{H}_j^N)}.$$

Then, with (3.5), Assumptions **(A5)** and **(B1)**, and $\mathbb{E}((\varepsilon_{\bullet}^j)^2) = \frac{1}{p_N}$, the result holds. □

Proof of Theorem 4.1 We have

$$\bar{I}_N(f(\cdot, \theta)) = \tilde{I}_N(f(\cdot, \theta)) + \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} f(Y_{\bullet}^{j-1}, \theta) \Delta_N (b(Y_{\bullet}^j) - b(Y_{\bullet}^{j-1})),$$

where $\tilde{I}_N(f(\cdot, \theta)) = \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} V_j^N(\theta)$ with $V_j^N(\theta) = f(Y_{\bullet}^{j-1}, \theta) (Y_{\bullet}^{j+1} - Y_{\bullet}^j - \Delta_N b(Y_{\bullet}^j))$. We only need to prove that $\tilde{I}_N(f(\cdot, \theta)) \rightarrow 0$ in probability, uniformly in $\theta \in \Theta$, as the second term is $o_P(1)$, uniformly in θ . As $V_j^N(\theta)$ is \mathcal{H}_{j+2}^N -measurable, we split the sum into three parts

$$\sum_{j=1}^{k_N-2} V_{j,N}(\theta) = \sum_{1 \leq 3j \leq k_N-2} V_{3j,N}(\theta) + \sum_{1 \leq 3j+1 \leq k_N-2} V_{3j+1,N}(\theta) + \sum_{1 \leq 3j+2 \leq k_N-2} V_{3j+2,N}(\theta).$$

We treat only the sum with indexes multiples of 3 and set:

$$V_{3j}^N(\theta) = v_{3j,N}^{(1)}(\theta) + v_{3j,N}^{(2)}(\theta) + v_{3j,N}^{(3)}(\theta)$$

where

$$\begin{aligned} v_{3j,N}^{(1)}(\theta) &= f(Y_{\bullet}^{3j-1}, \theta) \sigma(X_{3j\Delta_N}) (\zeta_{3j+1,N} + \zeta'_{3j+2,N}), \\ v_{3j,N}^{(2)}(\theta) &= f(Y_{\bullet}^{3j-1}, \theta) \rho_N (\varepsilon_{\bullet}^{3j+1} - \varepsilon_{\bullet}^{3j}), \\ v_{3j,N}^{(3)}(\theta) &= f(Y_{\bullet}^{3j-1}, \theta) \tau_{3j,N}. \end{aligned}$$

In order to prove the pointwise convergence in θ to zero, we use Lemma A.3. As Y_{\bullet}^{3j-1} , $X_{3j\Delta_N}$ are \mathcal{H}_{3j}^N -measurables and $\varepsilon_{\bullet}^{3j+1} - \varepsilon_{\bullet}^{3j}$ is independent of \mathcal{H}_{3j}^N , we have $\mathbb{E}(v_{3j,N}^{(1)}(\theta)|\mathcal{H}_{3j}^N) = 0$ and $\mathbb{E}(v_{3j,N}^{(2)}(\theta)|\mathcal{H}_{3j}^N) = 0$. By Proposition 3.3,

$$|\mathbb{E}(\tau_{3j,N}|\mathcal{H}_{3j}^N)| \leq c\Delta_N(1 + |X_{3j\Delta_N}|^2 + \rho_N^2 \mathbb{E}((\varepsilon_{\bullet}^{3j})^2))(\Delta_N(1 + |X_{3j\Delta_N}|^4) + \rho_N^2 \sqrt{\mathbb{E}((\varepsilon_{\bullet}^{3j})^4)}).$$

Using **(A4)**, this implies $\frac{1}{k_N\Delta_N} \sum_{1 \leq 3j \leq k_N-2} \mathbb{E}(v_{3j,N}^{(3)}(\theta)|\mathcal{H}_{3j}^N) = o_P(1)$. We also have to verify for $\ell = 1, 2, 3$,

$$\frac{1}{(k_N\Delta_N)^2} \sum_{j=1}^{k_N-2} \mathbb{E}((v_{3j,N}^{(\ell)}(\theta))^2|\mathcal{H}_{3j}^N) = o_P(1).$$

For $\ell = 1$, we have

$$\begin{aligned} & \frac{1}{(k_N\Delta_N)^2} \sum_{1 \leq 3j \leq k_N-2} \mathbb{E}((v_{3j,N}^{(1)}(\theta))^2|\mathcal{H}_{3j}^N) \\ &= \frac{1}{(k_N\Delta_N)^2} \sum_{1 \leq 3j \leq k_N-2} f(Y_{\bullet}^{3j-1}, \theta)^2 \sigma(X_{3j\Delta_N})^2 \mathbb{E}\left(\left(\zeta_{3j+1,N} + \zeta'_{3j+2,N}\right)^2 \middle| \mathcal{H}_{3j}^N\right) \\ &\leq \frac{1}{N\delta_N} \frac{2}{k_N} \sum_{1 \leq 3j \leq k_N-2} f(Y_{\bullet}^{3j-1}, \theta)^2 \sigma(X_{3j\Delta_N})^2 = o_P(1). \end{aligned}$$

For $\ell = 2$,

$$\begin{aligned} \frac{1}{(k_N\Delta_N)^2} \sum_{1 \leq 3j \leq k_N-2} \mathbb{E}((v_{3j,N}^{(2)}(\theta))^2|\mathcal{H}_{3j}^N) &= \frac{1}{(k_N\Delta_N)^2} \sum_{1 \leq 3j \leq k_N-2} f(Y_{\bullet}^{3j-1}, \theta)^2 \rho_N^2 \mathbb{E}((\varepsilon_{\bullet}^{3j+1} - \varepsilon_{\bullet}^{3j})^2) \\ &= \frac{2\rho_N^2}{N\delta_N p_N \Delta_N} \frac{1}{k_N} \sum_{1 \leq 3j \leq k_N-2} f(Y_{\bullet}^{3j-1}, \theta)^2. \end{aligned}$$

As $p_N \Delta_N = p_N^{2-\alpha}$, with $1 < \alpha \leq 2$, the above term is $o_P(1)$.

For $\ell = 3$,

$$\frac{1}{(k_N\Delta_N)^2} \sum_{j=1}^{k_N-2} \mathbb{E}((v_{3j,N}^{(3)}(\theta))^2|\mathcal{H}_{3j}^N) = \frac{1}{k_N} \frac{1}{k_N} \sum_{j=1}^{k_N-2} f_{\theta}(Y_{\bullet}^{3j-1})^2 \frac{1}{\Delta_N^2} \mathbb{E}(\tau_{j,N}^2|\mathcal{H}_{3j}^N) = o_P(1),$$

using that, by Proposition 3.3, $\Delta_N^{-2} \mathbb{E}(\tau_{j,N}^2|\mathcal{H}_{3j}^N)$ is $O_P(1)$.

To obtain uniformity in θ , we shall use Proposition A.2 and evaluate $\sup_{N \in \mathbb{N}} \mathbb{E}(\sup_{\theta \in \Theta} |\partial_\theta \tilde{I}_N(f_\theta)|)$. To study

$$\partial_\theta \tilde{I}_N(f_\theta) = \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} \partial_\theta V_j^N(\theta),$$

we use the same method, split the sum in three parts, and define:

$$S_N^{(\ell)}(\theta) = \frac{1}{k_N \Delta_N} \sum_{1 \leq 3j \leq k_N-2} v_{3j,N}^{(\ell)}(\theta).$$

The sum for $\ell = 3$ is the simplest. With assumption **(C1)** for $\partial_\theta f$, we deduce

$$\mathbb{E}(\sup_{\theta \in \Theta} |\partial_\theta v_{3j,N}^{(3)}(\theta)| | \mathcal{H}_{3j}^N) \leq c(1 + |Y_{\bullet}^{3j-1}|) \sqrt{\mathbb{E}(\tau_{3j,N}^2 | \mathcal{H}_{3j}^N)}.$$

With the Cauchy Schwarz inequality, we have

$$\begin{aligned} \mathbb{E}(\sup_{\theta \in \Theta} |\partial_\theta v_{3j,N}^{(3)}(\theta)| | \mathcal{H}_{3j}^N) &\leq c \sqrt{\Delta_N} (1 + |Y_{\bullet}^{3j-1}|) (1 + |X_{3j\Delta_N}| + \rho_N \sqrt{\mathbb{E}((\varepsilon_{\bullet}^{3j})^2)}) \\ &\quad \times (\sqrt{\Delta_N} (1 + |X_{3j\Delta_N}|^2) + \rho_N \left(\mathbb{E}((\varepsilon_{\bullet}^{3j})^4) \right)^{\frac{1}{4}}) \end{aligned}$$

and with Lemma A.2 and **(A4)-(A5)**, this implies $\sup_{N \in \mathbb{N}} \mathbb{E}(\sup_{\theta \in \Theta} |\partial_\theta S_N^{(3)}(\theta)|) < \infty$.

We cannot use the same method to study $S_N^{(\ell)}(\theta)$, $\ell = 1, 2$. Instead, we use Theorem 20 in Appendix 1 of Ibragimov and Has'minskiĭ (1979): it is enough to show that, for $\ell = 1, 2$, there exists two constants $M \geq 0$ and $\epsilon > 0$ such that:

$$\begin{aligned} \forall \theta \in \Theta, \forall N \in \mathbb{N}, \quad \mathbb{E}(|S_N^{(\ell)}|^{2+\epsilon}) &\leq M \\ \text{and } \forall \theta, \theta' \in \Theta, \forall N \in \mathbb{N}, \quad D_N(\theta, \theta') &\leq M |\theta - \theta'|^{2+\epsilon} \end{aligned} \quad (8.12)$$

where $D_N(\theta, \theta') = \mathbb{E}(|S_N^{(\ell)}(\theta) - S_N^{(\ell)}(\theta')|^{2+\epsilon})$.

For $\ell = 1$, using the Rosenthal inequality for martingales, we get, for any $\epsilon > 0$:

$$\begin{aligned} \mathbb{E}(|S_N^{(1)}(\theta)|^{2+\epsilon}) &\leq \frac{1}{(k_N \Delta_N)^{2+\epsilon}} \mathbb{E} \left(\left| \sum_{1 \leq 3j \leq k_N-2} \mathbb{E} \left((v_{3j,N}^{(1)}(\theta))^2 \mid \mathcal{H}_{3j}^N \right) \right|^{1+\frac{\epsilon}{2}} \right) \\ &\quad + \frac{1}{(k_N \Delta_N)^{2+\epsilon}} \sum_{1 \leq 3j \leq k_N-2} \mathbb{E}(|v_{3j,N}^{(1)}(\theta)|^{2+\epsilon}) \end{aligned}$$

Then it comes:

$$\mathbb{E} \left(\left| \sum_{1 \leq 3j \leq k_N-2} \mathbb{E} \left((v_{3j,N}^{(1)}(\theta))^2 \mid \mathcal{H}_{3j}^N \right) \right|^{1+\frac{\epsilon}{2}} \right) \leq k_N^{\frac{\epsilon}{2}} \sum_{1 \leq 3j \leq k_N-2} \mathbb{E} \left(\left| \mathbb{E} \left((v_{3j,N}^{(1)}(\theta))^2 \mid \mathcal{H}_{3j}^N \right) \right|^{1+\frac{\epsilon}{2}} \right)$$

With $\mathbb{E}((\zeta_{3j+1,N} + \zeta'_{3j+2,N})^2 | \mathcal{H}_{3j}^N) = \Delta_N \left(1 - \frac{1}{3} \left(\frac{p_N^2 - 1}{p_N^2}\right)\right)$, Assumption **(A5)** and **(C1)**, we derive

$$\sup_{j,N} \mathbb{E} \left(\left| \mathbb{E} \left((v_{3j,N}^{(1)}(\theta))^2 \middle| \mathcal{H}_{3j}^N \right) \right|^{1+\frac{\epsilon}{2}} \right) \leq c \Delta_N^{1+\frac{\epsilon}{2}} \quad \text{and} \quad \sup_{j,N} \mathbb{E} \left(\left| v_{3j,N}^{(1)}(\theta) \right|^{2+\epsilon} \right) \leq c \Delta_N^{1+\frac{\epsilon}{2}}.$$

Hence

$$\mathbb{E} \left(\left| S_N^{(1)}(\theta) \right|^{2+\epsilon} \right) \leq c \left(\frac{1}{(k_N \Delta_N)^{1+\frac{\epsilon}{2}}} + \frac{1}{(k_N \Delta_N)^{1+\frac{\epsilon}{2}}} \frac{1}{k_N^{\frac{\epsilon}{2}}} \right).$$

The study of $D_N(\theta, \theta')$ is analogous, so (8.12) holds. This implies $S_N^{(1)}(\theta) = o_P(1)$ uniformly in θ .

We use similar tools for $S_N^{(2)}$. With the Rosenthal inequality, we have

$$\begin{aligned} \mathbb{E}(|S_N^{(2)}(\theta)|^{2+\epsilon}) &\leq \frac{1}{(k_N \Delta_N)^{2+\epsilon}} \mathbb{E} \left(\left| \sum_{1 \leq 3j \leq k_N - 2} \mathbb{E} \left((v_{3j,N}^{(2)}(\theta))^2 \middle| \mathcal{H}_{3j}^N \right) \right|^{1+\frac{\epsilon}{2}} \right) \\ &\quad + \frac{1}{(k_N \Delta_N)^{2+\epsilon}} \sum_{1 \leq 3j \leq k_N - 2} \mathbb{E}(|v_{3j,N}^{(2)}(\theta)|^{2+\epsilon}). \end{aligned}$$

Hence, with $\mathbb{E} \left((v_{3j,N}^{(2)}(\theta))^2 \middle| \mathcal{H}_{3j}^N \right) = 2\rho_N^2 f(Y_{\bullet}^{3j-1}, \theta)^2 \sigma(X_{3j\Delta_N})^2 \mathbb{E}((\varepsilon_{\bullet}^{3j})^2)$ and $\mathbb{E}((\varepsilon_{\bullet}^{3j})^2) = \frac{1}{p_N}$, and $\Delta_N = p_N^{1-\alpha}$, we obtain (8.12). Finally $\tilde{I}_N(f_\theta) = o_P(1)$, uniformly in θ . \square

Proof of Theorem 4.2 Let $W_{j,N}(\theta) = f(Y_{\bullet}^{j-1}, \theta)(Y_{\bullet}^{j+1} - Y_{\bullet}^j)^2$. By Proposition 3.3, we have $W_{j,N}(\theta) = \sum_{i=1}^6 w_{j,N}^{(i)}(\theta)$ with

$$\begin{aligned} w_{j,N}^{(1)}(\theta) &= f(Y_{\bullet}^{j-1}, \theta) \sigma(X_{j\Delta_N})^2 (\zeta_{j+1,N} + \zeta'_{j+2,N})^2 \\ w_{j,N}^{(2)}(\theta) &= f(Y_{\bullet}^{j-1}, \theta) \rho_N^2 (\varepsilon_{\bullet}^{j+1} - \varepsilon_{\bullet}^j)^2 \\ w_{j,N}^{(3)}(\theta) &= f(Y_{\bullet}^{j-1}, \theta) (\Delta_N b(Y_{\bullet}^j) + \tau_{j,N})^2 \\ w_{j,N}^{(4)}(\theta) &= f(Y_{\bullet}^{j-1}, \theta) 2\sigma(X_{j\Delta_N}) (\zeta_{j+1,N} + \zeta'_{j+2,N}) \rho_N (\varepsilon_{\bullet}^{j+1} - \varepsilon_{\bullet}^j) \\ w_{j,N}^{(5)}(\theta) &= f(Y_{\bullet}^{j-1}, \theta) 2\sigma(X_{j\Delta_N}) (\zeta_{j+1,N} + \zeta'_{j+2,N}) (\Delta_N b(Y_{\bullet}^j) + \tau_{j,N}) \\ w_{j,N}^{(6)}(\theta) &= f(Y_{\bullet}^{j-1}, \theta) 2\rho_N (\varepsilon_{\bullet}^{j+1} - \varepsilon_{\bullet}^j) (\Delta_N b(Y_{\bullet}^j) + \tau_{j,N}), \end{aligned}$$

where we recall that $Y_{\bullet}^{j-1}, X_{j\Delta_N}$ are \mathcal{H}_j^N -measurable and $\varepsilon_{\bullet}^{j+1} - \varepsilon_{\bullet}^j$ is independent of \mathcal{H}_j^N . Therefore, splitting again into three parts, we consider, for $\ell = 0, 1, 2$,

$$T_{\ell,N}^{(i)}(\theta) = \frac{1}{k_N \Delta_N} \sum_{1 \leq 3j + \ell \leq k_N - 2} w_{3j+\ell,N}^{(i)}(\theta) \quad \text{for } i = 1, \dots, 6.$$

We start by studying $T_{0,N}^{(1)}(\theta)$:

$$\mathbb{E}(w_{3j,N}^{(1)}(\theta) | \mathcal{H}_{3j}^N) = f(Y_{\bullet}^{3j-1}, \theta) \sigma(X_{3j\Delta_N})^2 \Delta_N \left(1 - \frac{1}{3} \left(\frac{p_N^2 - 1}{p_N^2}\right)\right)$$

and

$$\mathbb{E}((w_{3j,N}^{(1)}(\theta))^2 | \mathcal{H}_{3j}^N) = 3f(Y_{\bullet}^{3j-1}, \theta)^2 \sigma(X_{3j\Delta_N})^4 \Delta_N^2 \left(\frac{2}{3} + \frac{1}{3p_N^2} \right)^2.$$

Applying Lemma A.3 with Lemma A.2, we get, for all θ , $T_{0,N}^{(1)}(\theta) = \frac{1}{3} \times \frac{2}{3} \nu_0(f(\cdot, \theta) \sigma^2) + o_P(1)$. Thus

$$T_{0,N}^{(1)}(\theta) + T_{1,N}^{(1)}(\theta) + T_{2,N}^{(1)}(\theta) = \frac{2}{3} \nu_0(f(\cdot, \theta) \sigma^2) + o_P(1).$$

Then, we study $T_{0,N}^{(2)}(\theta)$:

$$\begin{aligned} \mathbb{E}(w_{3j,N}^{(2)}(\theta) | \mathcal{H}_{3j}^N) &= f(Y_{\bullet}^{3j-1}, \theta) \rho_N^2 \mathbb{E}((\varepsilon_{\bullet}^{3j+1} - \varepsilon_{\bullet}^{3j})^2) \\ &= 2f(Y_{\bullet}^{3j-1}, \theta) \rho_N^2 p_N^{-1} \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}((w_{3j,N}^{(2)}(\theta))^4 | \mathcal{H}_{3j}^N) &= f(Y_{\bullet}^{3j-1}, \theta)^2 \rho_N^4 \mathbb{E}((\varepsilon_{\bullet}^{3j+1} - \varepsilon_{\bullet}^{3j})^4) \\ &= f(Y_{\bullet}^{3j-1}, \theta)^2 \rho_N^4 (12p_N^{-2}(1 + o(1))) \end{aligned}$$

Recall that $\Delta_N = p_N^{1-\alpha}$, $1 < \alpha \leq 2$. If $\alpha < 2$, with Lemma A.3, $T_{0,N}^{(2)} = o_P(1)$. But if $\alpha = 2$, *i.e.* $\Delta_N = \frac{1}{p_N}$, and $\rho_N = \rho$, we have $T_{0,N}^{(2)}(\theta) = \frac{1}{3} \times 2\rho^2 \nu_0(f(\cdot, \theta)^2) + o_P(1)$. and

$$T_{0,N}^{(2)}(\theta) + T_{1,N}^{(2)}(\theta) + T_{2,N}^{(2)}(\theta) = 2\rho^2 \nu_0(f(\cdot, \theta)^2) + o_P(1).$$

We easily deduce from Proposition 3.3, Lemma A.3 and Lemma A.2 that $T_{0,N}^{(3)}(\theta) = o_P(1)$.

For $T_{0,N}^{(4)}(\theta)$, we have

$$\mathbb{E}(w_{3j,N}^{(4)}(\theta) | \mathcal{H}_{3j}^N) = 2f(Y_{\bullet}^{3j-1}, \theta) \sigma(X_{3j\Delta_N}) \rho_N \mathbb{E}((\zeta_{3j+1,N} + \zeta'_{3j+2,N})(\varepsilon_{\bullet}^{3j+1} - \varepsilon_{\bullet}^{3j}) | \mathcal{H}_{3j}^N)$$

Given \mathcal{H}_{3j}^N , the random variables $(\zeta_{3j+1,N} + \zeta'_{3j+2,N})$ and $(\varepsilon_{\bullet}^{3j+1} - \varepsilon_{\bullet}^{3j})$ are independent, so $\mathbb{E}(w_{3j,N}^{(4)}(\theta) | \mathcal{H}_{3j}^N) = 0$. Furthermore

$$\begin{aligned} \mathbb{E}((w_{3j,N}^{(4)}(\theta))^2 | \mathcal{H}_{3j}^N) &= 4f(Y_{\bullet}^{3j-1}, \theta)^2 \sigma(X_{3j\Delta_N})^2 \rho_N^2 \mathbb{E}((\zeta_{3j+1,N} + \zeta'_{3j+2,N})^2 (\varepsilon_{\bullet}^{3j+1} - \varepsilon_{\bullet}^{3j})^2 | \mathcal{H}_{3j}^N) \\ &= 8f(Y_{\bullet}^{3j-1}, \theta)^2 \sigma(X_{3j\Delta_N})^2 \rho_N^2 \Delta_N \left(\frac{2}{3} + \frac{1}{3p_N^2} \right) \frac{1}{p_N}. \end{aligned}$$

Then, with Proposition 3.3, Lemma A.3 and Lemma A.2, $T_{0,N}^{(4)}(\theta) = o_P(1)$.

We have

$$\mathbb{E}(w_{3j,N}^{(5)}(\theta) | \mathcal{H}_{3j}^N) = 2f(Y_{\bullet}^{3j-1}, \theta) \sigma(X_{3j\Delta_N}) \mathbb{E}((\zeta_{3j+1,N} + \zeta'_{3j+2,N})(\Delta_N b(Y_{\bullet}^{3j}) + \tau_{3j,N}) | \mathcal{H}_{3j}^N).$$

With the Cauchy Schwarz inequality,

$$\begin{aligned} |\mathbb{E}(w_{3j,N}^{(5)}(\theta) | \mathcal{H}_{3j}^N)| &\leq c|f(Y_{\bullet}^{3j-1}, \theta)| \sigma(X_{3j\Delta_N}) \sqrt{\Delta_N} \sqrt{\mathbb{E}((\Delta_N b(Y_{\bullet}^{3j}) + \tau_{3j,N})^2 | \mathcal{H}_{3j}^N)} \\ &\leq c|f(Y_{\bullet}^{3j-1}, \theta)| \sigma(X_{3j\Delta_N}) \sqrt{\Delta_N} \sqrt{\Delta_N^2 \mathbb{E}(b(Y_{\bullet}^{3j})^2 | \mathcal{H}_{3j}^N) + \mathbb{E}(\tau_{3j,N}^2 | \mathcal{H}_{3j}^N)}. \end{aligned}$$

Moreover, with the Cauchy Schwarz inequality,

$$\begin{aligned}\mathbb{E}((w_{3j,N}^{(5)}(\theta))^2 | \mathcal{H}_{3j}^N) &= 4f(Y_{\bullet}^{3j-1}, \theta)^2 \sigma(X_{3j\Delta_N})^2 \mathbb{E}((\zeta_{3j+1,N} + \zeta'_{3j+2,N})^2 (\Delta_N b(Y_{\bullet}^{3j}) + \tau_{3j,N})^2 | \mathcal{H}_{3j}^N) \\ &\leq cf(Y_{\bullet}^{3j-1}, \theta)^2 \sigma(X_{3j\Delta_N})^2 \Delta_N^2 \sqrt{\mathbb{E}((\Delta_N b(Y_{\bullet}^{3j}) + \tau_{3j,N})^4 | \mathcal{H}_{3j}^N)}.\end{aligned}$$

Then, with Proposition 3.3, Lemma A.3 and Lemma A.2 $T_{0,N}^{(5)} = o_P(1)$.

With some straightforward computations, $T_{3j,N}^{(6)} = o_P(1)$.

We prove now uniformity in θ in these convergences, using Proposition A.2. For $w_{j,N}^{(1)}(\theta)$, we get

$$\mathbb{E} \left(\sup_{\theta \in \Theta} \left| \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} \partial_{\theta} w_{j,N}^{(1)}(\theta) \right| \right) < \infty$$

with

$$\mathbb{E}(\sigma(X_{j\Delta_N})^2 (\zeta_{j+1,N} + \zeta'_{j+2,N})^2 | \mathcal{H}_j^N) \leq c \Delta_N \sigma(X_{j\Delta_N})^2.$$

With similar arguments for $w_{j,N}^{(i)}(\theta)$, $i = 2 \dots 6$, we derive uniformity in θ .

□

Proof of Lemma 5.1 We have $\hat{\rho}_N^2 - \rho^2 = a_{1,N} + a_{2,N} + a_{3,N}$ where

$$\begin{aligned}a_{1,N} &= \frac{\rho^2}{2N} \sum_{i=0}^{N-1} \{(\varepsilon_{(i+1)\delta_N} - \varepsilon_{i\delta_N})^2 - 2\}, & a_{2,N} &= \frac{1}{2N} \sum_{i=0}^{N-1} (X_{(i+1)\delta_N} - X_{i\delta_N})^2, \\ a_{3,N} &= \frac{\rho}{N} \sum_{i=0}^{N-1} (X_{(i+1)\delta_N} - X_{i\delta_N})(\varepsilon_{(i+1)\delta_N} - \varepsilon_{i\delta_N}).\end{aligned}$$

With the usual law of large numbers, $a_{1,N} = o_P(1)$. With Proposition A.1,

$$\mathbb{E}(a_{2,N}) \leq c\delta_N(1 + \sup_{t \geq 0} \mathbb{E}(X_t^2)) = \delta_N O(1), \quad \mathbb{E}((a_{2,N})^2) \leq \frac{c\delta_N}{N}.$$

Hence $\hat{\rho}_N - \rho^2 = o_P(1)$. Moreover, $\sqrt{N}a_{2,N} = \sqrt{N}\delta_N O_P(1)$ and $\sqrt{N}a_{3,N} = \sqrt{\delta_N} O_P(1)$ tend to 0 as $N \rightarrow \infty$ for $N\delta_N^2 = o(1)$. To study the main term, let us set $u_i = \frac{\rho^2}{\sqrt{N}}(\varepsilon_{i\delta_N}^2 - 1 - \varepsilon_{(i-1)\delta_N} \varepsilon_{i\delta_N})$ so that $\sqrt{N}a_{1,N} = \sum_{i=1}^{N-1} u_i + o_P(1)$. With

$$\begin{aligned}\mathbb{E}(u_i | \varepsilon_{\ell\delta_N}, \ell \leq i-1) &= 0, & \sum_{i=1}^{N-1} \mathbb{E}(u_i^2 | \varepsilon_{\ell\delta_N}, \ell \leq i-1) &= 3\rho^4 + o_P(1), \\ \mathbb{E}(u_i^4 | \varepsilon_{\ell\delta_N}, \ell \leq i-1) &= o_P(1),\end{aligned}$$

we conclude by the Central Limit Theorem for martingale arrays. □

Proof of Theorem 5.1. For this proof, recall that $b(\cdot) = b(\cdot, \kappa_0)$, $c(\cdot) = c(\cdot, \lambda_0)$ denote the drift and diffusion coefficients at the true value θ_0 . Developing $\mathcal{E}_N(\theta)$ (see (5.1))

yields:

$$\begin{aligned} \mathcal{E}_N(\theta) &= k_N \left\{ \frac{3}{2} \bar{Q}_N \left(\frac{1}{c(\cdot, \lambda)} + \bar{\nu}_N(\log(c(\cdot, \lambda))) \right) \right\} \\ &\quad + 3k_N \Delta_N \left\{ \frac{1}{2} \bar{\nu}_N \left(\frac{b(\cdot, \kappa)^2 - 2b(\cdot, \kappa)b(\cdot, \kappa_0)}{c(\cdot, \lambda)} \right) - \bar{I}_N \left(\frac{b(\cdot, \kappa)}{c(\cdot, \lambda)} \right) \right\}. \end{aligned}$$

Proposition 4.1, Theorem 4.1 and Theorem 4.2 imply that $\mathcal{E}_N(\theta)$ is the sum of two terms with different rates of convergence. Therefore, to prove consistency of $\hat{\theta}_N$, we must proceed in two steps as in Kessler (1997) and Gloter (2006). It is enough to prove that, first,

$$\frac{1}{k_N} \mathcal{E}_N(\theta) \xrightarrow{N \rightarrow \infty} \nu_0 \left(\frac{c(\cdot, \lambda_0)}{c(\cdot, \lambda)} + \log(c(\cdot, \lambda)) \right) \quad (8.13)$$

in probability, uniformly in θ . This will ensure the convergence of $\hat{\lambda}_N$ to λ_0 . Second, we prove that

$$\frac{1}{k_N \Delta_N} (\mathcal{E}_N(\kappa, \lambda) - \mathcal{E}_N(\kappa_0, \lambda)) \xrightarrow{N \rightarrow \infty} \frac{3}{2} \nu_0 \left(\frac{(b(\cdot, \kappa) - b(\cdot, \kappa_0))^2}{c(\cdot, \lambda)} \right) \quad (8.14)$$

in probability, uniformly in θ .

Using Theorem 4.1, Theorem 4.2 and Proposition 4.1, with $\Delta_N \rightarrow 0$ we obtain (8.13) and (8.14).

For the second case, we have $\|c_{N,\rho}(\cdot, \lambda) - c_\rho(\cdot, \lambda)\|_\infty = 0$ if $\alpha = 2$, and

$$\|c_{N,\rho}(\cdot, \lambda) - c_\rho(\cdot, \lambda)\|_\infty \leq 3\Delta_N^{\frac{2-\alpha}{\alpha-1}} \rho^2 \text{ if } \alpha \in (1, 2).$$

Then, $c_{N,\rho}$ converges uniformly (in (x, λ)) to c_ρ . Moreover, by Assumption **(A7)**, c^{-1} satisfies **(C1)**. Thus

$$|c_{N,\rho}(x, \lambda)^{-1} - c_\rho(x, \lambda)^{-1}| \leq c \|c_{N,\rho}(\cdot, \lambda) - c_\rho(\cdot, \lambda)\|_\infty (1 + |x|^4)$$

and

$$|\log(c_{N,\rho}(x, \lambda)) - \log(c_\rho(x, \lambda))| \leq c \|c_{N,\rho}(\cdot, \lambda) - c_\rho(\cdot, \lambda)\|_\infty (1 + |x|^2).$$

The end of the proof is identical, replacing \mathcal{E}_N by \mathcal{E}_N^ρ and c by c_ρ in the limits (8.13)-(8.14). \square

Proof of Corollary 5.1. As formerly, we evaluate

$$\|c_{N,\hat{\rho}_N}(\cdot, \lambda) - c_\rho(\cdot, \lambda)\|_\infty \leq 3\Delta_N^{\frac{2-\alpha}{\alpha-1}} |\rho^2 - \hat{\rho}_N^2|.$$

We conclude using Lemma 5.1. \square

References

- Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, 70(1):223–262.
- Baltazar-Larios, F. and Sørensen, M. (2009). Maximum likelihood estimation for integrated diffusion processes. *Preprint*, 1(1).
- Beskos, A., Papaspiliopoulos, O., and Roberts, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli*, 12(6):1077–1098.
- Bibby, B. M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1(1-2):17–39.
- Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in hidden Markov models*. Springer Series in Statistics. Springer, New York. With Randal Douc’s contributions to Chapter 9 and Christian P. Robert’s to Chapters 6, 7 and 13, With Chapter 14 by Gersende Fort, Philippe Soulier and Moulines, and Chapter 15 by Stéphane Boucheron and Elisabeth Gassiat.
- Ditlevsen, S. and Ditlevsen, O. (2007). Parameter estimation from observations of first-passage times of the Ornstein-Uhlenbeck process and the Feller process. *Probabilistic Engineering Mechanics*, 23(1).
- Ditlevsen, S. and Lansky, P. (2005). Estimation of the input parameters in the ornstein-uhlenbeck neuronal model. *Phys. Rev. E*, 71(1):011907.
- Favetto, B. and Samson, A. (2009). Parameter estimation for a bidimensional partially observed Ornstein-Uhlenbeck process with biological application. *Scand J Statist*, to appear.
- Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, 20(4):547–557.
- Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations. *Statistics*, 21(1):99–116.
- Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 29(1):119–151.
- Gloter, A. (2000). Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Statist.*, 4:205–227 (electronic).
- Gloter, A. (2006). Parameter estimation for a discretely observed integrated diffusion process. *Scand. J. Statist.*, 33(1):83–104.
- Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. II. Optimal estimators. *ESAIM Probab. Statist.*, 5:243–260 (electronic).
- Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York. Probability and Mathematical Statistics.
- Höpfner, R. and Brodda, K. (2006). A stochastic model and a functional central limit theorem for information processing in large systems of neurons. *J. Math. Biol.*, 52(4):439–457.
- Ibragimov, I. A. and Has’minskiĭ, R. Z. (1979). *Asimptoticheskaya teoriya otsenivaniya*. “Nauka”, Moscow.

- Idoux, E., Serafin, M., Fort, P., Vidal, P., Beraneck, M., Vibert, N., MÃ¼hlethaler, M., and Moore, L. (2006). Oscillatory and intrinsic membrane properties of guinea pig nucleus prepositus hypoglossi neurons in vitro. *J Neurophysiol*, 96(1):175–196.
- Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Process. Appl.*, 119(7):2249–2276.
- Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, 24(2):211–229.
- Pedersen, A. (1994). Statistical analysis of gaussian diffusion processes based on incomplete discrete observations. *Research Report, Department of Theoretical Statistics, University of Aarhus*, 297.
- Pedersen, A. R. (1995a). Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli*, 1(3):257–279.
- Pedersen, A. R. (1995b). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.*, 22(1):55–71.
- Rogers, L. C. G. and Williams, D. (2000). *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge. Itô calculus, Reprint of the second (1994) edition.
- Sørensen, M. (2009). Efficient estimation for ergodic diffusions sampled at high frequency. *Preprint*, 1(1).
- Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.*, 41(2):220–242.
- Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 100(472):1394–1411.

Appendix

The following lemma can be found in Gloter (2006), and precises a result from Kessler (1997) :

Lemma A.2. *Assume (A1)-(A3). Let $f \in \mathcal{C}^1(\mathbb{R} \times O)$, where O is an open neighbourhood of Θ , satisfy*

$$\sup_{\theta \in \Theta} \{|f(x, \theta)| + |\partial_x f(x, \theta)| + |\partial_\theta f(x, \theta)|\} \leq C(1 + |x|)$$

then:

$$\frac{1}{k_N} \sum_{j=0}^{k_N-1} f(X_{j\Delta_N}, \theta) \xrightarrow[k_N \rightarrow \infty]{} \nu_0(f(\cdot, \theta)) \quad (\text{A.15})$$

uniformly in θ , in probability.

The following proposition can be found in Gloter (2000) and Gloter (2006), and the numerical constant c may varies.

Proposition A.1. *Assume (A1) and let $f \in \mathcal{C}^1(\mathbb{R})$ satisfy:*

$$\exists \gamma \geq 0, \exists c > 0, \forall x \in \mathbb{R} |f'(x)| \leq c(1 + |x|).$$

Then for all integer $k \geq 1$, there exists $c > 0$ such that, for all $j \geq 0$:

$$\mathbb{E} \left(\sup_{s \in [j\Delta_N, (j+1)\Delta_N]} |f(X_s) - f(X_{j\Delta_N})|^k | \mathcal{G}_j^N \right) \leq c \Delta_N^{\frac{k}{2}} (1 + |X_{j\Delta_N}|^{1+k})$$

In particular, with $f(x) = x$, we have:

$$\mathbb{E} \left(\sup_{s \in [j\Delta_N, (j+1)\Delta_N]} |X_s - X_{j\Delta_N}|^k \middle| \mathcal{G}_j^N \right) \leq c \Delta_N^{k/2} (1 + |X_{j\Delta_N}|^k).$$

We also recall the following lemma which is given in Genon-Catalot and Jacod (1993), setting $\mathcal{G}_j^N = \mathcal{G}_{j\Delta_N}$

Lemma A.3. *Let χ_j^N, U be random variables, with χ_j^N being \mathcal{G}_j^N -measurable. The following two conditions:*

$$\begin{aligned} \sum_{j=0}^{k_N-1} \mathbb{E}(\chi_j^N | \mathcal{G}_{j-1}^N) &\xrightarrow{\mathbb{P}} U, \\ \sum_{j=0}^{k_N-1} \mathbb{E}((\chi_j^N)^2 | \mathcal{G}_{j-1}^N) &\xrightarrow{\mathbb{P}} 0 \end{aligned}$$

imply $\sum_{j=0}^{k_N-1} \chi_j^N \xrightarrow{\mathbb{P}} U$.

The following proposition is given in Gloter (2006), to obtain convergences in probability uniformly in θ .

Proposition A.2. *Let $S_n(\omega, \theta)$ be a sequence of measurable real valued functions defined on $\Omega \times \Theta$ where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and Θ is product of compact intervals of $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. We assume that $S_n(\cdot, \theta)$ converges to zero in probability for all $\theta \in \Theta$ and that there exists an open neighbourhood of Θ on which $S_n(\omega, \cdot)$ is continuously differentiable for all $\omega \in \Omega$. Furthermore, we suppose that $\sup_{n \in \mathbb{N}} \mathbb{E}(\sup_{\theta \in \Theta} |\nabla_{\theta} S_n(\theta)|) < \infty$. Then*

$$S_n(\theta) \rightarrow 0$$

uniformly in θ , in probability.

Lemma A.4. *The random variables $\xi_{j,N}$ and $\xi'_{j+1,N}$ are independent and gaussian; $\xi_{j,N}$ is \mathcal{G}_{j+1}^N measurable and independent of \mathcal{G}_j^N ; $\xi'_{j+1,N}$ is \mathcal{G}_{j+2}^N measurable and independent of \mathcal{G}_{j+1}^N . We will use the following expectations:*

$$\begin{aligned}\mathbb{E}(\xi_{j,N}|\mathcal{G}_j^N) &= \mathbb{E}(\xi'_{j+1,N}|\mathcal{G}_j^N) = 0, \\ \mathbb{E}(\xi_{j,N}^2|\mathcal{G}_j^N) &= \mathbb{E}(\xi'^2_{j+1,N}|\mathcal{G}_j^N) = \frac{1}{3}, \\ \mathbb{E}((\xi_{j,N} - \frac{1}{3})^2|\mathcal{G}_j^N) &= \mathbb{E}((\xi'^2_{j+1,N} - \frac{1}{3})^2|\mathcal{G}_j^N) = \frac{2}{9}, \\ \mathbb{E}((\xi_{j,N} - \frac{1}{3})\xi'_{j,N}|\mathcal{G}_j^N) &= \mathbb{E}((\xi'^2_{j+1,N} - \frac{1}{3})\xi'_{j,N}|\mathcal{G}_j^N) = 0, \\ \mathbb{E}(\xi_{j,N}\xi'_{j,N}|\mathcal{G}_j^N) &= \frac{1}{6}.\end{aligned}$$

This lemma, based on elementary computations, is mentioned in Gloter (2000).