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

Benjamin Favetto

*Statistics*
Vol. 00, No. 00, December 2011, 1–29

ARTICLE

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

Benjamin Favetto[a*]

[a] *Université Paris Descartes*

(*v2 released November 2012*)

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. Some examples are discussed with numerical simulations.

**Keywords:** contrast function; diffusion process; hidden Markov models; parametric inference; discrete time noisy observations

**AMS Subject Classification**: 62M09; 62F12

## 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. More precisely, let $(X_t)$ be given by the stochastic differential equation:

$$dX_t = b(X_t, \kappa)dt + \sigma(X_t, \lambda)dB_t, \qquad X_0 = \eta \tag{1}$$

with $B$ a standard Wiener process and $\eta$ a random variable independent of $B$, and $b(., \kappa), \sigma(., \lambda)$ real valued functions, defined on $\mathbb{R}$, depending on unknown parameters $(\kappa, \lambda) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. The estimation of $\theta = (\kappa, \lambda)$ based on a discrete sample $(X_{i\delta}, i \leq N)$ with small sampling interval $\delta$ has been largely investigated. (see *e.g.* [6], [7] for contrast-based estimator of the drift parameter, [22] for maximum likelihood estimator, [8] for the estimation of the diffusion coefficient of multidimensional diffusion process, [16] for the case of an ergodic diffusion observed on a long-time interval, [2], [21], [1] ...)

*Email: benjamin.favetto@gmail.com

ISSN: 0233-1888 print/ISSN 1029-4910 online

DOI: 10.1080/0233188YYxxxxxxxx
http://www.informaworld.com

In this paper, we suppose that, instead of observing exactly $X_{i\delta}$, the observation at time $i\delta$ is given by

$$Y_{i\delta} = X_{i\delta} + \rho\varepsilon_{i\delta} \tag{2}$$

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. In this context, the estimation of the integrated volatility has been widely investigated (see *e.g* [23]). Jacod *et al.* in [15] consider the same kind of observations for $\delta = \delta_N$, over an interval of length $N\delta_N = t$ fixed, to estimate the integrated volatility $\int_0^t \sigma(X_s, \lambda)^2 ds$.

From now on, our concern is the joint estimation of $\theta = (\kappa, \lambda)$ using discrete observations 2 over a long-time interval.

The exact likelihood of $(Y_{i\delta}, i \leq N)$ given by (1)-(2) is generally intractable except for few models (essentially for Gaussian diffusions with additive Gaussian noise, see *e.g.* [3], [17], [5]). 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 [12]. The authors use a contrast method and only diffusion coefficient parameters can be consistently estimated in this case. For the nonparametric case, the inference of the drift function and the diffusion coefficient have been studied in [20] and [19].

In this paper, we study observations given by (1)-(2) where $\delta = \delta_N \to 0$ while $N\delta_N \to \infty$, under ergodic properties for the hidden diffusion $X$ and propose consistent estimators of both the drift and diffusion coefficient parameters $(\kappa, \lambda)$. The noise distribution is unknown, the variance $\rho^2$ of the noise term may be known or unknown and we assume that $\rho$ is fixed.

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, the sample is split into $k$ blocks of size $p$, 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 $\delta_N$ are chosen such that $\Delta_N \to 0$, we build the empirical mean of the $j^{th}$ block:

$$Y_{\bullet}^{j} = X_{\bullet}^{j} + \rho\varepsilon_{\bullet}^{j}, \quad j = 0, 1 \ldots k_N - 1, \tag{3}$$

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

$$Z_{\bullet}^{j} = \frac{1}{p_N} \sum_{i=0}^{p_N - 1} Z_{j\Delta_N + i\delta_N}. \tag{4}$$

Thus, $\Delta_N$ defines a coarser sampling interval than $\delta_N$, still tending to 0 while $N\delta_N = k_N\Delta_N \to \infty$.

Our statistical procedure is based on the $k_N-$ sample $(Y_{\bullet}^{j}, j = 0 \ldots k_N - 1)$ and follows a scheme analogous to the one in [11]. Hence, the empirical mean $X_{\bullet}^{j} = \frac{1}{p_N}\sum_{i=0}^{p_N-1} X_{j\Delta_N + i\delta_N}$ of the diffusion is closed to the integrated process $\frac{1}{\Delta_N}\int_{j\Delta_N}^{(j+1)\Delta_N} X_s ds$ as $\delta_N$ is sufficiently small. The parameter estimation of $\kappa$ and $\lambda$ based of the observations of an integrated diffusion process has been investigated by Gloter in [9], [10] and [11]. Our approach is based on these considerations.

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 3.4) which is the base of the statistical applications. These results allow us to prove limit theorems for the variation and the quadratic variation of $(Y_\bullet^j)$ which hold by setting $\delta_N = p_N^{-\alpha}$ with $1 < \alpha \leq 2$ (Theorems 4.2 and 4.3). We introduce contrasts and prove the consistency of the associated minimum contrast estimators. The study of the asymptotic distributions of the minimum contrast estimators is studied in another paper, as it requires further developments (see [4]).

The paper is organised as follows. In Section 2, notations and assumptions on the model are precised. 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 $\rho$ unknown and prove that $\rho^2$ can be replaced by an estimator in the contrast formula. Section 6 is devoted to examples and numerical results. For several models, we implement our estimators on simulated data for different choices of $(N, \delta_N, p_N)$ and of the 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

$$dX_t = b(X_t, \kappa_0)dt + \sigma(X_t, \lambda_0)dB_t, \qquad X_0 = \eta \tag{5}$$

where $B$ is a standard Brownian motion and $\eta$ 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 $\Theta_1$ (resp. $\Theta_2$) 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 $\sigma$ ensuring that (5) 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 $\sigma$ belong to $\mathcal{C}^2(\mathbb{R})$, $\sigma(x) > 0$ for all $x$, and there exists $c > 0$ such that for all $x \in \mathbb{R}$:

$$|b(x)| + |b'(x)| + |b''(x)| \leq c(1 + |x|),$$
$$\sigma(x) + |\sigma'(x)| + |\sigma''(x)| \leq c(1 + |x|).$$

**(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 $\int_{-\infty} s(x)dx = \int^{+\infty} s(x)dx = \infty$ and $\int_{-\infty}^{+\infty} m(x)dx = M < \infty$.

**(A3)** Let $\nu_0(dx) = \frac{1}{M}m(x)dx$. For all $k > 0$, $\nu_0$ 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.

Assumption **(A1)** implies that (1) admits a unique strong solution on $\mathbb{R}$. Under **(A1)** and **(A2)**, $\nu_0$ is the unique invariant probability of (5) and $(X_t)$ satisfies the classical ergodic theorem (see *e.g.* [18])

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

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

$$\mathbb{E}\left(\sup_{s\in[t,t+1]} |X_s|^k \middle| \mathcal{G}_t\right) \leq c(k)(1+|X_t|^k). \tag{6}$$

where $\mathcal{G}_t = \sigma(B_s, s \leq t; \eta)$. (See *e.g* [9]). Furthermore, Assumptions **(A1)-(A3)** imply **(A4)** if $\eta$ has distribution $\nu_0$ or $\eta$ is deterministic (for the latter case, see [11], Proposition 3). Below, we first assume that the noise level $\rho$ is known and discuss later the case where $\rho$ is unknown.

Define the $\sigma$-fields

$$\begin{aligned} \mathcal{G}_j^N = \mathcal{G}_{j\Delta_N} = \sigma(B_s, s \leq j\Delta_N; \eta), \quad \mathcal{H}_j^N = \mathcal{G}_j^N \vee \mathcal{A}_j^N, \\ \mathcal{A}_j^N = \sigma(\varepsilon_{k\Delta_N + i\delta_N}, i \leq p_N - 1, k \leq j-1) = \sigma(\varepsilon_{l\delta_N}, l \leq j\Delta_N - \delta_N) \end{aligned} \tag{7}$$

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 $g : \mathbb{R} \times \Theta \longrightarrow \mathbb{R}$:

**(C1)** The function $g$ is the restriction of a function defined on $\mathbb{R} \times \mathcal{O}$ with $\mathcal{O}$ an open neighbourhood of $\Theta$ 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 and **(A7)** is a smoothness condition for the contrast):

**(A6)**

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

**(A7)** The partial derivatives $\partial_x b$, $\partial_\kappa b$, $\partial_x \sigma$, $\partial_\lambda \sigma$, $\partial^2_{xx} b$, $\partial^2_{\kappa\kappa} b$, $\partial^2_{x\kappa} b$, $\partial_{xx}\sigma$, $\partial^2_{\lambda\lambda}\sigma$ and $\partial^2_{x\lambda}\sigma$ exist, are continuous and satisfy Condition **(C1)**.

## 3. Small sample properties of the local means sample

In this section, some local properties of the local means are gathered ton enlight first order approximation of $Y_\bullet^j - X_{j\Delta_N}$ and $Y_\bullet^{j+1} - Y_\bullet^j$.

The following random variables appear in the expansions below:

$$\zeta_{j+1,N} = \frac{1}{p_N}\sum_{i=0}^{p_N-1}\int_{j\Delta_N+i\delta_N}^{(j+1)\Delta_N} dB_s, \qquad \zeta'_{j+2,N} = \frac{1}{p_N}\sum_{i=0}^{p_N-1}\int_{(j+1)\Delta_N}^{(j+1)\Delta_N+i\delta_N} dB_s, \tag{8}$$

Consider also the following random variables which will appear in further expansions:

$$\xi'_{j+1,N} = \frac{1}{\Delta_N^{3/2}} \int_{(j+1)\Delta_N}^{(j+2)\Delta_N} ((j+2)\Delta_N - s) dB_s, \tag{9}$$

$$\xi'_{i+1,j,N} = \frac{1}{\delta_N^{3/2}} \int_{j\Delta_N+(i+1)\delta_N}^{j\Delta_N+(i+2)\delta_N} (j\Delta_N + (i+2)\delta_N - s) dB_s. \tag{10}$$

Some basic properties of these random variables are summarized in Lemma 8.1 and in Lemma 8.2 in Section 8.

**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 (7))*

$$\exists c > 0, \quad |\mathbb{E}(e_{j,N}|\mathcal{H}_j^N)| \leq \delta_N c(1 + |X_{j\Delta_N}|), \qquad \mathbb{E}(e_{j,N}^2|\mathcal{H}_j^N) \leq \delta_N^2 c(1 + |X_{j\Delta_N}|^4).$$

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 [9].

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

$$Y_\bullet^j - X_{j\Delta_N} = \sigma(X_{j\Delta_N}) \sqrt{\Delta_N} \xi'_{j,N} + e'_{j,N} + \rho \varepsilon_\bullet^j \tag{11}$$

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

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

*If moreover* **(A5)** *holds, for* $k \leq 8$,

$$\exists c > 0, \forall j \leq k_N - 1, \mathbb{E}\left( |Y_\bullet^j - X_{j\Delta_N}|^k \middle| \mathcal{H}_j^N \right) \leq c \left( \Delta_N^{k/2}(1 + |X_{j\Delta_N}|^k) + \rho^k \mathbb{E}\left(|\varepsilon_\bullet^j|^k\right) \right). \tag{12}$$

We deduce:

**Corollary 3.3:** *Assume* **(A1)** *and* **(A5)**, *and consider* $f : \mathbb{R}^2 \times \Theta \to \mathbb{R}$ *such that* $f, \partial_x f, \partial_{xx}^2 f$ *satisfy* **(C1)**. *Then*

$$\exists c > 0, \forall j \geq 0, \forall \theta \in \Theta, \left| \mathbb{E}\left( f(Y_\bullet^j, \theta) - f(X_{j\Delta_N}, \theta) \middle| \mathcal{H}_j^N \right) \right| \leq c(\Delta_N(1 + |X_{j\Delta_N}|^2) + \rho^2 \sqrt{\mathbb{E}((\varepsilon_\bullet^j)^4)}) \tag{13}$$

*and for* $l = 1, 2$

$$\begin{aligned} \mathbb{E}\left( (f(Y_\bullet^j, \theta) - f(X_{j\Delta_N}, \theta))^{2l} \middle| \mathcal{H}_j^N \right) &\leq c(1 + |X_{j\Delta_N}|^{2l} + \rho^{2l} \mathbb{E}((\varepsilon_\bullet^j)^{2l})) \\ &\quad \times (\Delta_N^l(1 + |X_{j\Delta_N}|^{2l}) + \rho^{2l} \sqrt{\mathbb{E}((\varepsilon_\bullet^j)^{4l})}). \end{aligned} \tag{14}$$

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

**Proposition 3.4:** *Under Assumptions* **(A1)** *and* **(A5)**, *we have*

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

*where* $\tau_{j,N}$ *is* $\mathcal{H}_{j+2}^N$ *mesurable, and there exists a constant* $c > 0$ *such that*

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

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

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

Note that, for $i = 1, 2$, by the Rosenthal inequality $\rho^{2i}\sqrt{\mathbb{E}((\varepsilon_\bullet^j)^{4i})} = O(\frac{\rho^{2i}}{p_N^i})$.

**3.0.0.1. Remark: .** In [9], Theorem 2.3., it is proved that

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

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 3.4, additionnal terms due to the noise appear.

## 4. Uniform convergence in probability results

In this section, asymptotic results for functionals of local means are stated. They are involved in the asymptotic study of the minimum contrast estimators described in Section 5.

From now on, $f : \mathbb{R} \times \Theta \to \mathbb{R}$ denotes a $\mathcal{C}^2$ function, such that $f$, $\partial_x f$, $\partial_{xx}^2 f$ and $\partial_\theta f$ satisfy **(C1)**. The assumptions on asymptotics are denoted **(AH)** :

**(AH)** The number of observations $N \to \infty$, with $\delta_N \to 0$, $p_N \to \infty$, $k_N \to \infty$, $\Delta_N = p_N\delta_N \to 0$ and $N\delta_N = k_N\Delta_N \to \infty$.

The first result is an ergodic theorem for the local means.

**Proposition 4.1:** *Under Assumptions* **(A1)-(A5)** *and* **(AH)**, *we have*

$$\bar{\nu}_N(f(.,\theta)) = \frac{1}{k_N}\sum_{j=0}^{k_N-1} f(Y_\bullet^j, \theta) \longrightarrow \nu_0(f(.,\theta)) \tag{15}$$

*uniformly in* $\theta$, *in probability.*

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

**Theorem 4.2:** *Under Assumptions* **(A1)-(A5)** *and* **(AH)**, *with* $\delta_N = p_N^{-\alpha}, \alpha \in (1,2]$, *(*$\Delta_N = p_N^{1-\alpha}$*) we have*

$$\bar{I}_N(f(.,\theta)) = \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1},\theta)(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1})) \xrightarrow{\mathbb{P}} 0 \tag{16}$$

*uniformly in* $\theta$.

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

**Theorem 4.3:** *Assume* **(A1)-(A5)** *and* **(AH)**.

*(1) If* $\delta_N = p_N^{-\alpha}$ *with* $\alpha \in (1,2)$ *(*$\Delta_N = p_N^{1-\alpha}$*), then*

$$\bar{Q}_N(f(.,\theta)) = \frac{1}{k_N \Delta_N} \sum_{j=1}^{k_N-2} f(Y_\bullet^{j-1},\theta)(Y_\bullet^{j+1} - Y_\bullet^j)^2 \xrightarrow{\mathbb{P}} \frac{2}{3}\nu_0(f(.,\theta)\sigma^2), \tag{17}$$

*(2) If* $\delta_N = p_N^{-2}$ *(*$\Delta_N = \frac{1}{p_N}$*), then*

$$\bar{Q}_N(f(.,\theta)) \xrightarrow{\mathbb{P}} \frac{2}{3}\nu_0(f(.,\theta)\sigma^2) + 2\rho^2\nu_0(f(.,\theta)), \tag{18}$$

*uniformly in* $\theta \in \Theta$.

The proofs of these two last theorems are based on the results of Proposition 3.4 and Lemma A.3 in the Appendix. Theorems 4.2 and 4.3 can be compared to the following results from [16]:

$$\frac{1}{k_N \Delta_N} \sum_{j=0}^{k_N-1} f(X_{j\Delta_N},\theta)(X_{(j+1)\Delta_N} - X_{j\Delta_N} - \Delta_N b(X_{j\Delta_N})) = o_P(1), \tag{19}$$

$$\frac{1}{k_N \Delta_N} \sum_{j=0}^{k_N-1} f(X_{j\Delta_N},\theta)(X_{(j+1)\Delta_N} - X_{j\Delta_N})^2 = \nu_0(f(.,\theta)\sigma^2) + o_P(1). \tag{20}$$

Theorem 4.2 gives the analogous result as (19), 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 $\Delta_N$ (if no lag, the limit is not 0, see for instance [11]). Theorem 4.3 underestimates $\nu_0(f(.,\theta)\sigma^2)$ because the variance of $\zeta_{j+1,N} + \zeta'_{j+2,N}$ (see Proposition 3.4) is equivalent to $\frac{2}{3}\Delta_N$ and not to $\Delta_N$. For $\delta_N = p_N^{-2}$, an additional bias appears due to the noise.

## 5. Estimation by contrast minimization

The main results about minimum contrast estimators using local means are described here. The contrasts presented in this section are inspired by the works of Kessler (see [16]) and Gloter (see [9] and [11]). They derive from the log-likelihood of independent Gaussian random variables of mean $X_{(j+1)\Delta_N} - X_{j\Delta_N} - \Delta_N b(X_{j\Delta_N}, \kappa)$ and variance $\Delta_N \sigma^2(X_{j\Delta_N}, \lambda)$ previously used to build a contrast for directly observed diffusions. Some corrections are needed to deal with the local means $(Y_\bullet^j)$, mainly justified by the asymptotic behaviour of the quadratic variation in Theorem 4.3. These constrasts have been modified in [11] to deal with parameter estimation for integrated diffusion processes.

### 5.1. *Definition of the contrasts*

Define

$$\mathcal{E}_N(\theta) = \sum_{j=1}^{k_N-2} \left\{ \frac{3}{2\Delta_N} \frac{(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa))^2}{\sigma^2(Y_\bullet^{j-1},\lambda)} + \log(\sigma^2(Y_\bullet^{j-1},\lambda)) \right\}. \quad (21)$$

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

$$\mathcal{E}_N^\rho(\theta) = \sum_{j=1}^{k_N-2} \left\{ \frac{3}{2\Delta_N} \frac{(Y_\bullet^{j+1} - Y_\bullet^j - \Delta_N b(Y_\bullet^{j-1},\kappa))^2}{c_{N,\rho}(Y_\bullet^{j-1},\lambda)} + \log(c_{N,\rho}(Y_\bullet^{j-1},\lambda)) \right\} \quad (22)$$

We have $\lim_{N\to\infty} c_{N,\rho}(x,\lambda) = c_\rho(x,\lambda)$ with $c_\rho(x,\lambda) = \sigma^2(x,\lambda)$ if $1 < \alpha < 2$ and $c_\rho(x,\lambda) = \sigma^2(x,\lambda) + 3\rho^2$ if $\alpha = 2$. Let $\hat\theta_N$ and $\hat\theta_N^\rho$ be the associated minimum contrast estimators, defined as any solution of

$$\hat\theta_N = \operatorname*{arginf}_{\theta\in\Theta} \mathcal{E}_N(\theta) \text{ and } \hat\theta_N^\rho = \operatorname*{arginf}_{\theta\in\Theta} \mathcal{E}_N^\rho(\theta). \quad (23)$$

**Theorem 5.1:** *Assume* **(A1)-(A7)**, *and consider* $\hat\theta_N$ *and* $\hat\theta_N^\rho$ *defined by (23).*

*(1) If* $\delta_N = p_N^{-\alpha}$, $\alpha \in (1,2)$, *the estimator* $\hat\theta_N$ *is consistent:* $\hat\theta_N \to \theta_0$ *in* $\mathbb{P}_{\theta_0}$ *probability.*
*(2) If* $\alpha \in (1,2]$, *the estimator* $\hat\theta_N^\rho$ *is consistent.*

Note that point 1 does not require the knowledge of $\rho$.

The parameter $\alpha$ links the number of observations $p_N$ in a subsample and the length $\Delta_N$ in the time-interval of this subsample, as $\Delta_N = p_N\delta_N = \delta_N^{1-\frac{1}{\alpha}}$. Tuning $\alpha$ depends on the total number of observations, to deal with a rather large number of observations in each subsample and denoise sufficiently each local mean (See Table 1 in Section 6 for a numerical example).

The limit value $\alpha = 2$ is determined by the apparition of the variance of the additional noise: there is not enough observations in each subsample to neglect $\rho^2$. Hence, the choice of the contrast $\mathcal{E}_N^\rho$ is motivated by the second result in Theorem 5.1.

### 5.2. *Estimation with unknown observation noise level*

Assuming **(B1)** with unknown $\rho$, 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 5.2:** *Assume* **(A1)-(A5)** *and* **(B1)**. *Then we have* $\hat\rho_N^2 \xrightarrow{\mathbb{P}} \rho^2$, *when* $N \to \infty$, *with* $\delta_N \to 0$ *and* $N\delta_N \to \infty$. *If, moreover,* $N\delta_N^2 \to 0$, $\sqrt{N}(\hat\rho_N^2 - \rho^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 3\rho^4)$.

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

**Corollary 5.3:** *Assume* **(A1)-(A7)**, **(B1)** *and* $\delta_N = p_N^{-\alpha}$ *with* $\alpha \in (1,2]$. *The estimator* $\hat{\theta}_N^{\hat{\rho}_N}$ *is consistent.*

## 6. Examples

In this section, simulation results are given for several examples of diffusion models on simulated data.

### 6.1. *Example 1. The Ornstein-Uhlenbeck process*

The hidden diffusion solves

$$dX_t = \kappa X_t dt + \lambda dB_t \tag{24}$$

with $\kappa < 0$ and $\lambda > 0$, and $X_0$ is deterministic or follows the stationary distribution of $X$. We 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

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

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

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

In Tables 1-5, the common distribution of $\varepsilon_{i\delta}$ is $\mathcal{N}(0,1)$ and Table 6 presents some results with other distributions. Tables 1, 2 and 3 give mean and variance of $\hat{\theta}_N$ for different values of $\delta, \alpha$ and $N$ ($\delta = p^{-\alpha}$). 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 variance in parenthesis.

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

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

First, we remark that the empirical variance is larger in the case $\alpha = 2$ than in the other cases. The parameter $\kappa_0$ is always underestimated, but the estimation of $\kappa_0$ is clearly improved as $N$ grows, and $\delta$ is close to 0. The estimation of $\lambda_0$ is better in Table 2 than in Table 1, and similar in Tables 2 and 3. The variance decreases strongly in the case $\alpha = 2$.

| $N = 20000, \delta = 0.005$ $(N\delta = 100, N\delta^2 = 0.5)$ $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$ | | | |
|---|---|---|---|
| | $\alpha = 1.35(p = 50, k = 400)$ | $\alpha = 1.5(p = 34, k = 588)$ | $\alpha = 2(p = 14, k = 1428)$ |
| $\hat{\kappa}_N$ ($10^2$ Var) | -0.74 (1.08) | -0.81 (1.47) | -0.87 (1.51) |
| $\hat{\lambda}_N^2$ ($10^3$ 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.

| $N = 100000, \delta = 0.001$ $(N\delta = 100, N\delta^2 = 0.1)$ $\kappa_0 = -1, \lambda_0 = 1, \rho^2 = 0.5$ | | | |
|---|---|---|---|
| | $\alpha = 1.3(p = 200, k = 500)$ | $\alpha = 1.5(p = 100, k = 1000)$ | $\alpha = 2(p = 32, k = 3125)$ |
| $\hat{\kappa}_N$ ($10^2$ Var) | -0.81 (1.36) | -0.89 (1.49) | -0.96 (1.95) |
| $\hat{\lambda}_N^2$ ($10^3$ 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.

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.

| $N = 10^5, \delta = 10^{-3}, \alpha = 1.5, \kappa_0 = -1, \lambda_0 = 1$ | | | | |
|---|---|---|---|---|
| | $\rho^2 = 0.1$ | $\rho^2 = 1$ | $\rho^2 = 2$ | $\rho^2 = 5$ |
| $\hat{\kappa}_N$ ($10^2$ Var) | -0.91 (1.49) | -0.89 (1.50) | -0.86 (1.75) | -0.83 (1.52) |
| $\hat{\lambda}_N^2$ ($10^3$ 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.

A strong bias appears for $\hat{\lambda}_N$ when $\rho^2$ is bigger than 1, whereas there are no significant changes in the estimation of the drift parameter $\kappa_0$. The empirical variances for the estimation of $\lambda_0$ also increases: the presence of noise in the observations contaminates the estimation of the diffusion coefficient in this case.

In Table 5, we study 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 variance 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.

The smallest value of $\lambda_0^2$ 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 $\lambda_0^2$, 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.

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 larger in this case than in the case of Gaussian distribution.

These simulations point out 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 $\kappa$ lower than the bias observed for values of $\alpha$ close to 1, and an empirical 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.

on simulations lower than the variance observed for $\alpha = 2$. The second remark concerns the number of observations: for $N = 5000$ observations, $\kappa$ is underestimated, for all the values of $\alpha$ considered. Thus, 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*

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. [7] and [16]). To compare the quality of the noise reduction and its influence on the estimation of the parameters, we compare the results for discrete observations with noise, based on the minimization of the contrast built on the $(Y_\bullet^j)$ (Tables 1, 2 and 3) 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 $\delta_N$-step of discretization are considered. The hidden diffusion $(X_t)$ is an Ornstein-Uhlenbeck process (24). The results based on the direct observations are given in Table 7.

| $\alpha = 1.5, \kappa_0 = -1, \lambda_0 = 1$, 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 $(1.98 \times 10^{-2})$ | 0.99 $(9.80 \times 10^{-3})$ | 1.00 $(4.30 \times 10^{-3})$ |

Table 7. Parameter estimation with direct observations of the Ornstein-Uhlenbeck model, for several numbers of observations.

The estimation of $\kappa_0$ 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. *Example 2. The Cox-Ingersoll-Ross process*

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

$$dX_t = (\kappa X_t + \kappa')dt + \lambda\sqrt{X_t}dB_t, \qquad X_0 = \eta, \tag{25}$$

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

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

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

where $(\varepsilon_{t_i})$ 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

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

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 consistency can be proved directly.

Explicit expressions for the estimator $\hat{\theta}_N = (\hat{\kappa}_N, \hat{\kappa''}_N, \hat{\lambda}^2_N)$ 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

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

Recall that the following explicit expressions for the estimator $\tilde{\theta}_N = (\tilde{\kappa}_N, \tilde{\kappa'}_N, \tilde{\lambda}^2_N)$ are available when the diffusion $(X_t)$ is directly observed ([16]):

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

and

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

Simulation results are presented in Table 8 (with noise) and Table 9 (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, we have chosen the value $\alpha = \frac{3}{2}$ as local mean size parameter.

$\kappa_0 = -2, \kappa''_0 = 1, \lambda_0 = 4, \rho^2 = 0.5, \alpha = 1.5$

| | $N = 5.10^3, \delta = 10^{-2}$ | $N = 2.10^4, \delta = 5.10^{-3}$ | $N = 10^5, \delta = 10^{-3}$ |
|---|---|---|---|
| $\hat{\kappa}_N$ ($10^2$ 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) |
| $\hat{\lambda}^2_N$ ($10^2$ Var) | 4.23 (37.61) | 4.35 (15.15) | 4.40 (8.15) |

Table 8. Parameter estimation for the Cox-Ingersoll-Ross process with a multiplicative noise for different values of $\alpha$.

In Table 8, we observe that $\kappa''_0 = 1$ is well estimated, whereas the estimation of $\kappa_0$ 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 $\lambda_0$, 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 9.

$\kappa_0 = -2, \kappa'_0 = 3, \lambda_0 = 4, \alpha = 1.5, \rho^2 = 0.5$

| | $N = 5.10^3, \delta = 10^{-2}$ | $N = 2.10^4, \delta = 5.10^{-3}$ | $N = 10^5, \delta = 10^{-3}$ |
|---|---|---|---|
| $\hat{\kappa}_N$ ($10^2$ 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) |
| $\hat{\lambda}^2_N$ ($10^2$ Var) | 4.11 (0.95) | 4.05 (0.20) | 4.01 (0.36) |

Table 9. Parameter estimation for the Cox-Ingersoll-Ross process with direct observations for different values of $\alpha$.

Notice that there is no bias in the estimation of $\kappa_0$ and $\kappa'_0$ for $N = 5000$ and $N = 20000$, contrary to the noisy case. Moreover, the estimation of $\lambda_0^2$ is more accurate, with a lower empirical variance for $\hat{\lambda}^2_N$.

## 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, without specific assumption on the distribution of the noise. The asymptotic normality is studied in a companion paper [4].

## Acknowledgements

The author is greatly indebtful to Valentine Genon-Catalot, his PhD advisor, for her help and her numerous comments, and Adeline Samson for the discussions about the numerical results and her insightful suggestions. He thanks also the anonymous referees for their comments.

## 8. Proofs

The following lemma, based on elementary computations, is mentioned in [9] and summarize the properties of the random variables $\xi_{j,N}$ and $\xi'_{j+1,N}$ defined in Section 3.

**Lemma 8.1:** *The random variables $\xi_{j,N}$ and $\xi'_{j+1,N}$ are independent and gaussian; $\xi_{j,N}$ is $\mathcal{G}^N_{j+1}$ measurable and independent of $\mathcal{G}^N_j$; $\xi'_{j+1,N}$ is $\mathcal{G}^N_{j+2}$ measurable and independent of $\mathcal{G}^N_{j+1}$. We will use the following expectations:*

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

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 8.2:** *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:*

$$\zeta_{j+1,N} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(k+1)\int_{I_{j,k}} dB_s, \quad \zeta'_{j+2,N} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(p_N-1-k)\int_{I_{j+1,k}} dB_s. \quad (26)$$

*Moreover, we have*

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

**Proof of Lemma 8.2** Using (8), we can rearrange terms to exhibit non-overlapping intervals, hence conditionally independent variables, and obtain (26). 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 $l = 1, 2$,

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

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

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

Then,

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

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

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

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 (6),

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

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

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

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

**Proof of Proposition 3.2** We have

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

where $\varepsilon\bullet^j$ is independent of $\mathcal{H}_j^N$. Proposition 2.2 in [9] states that, using the random variables (9),

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

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 $c > 0$

$$\mathbb{E}\left(|r_{j,N}|^k\middle|\,\mathcal{H}_j^N\right) = \mathbb{E}\left(|r_{j,N}|^k\middle|\,\mathcal{G}_j^N\right) \leq c(1 + |X_{j\Delta_N}|^k) \tag{27}$$

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 (10). With elementary computations on conditional expectation, we get (see notation (7))

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

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

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

which implies (27). Finally, $\mathbb{E}(|\varepsilon_\bullet^j|^k|\mathcal{H}_j^N) = \mathbb{E}(|\varepsilon_\bullet^j|^k)$ because $\varepsilon_\bullet^j$ is independent of $\mathcal{H}_j^N$. □

**Proof of Corollary 3.3** 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 $\theta \in \Theta$,

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

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 $\theta$ such that $D_j^2 = (\partial_x f(\tilde{Z},\theta))^2(Y_\bullet^j - X_{j\Delta_N})^2$ and

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

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

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

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

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

with $c$ independent of $\theta$. Using the Cauchy-Schwarz inequality, it comes

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

□

**Proof of Proposition 3.4** 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

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

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

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

By splitting $\Delta_N$ into $\Delta_N = (k+1)\delta_N + (p_N-k-1)\delta_N$ for all $k$, we get (see notation 8)

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

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

$$r_{j,N}^{(1)} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(k+1)\delta_N(b(X_{j\Delta_N+k\delta_N}) - b(Y_\bullet^j)), \tag{28}$$

$$s_{j,N}^{(1)} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(p_N-k-1)\delta_N(b(X_{(j+1)\Delta_N+k\delta_N}) - b(Y_\bullet^j)), \tag{29}$$

$$r_{j,N}^{(2)} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(k+1)\sigma(X_{j\Delta_N+k\delta_N})\int_{I_{j,k}} dB_s - \sigma(X_{j\Delta_N})\zeta_{j+1,N}, \tag{30}$$

$$s_{j,N}^{(2)} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(p_N-k-1)\sigma(X_{(j+1)\Delta_N+k\delta_N})\int_{I_{j+1,k}} dB_s - \sigma(X_{j\Delta_N})\zeta'_{j+2,N}, \tag{31}$$

$$r_{j,N}^{(3)} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(k+1)\int_{I_{j,k}}(b(X_s) - b(X_{j\Delta_N+k\delta_N}))ds, \tag{32}$$

$$s_{j,N}^{(3)} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(p_N-k-1)\int_{I_{j+1,k}}(b(X_s) - b(X_{(j+1)\Delta_N+k\delta_N}))ds, \tag{33}$$

$$r_{j,N}^{(4)} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(k+1)\int_{I_{j,k}}(\sigma(X_s) - \sigma(X_{j\Delta_N+k\delta_N}))dB_s, \tag{34}$$

$$s_{j,N}^{(4)} = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(p_N-k-1)\int_{I_{j+1,k}}(\sigma(X_s) - \sigma(X_{(j+1)\Delta_N+k\delta_N}))dB_s. \tag{35}$$

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

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

Now we deal with $\mathbb{E}((r_{j,N}^{(1)})^2|\mathcal{G}_j^N)$. With Proposition 3.3 and the Cauchy-Schwarz inequality, it comes

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

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

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

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 8.2, 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

$$f(s) = \frac{1}{p_N}\sum_{k=0}^{p_N-1}(k+1)(\sigma(X_{j\Delta_N+k\delta_N}) - \sigma(X_{j\Delta_N}))\mathbf{1}_{I_{j,k}}(s) \qquad (36)$$

With the Ito isometry and the Fubini theorem, we have

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

We use similar techniques with $r^{(4)}_{j,N}$ and $s^{(4)}_{j,N}$ to obtain

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

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

Now, using (28), (8), Lemma 8.2 and the Cauchy Schwarz inequality we have

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

Corollary 3.3 implies

$$|\mathbb{E}(r^{(1)}_{j,N}\zeta_{j+1,N}|\mathcal{G}^N_j)|\leq c\Delta_N^{\frac{3}{2}}(1+|X_{j\Delta_N}|+\rho\sqrt{\mathbb{E}((\varepsilon^j_\bullet)^2)})(\sqrt{\Delta_N}(1+|X_{j\Delta_N}|)+\rho(\mathbb{E}((\varepsilon^j_\bullet)^4))^{\frac{1}{4}}).$$

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

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 (36) and Corollary 3.3, we obtain

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

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

For $r^{(3)}_{j,N}$ (see (32)), we use the Cauchy Schwarz inequality:

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

Hence

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

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

With the Fubini theorem and the Ito isometry, we have

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

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 (36), the Burkholder - Davies - Gundy inequality and Proposition A.2, we have

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

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 3.3, we get

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

Analogously, using Proposition A.2, $\mathbb{E}((r_{j,N}^{(3)})^4|\mathcal{G}_j^N) \leq c\Delta_N^6(1+|X_{j\Delta_N}|^4)$. Finally, we get the bound for $\mathbb{E}(\tau_{j,N}^4|\mathcal{G}_j^N)$. □

**Proof of Proposition 4.1** By Lemma A.1, 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

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

Then, with (12), Assumptions **(A5)** and **(B1)**, and $\mathbb{E}((\varepsilon_\bullet^j)^2) = \frac{1}{p_N}$, the result holds.
□

**Proof of Theorem 4.2** We have

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

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

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

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

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

In order to prove the pointwise convergence in $\theta$ 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 3.4,

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

Using **(A4)**, this implies $\frac{1}{k_N\Delta_N}\sum_{1\le 3j\le k_N-2}\mathbb{E}(v_{3j,N}^{(3)}(\theta)|\mathcal{H}_{3j}^N) = o_P(1)$. 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

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

For $\ell = 2$,

$$\frac{1}{(k_N\Delta_N)^2}\sum_{1\le 3j\le k_N-2}\mathbb{E}((v^{(2)}_{3j,N})^2(\theta)|\mathcal{H}^N_{3j}) = \frac{1}{(k_N\Delta_N)^2}\sum_{1\le 3j\le k_N-2} f(Y^{3j-1}_\bullet,\theta)^2\rho^2\mathbb{E}((\varepsilon^{3j+1}_\bullet-\varepsilon^{3j}_\bullet)^2)$$
$$= \frac{2\rho^2}{N\delta_N p_N\Delta_N}\frac{1}{k_N}\sum_{1\le 3j\le k_N-2} f(Y^{3j-1}_\bullet,\theta)^2.$$

As $p_N\Delta_N = p_N^{2-\alpha}$, with $1<\alpha\le 2$, the above term is $o_P(1)$.
For $\ell = 3$,

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

using that, by Proposition 3.4, $\Delta_N^{-2}\mathbb{E}(\tau^2_{j,N}|\mathcal{H}^N_j)$ is $O_P(1)$.

To obtain uniformity in $\theta$, we shall use Proposition A.4 and evaluate $\sup_{N\in\mathbb{N}}\mathbb{E}(\sup_{\theta\in\Theta}|\partial_\theta\tilde{I}_N(f_\theta)|)$. To study

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

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

$$S^{(\ell)}_N(\theta) = \frac{1}{k_N\Delta_N}\sum_{1\le 3j\le k_N-2} v^{(\ell)}_{3j,N}(\theta).$$

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

$$\mathbb{E}(\sup_{\theta\in\Theta}|\partial_\theta v^{(3)}_{3j,N}(\theta)||\mathcal{H}^N_{3j}) \le c(1+|Y^{3j-1}_\bullet|)\sqrt{\mathbb{E}(\tau^2_{3j,N}|\mathcal{H}^N_{3j})}.$$

With the Cauchy Schwarz inequality, we have

$$\mathbb{E}(\sup_{\theta\in\Theta}|\partial_\theta v^{(3)}_{3j,N}(\theta)||\mathcal{H}^N_{3j}) \le c\sqrt{\Delta_N}(1+|Y^{3j-1}_\bullet|)(1+|X_{3j\Delta_N}|+\rho\sqrt{\mathbb{E}((\varepsilon^{3j}_\bullet)^2)})$$
$$\times(\sqrt{\Delta_N}(1+|X_{3j\Delta_N}|^2)+\rho\left(\mathbb{E}\left(\left(\varepsilon^{3j}_\bullet\right)^4\right)\right)^{\frac{1}{4}})$$

and with Lemma A.1 and **(A4)-(A5)**, this implies

$$\sup_{N\in\mathbb{N}}\mathbb{E}(\sup_{\theta\in\Theta}|\partial_\theta S^{(3)}_N(\theta)|) < \infty.$$

We cannot use the same method to study $S^{(\ell)}_N(\theta), \ell = 1,2$. Instead, we use Theorem 20 in Appendix 1 of [14]: it is enough to show that, for $\ell = 1,2$, there exists two constants $M\ge 0$ and $\epsilon>0$ such that:

$$\begin{aligned} &\forall\theta\in\Theta,\forall N\in\mathbb{N},\quad \mathbb{E}(|S^{(\ell)}_N|^{2+\epsilon})\le M\\ \text{and}\quad &\forall\theta,\theta'\in\Theta,\forall N\in\mathbb{N},\quad D_N(\theta,\theta')\le M|\theta-\theta'|^{2+\epsilon}\end{aligned} \tag{37}$$

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

For $\ell = 1$, using the Rosenthal inequality for martingales (see [13]), we get, for any $\epsilon > 0$:

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

Then it comes:

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

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

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

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 (37) holds. This implies $S_N^{(1)}(\theta) = o_P(1)$ uniformly in $\theta$.

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

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

Hence, with $\mathbb{E}\left((v_{3j,N}^{(2)}(\theta))^2\middle|\mathcal{H}_{3j}^N\right) = 2\rho^2 f(Y_\bullet^{3j-1},\theta)^2\sigma(X_{3j\Delta_N})^2\mathbb{E}((\varepsilon_\bullet^{3j})^2)$ and $\mathbb{E}((\varepsilon_\bullet^{3j})^2) = \frac{1}{p_N}$, and $\Delta_N = p_N^{1-\alpha}$, we obtain (37). Finally $\tilde{I}_N(f_\theta) = o_P(1)$, uniformly in $\theta$. □

**Proof of Theorem 4.3** Let $W_{j,N}(\theta) = f(Y_\bullet^{j-1},\theta)(Y_\bullet^{j+1} - Y_\bullet^j)^2$. By Proposition

3.4, we have $W_{j,N}(\theta) = \sum_{i=1}^{6} w_{j,N}^{(i)}(\theta)$ with

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

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 $\ell = 0,1,2$,

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

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

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

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

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

and

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

Recall that $\Delta_N = p_N^{1-\alpha}$, $1<\alpha\le 2$. 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 $\rho=\rho$, we have $T_{0,N}^{(2)}(\theta) = \frac{1}{3}\times 2\rho^2\nu_0(f(.,\theta)^2)+o_P(1)$. and

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

We easily deduce from Proposition 3.4, Lemma A.3 and Lemma A.1 that $T_{0,N}^{(3)}(\theta) = o_P(1)$. For $T_{0,N}^{(4)}(\theta)$, we have

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

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

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

Then, with Proposition 3.4, Lemma A.3 and Lemma A.1, $T_{0,N}^{(4)}(\theta) = o_P(1)$.

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,

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

Moreover, with the Cauchy Schwarz inequality,

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

Then, with Proposition 3.4, Lemma A.3 and Lemma A.1 $T_{0,N}^{(5)} = o_P(1)$.

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

We prove now uniformity in $\theta$ in these convergences, using Proposition A.4. 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

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

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

**Proof of Lemma 5.2** We have $\hat{\rho^2}_N - \rho^2 = a_{1,N} + a_{2,N} + a_{3,N}$ where

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

With the usual law of large numbers, $a_{1,N} = o_P(1)$. With Proposition A.2,

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

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 $N \to \infty$ 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

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

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

**Proof of Theorem 5.1.** For this proof, recall that $b(.) = b(., \kappa_0)$, $c(.) = c(., \lambda_0)$ denote the drift and diffusion coefficients at the true value $\theta_0$.

The steps of the proof of the convergence of $\hat{\theta}_N$ to $\theta_0$ are similar to Section 4 of [16], and we only give details here for the convergence of $\frac{1}{k_N}\mathcal{E}_N(\theta)$ and $\frac{1}{k_N\Delta_N}(\mathcal{E}_N(\kappa, \lambda) - \mathcal{E}_N(\kappa_0, \lambda))$.

Developping $\mathcal{E}_N(\theta)$ (see (21)) yields:

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

Proposition 4.1, Theorem 4.2 and Theorem 4.3 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 [16] and [11]. It is enough to prove that, first,

$$\frac{1}{k_N}\mathcal{E}_N(\theta) \underset{N\to\infty}{\longrightarrow} \nu_0\left(\frac{c(., \lambda_0)}{c(., \lambda)} + \log(c(., \lambda))\right) \tag{38}$$

in probability, uniformly in $\theta$. This will ensure the convergence of $\hat{\lambda}_N$ to $\lambda_0$ (as in Theorem 1 of [16]). Second, we prove that

$$\frac{1}{k_N\Delta_N}(\mathcal{E}_N(\kappa, \lambda) - \mathcal{E}_N(\kappa_0, \lambda)) \underset{N\to\infty}{\longrightarrow} \frac{3}{2}\nu_0\left(\frac{(b(., \kappa) - b(., \kappa_0))^2}{c(., \lambda)}\right) \tag{39}$$

in probability, uniformly in $\theta$. Using Theorem 4.2, Theorem 4.3 and Proposition 4.1, with $\Delta_N \to 0$ we obtain (38) and (39).

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

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

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

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

and

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

The end of the proof is identical, replacing $\mathcal{E}_N$ by $\mathcal{E}_N^\rho$ and $c$ by $c_\rho$ in the limits (38)-(39).□

**Proof of Corollary 5.3.** As formerly, we evaluate

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

We conclude using Lemma 5.2. □

## References


[1] Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, 70(1):223–262.

[2] Bibby, B. M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1(1-2):17–39.

[3] 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.

[4] Favetto, B. (2010). Estimating functions for noisy observations of ergodic diffusions. Preprint MAP5 2010-30 available at http://hal.archives-ouvertes.fr/hal-00531096/fr/.

[5] Favetto, B. and Samson, A. (2010). Parameter estimation for a bidimensional partially observed ornstein–uhlenbeck process with biological application. *Scandinavian Journal of Statistics*, 37(2):200–220.

[6] Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, 20(4):547–557.

[7] Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations. *Statistics*, 21(1):99–116.

[8] 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.

[9] Gloter, A. (2000). Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Statist.*, 4:205–227 (electronic).

[10] Gloter, A. (2001). Parameter estimation for a discrete sampling of an integrated Ornstein-Uhlenbeck process. *Statistics*, 35(3):225–243.

[11] Gloter, A. (2006). Parameter estimation for a discretely observed integrated diffusion process. *Scand. J. Statist.*, 33(1):83–104.

[12] Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. II. Optimal estimators. *ESAIM Probab. Statist.*, 5:243–260 (electronic).

[13] 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.

[14] Ibragimov, I. A. and Has′minskiĭ, R. Z. (1979). *Asimptoticheskaya teoriya otsenivaniya*. "Nauka", Moscow.

[15] 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.

[16] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, 24(2):211–229.

[17] Pedersen, A. (1994). Statistical analysis of gaussian diffusion processes based on incomplete discrete observations. *Research Report, Department of Theoretical Statistics, University of Aarhus*, 297.

[18] 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.

[19] Schmisser, E. (2010). Non-parametric estimation of the diffusion coefficient from noisy data. Preprint available at http://hal.archives-ouvertes.fr/hal-00443993/en/.
[20] Schmisser, E. (2011). Nonparametric drift estimation for diffusions from noisy data. *Statistics & Decisions*, 28(2):119–150.
[21] Sørensen, M. (2009). Efficient estimation for ergodic diffusions sampled at high frequency. Preprint available at http://www.math.ku.dk/˜michael/efficient.pdf.
[22] Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.*, 41(2):220–242.
[23] 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 A. Appendix

The following lemma can be found in [11], and precises a result from [16] :

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

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

*then:*

$$\frac{1}{k_N} \sum_{j=0}^{k_N-1} f(X_{j\Delta_N}, \theta) \underset{k_N \to \infty}{\longrightarrow} \nu_0(f(.,\theta)) \tag{A1}$$

*uniformly in $\theta$, in probability.*

The following proposition can be found in [9] and [11], and the numerical constant $c$ may varies.

**Proposition A.2:** *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 [8], setting $\mathcal{G}_j^N = \mathcal{G}_{j\Delta_N}$

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

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

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

The following proposition is given in [11], to obtain convergences in probability uniformly in $\theta$.

**Proposition A.4:** *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 $\Theta$ is product of compact intervals of $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. We assume that $S_n(.,\theta)$ converges to zero in probability for all $\theta \in \Theta$ and that there exists an open neigbourhood of $\Theta$ on which $S_n(\omega, .)$ 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*

$$S_n(\theta) \to 0$$

*uniformly in $\theta$, in probability.*